



**HAL**  
open science

## ESPON Database 2013 Technical reports

Claude Grasland, Maher Ben Rebah, Ronan Ysebaert, Christine Zanin,  
Nicolas Lambert, Bernard Corminboeuf, Isabelle Salmon, Jérôme Gensel,  
Bogdan Moisuc, Marlène Villanova-Oliver, et al.

► **To cite this version:**

Claude Grasland, Maher Ben Rebah, Ronan Ysebaert, Christine Zanin, Nicolas Lambert, et al..  
ESPON Database 2013 Technical reports. [Research Report] ESPON | Inspire Policy Making with  
Territorial Evidence. 2011. hal-03595672

**HAL Id: hal-03595672**

**<https://hal.archives-ouvertes.fr/hal-03595672>**

Submitted on 3 Mar 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

# TECHNICAL REPORTS

## Annexed to the Final Report of the ESPON 2013 Database Project



### METHODS

#### ❖ The ESPON DB Application

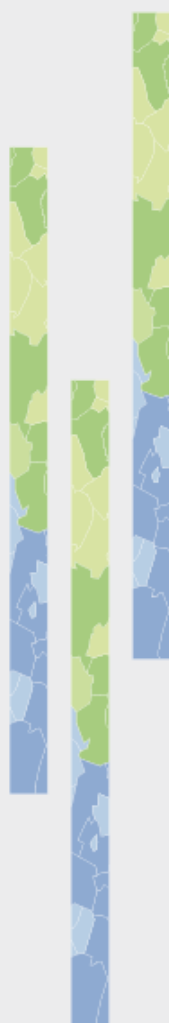
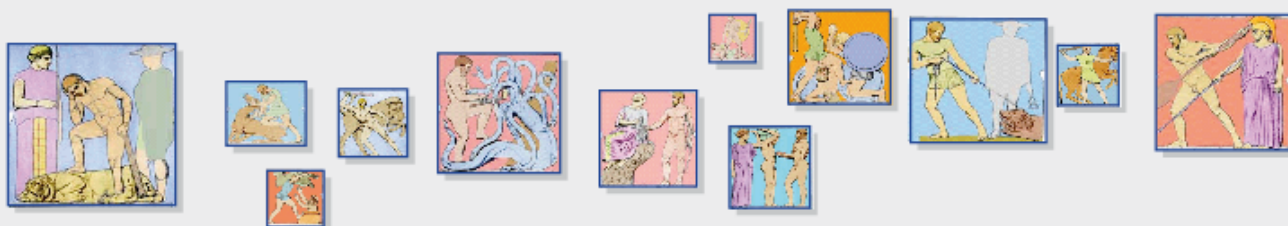
- *The ESPON Database application.*
- *A two-step approach to structure the ESPON 2013 DB by themes and sub-themes*
- *Coding scheme to label indicators: Guiding principles to TPGs involved in ongoing and future ESPON projects.*
- *Towards an ESPON thesaurus? Some preliminary considerations for the thematic structuring of the ESPON database.*
- *UPDATE OF ESPON DATABASE 2006 INTO THE 2013 VERSION.*
- *Listing of ESPON Database indicators.*

#### ❖ Methodological aspects

- *Modeling territorial changes and time series database building process: empirical approach and applications.*
- *Disaggregation of socioeconomic data into a regular grid and combination with other types of data.*
- *Spatial analysis for quality control (v1 & v2).*
- *Using downscaled population in local data generation.*
- *Spatial data creation for Europe.*
- *Mapping guide – Cartography for ESPON projects.*
- *Mapping guide – Cartography in ESPON 2013.*

#### ❖ Enlargement of the data collection

- *World database: Global data for the ESPON 2013 Database*
- *Analysis of the availability and the quality of data on Western Balkans and Turkey.*
- *Local & regional data First investigations in Romania, Bulgaria, Czech Republic and Slovakia.*
- *Local & regional data producing innovative indicators at local scale.*
- *Naming UMZ: a database now operational for urban studies;*
- *LUZ specifications (Urban Audit 2004)*
- *The Functional Urban Areas database.*



## The ESPON Database Application

### CONTENT

The ESPON 2013 Database Application is a complex information system dedicated to the management of statistical data about the European territory, spanning over a long period of time. The overall architecture relies on two databases: one is used for storing ontological data, and the other, called the ESPON Database, is meant to be queried by end-users. The latter only is made accessible to users through Web interfaces that each correspond to the four main functionalities offered by the ESPON 2013 Database Application: registration, administration, upload of both data and metadata, query and retrieval of such data and metadata.

**ESPON 2013 DATABASE**  
**MARCH 2011**



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

38 pages

# LIST OF AUTHORS

Bogdan Moisuc, LIG Steamer  
Jérôme Gensel, LIG Steamer  
Anton Telechev, LIG Steamer  
Benoit Le Rubrus, LIG Steamer

## Contact

[moisuc@imag.fr](mailto:moisuc@imag.fr)

[jerome.gensel@imag.fr](mailto:jerome.gensel@imag.fr)

[anton.telechev@imag.fr](mailto:anton.telechev@imag.fr)

[benoit.le-rubrus@imag.fr](mailto:benoit.le-rubrus@imag.fr)

.

LIG Grenoble  
Bâtiment ENSIMAG D  
681 rue de la Passerelle  
38400 Saint Martin d'Hères  
(+33) 4 76 82 72 11

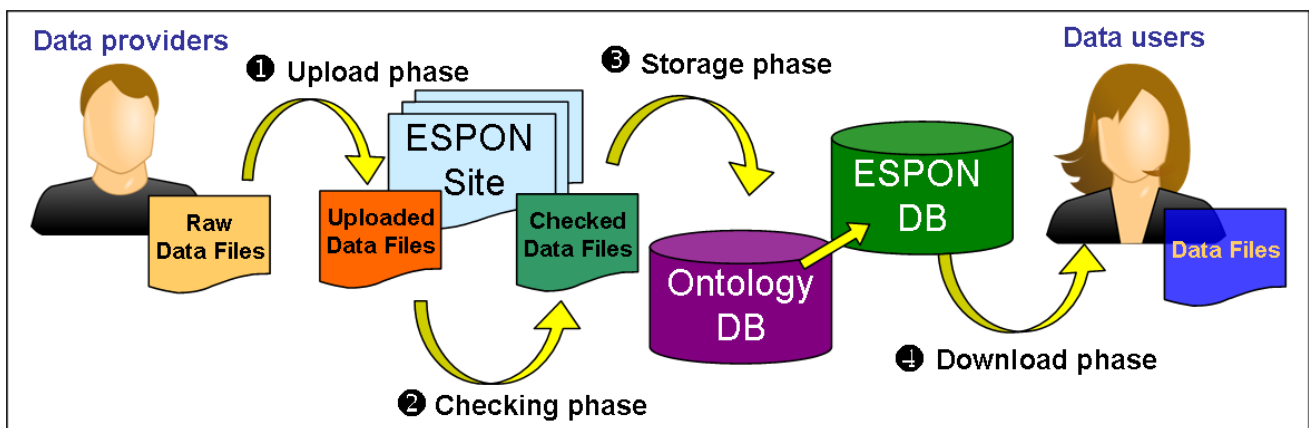
# TABLE OF CONTENTS

<b>Introduction.....</b>	<b>3</b>
<b>1 ESPON DB Web Interface .....</b>	<b>5</b>
1.1 Registration page.....	5
1.2 Login page .....	6
1.3 Account page (registered users only).....	8
1.4 Help page .....	9
1.5 Search page .....	10
1.6 Basket page .....	17
1.7 Upload page (registered users only) .....	19
1.8 Administration page (administrators only).....	24
<b>2 The Databases .....</b>	<b>26</b>
2.1 The ontology database .....	26
2.2 The Espon database .....	30
2.2.1 Core data structures in ESPON 2013 Database .....	31
2.2.2 Data search algorithm .....	33
2.2.3 Data completeness calculation .....	35
<b>Conclusion and future works.....</b>	<b>37</b>

# Introduction

This technical report describes the main features of the ESPON Database Application. The ESPON 2013 Database Application is a complex information system dedicated to the management of statistical data about the European territory, spanning over a long period of time. The part that is visible to the users is the Web interface of the application. Equally important is the hidden part, relying on two databases: one is used for storing ontology data, and the other, called the ESPON Database, is meant to be queried by end-users.

The ESPON 2013 is the path followed by both data and metadata from the moment they are entered in the ESPON DB Application, until they are output as answers to queries expressed by end-users (see **Figure 1**). Four phases are identified along this data flow:



**Figure 1:** The ESPON 2013 DB data flow.

1. The upload phase is handled by the upload Web interface through which users (here, data providers) are guided in the preparation and the transfer of both their data and metadata files to the ESPON Database server. During this phase, users are helped in providing well formatted and Inspire compliant metadata through the ESPON Metadata Editor.

2. The checking phase follows; it aims at validating both data and metadata files provided by users before they are stored in the ESPON Database. The checking process alternates between automatic and manual steps performed either by the application itself or by the expert members of the ESPON DB 2013 Project. If some of the errors detected cannot be corrected or need some additional information and precisions, then both data and metadata files are sent back to providers in order to be fixed. When the checking phase succeeds, then the validated data and metadata files are ready to be stored in the ESPON Database.

3. The storage phase deals with the management and the maintenance of both data and metadata in the ESPON Database. Flexible database schemas have been designed and built for handling long term storage of statistical and spatial data, considering that both data and metadata may evolve while stored in the ESPON Database, as a result of harmonization and gap filling processes.

4. During the download phase, end-users of the ESPON DB Application are invited to explore, search and retrieve both data and metadata through a

Web interface. Free data and metadata can be accessed and downloaded by any end-user, while data and metadata subject to copyright restrictions are made available for authorized and registered users only.

The first section of the document shows the screens and features of the Web application interface version 1.0, also known as the ESPON 2013 Database Web extraction tool. This gives a *vademecum* for the application by following a typical user session, from the authentication to the download of query results.

The second section of this document contains a description of the databases upon which this application is based: the ontology base and the ESPON database, and of the software that allows filling them.

# 1 ESPON DB Web Interface

The ESPON DB Web download interface is an on-line application designed to offer fast browsing and searching capabilities over the ESPON DB. The Web download interface implements several innovative elements that guaranties scalable performance to accommodate the fast growing size of the ESPON DB :

- The use of a server-side application cache system allows the application to avoid querying the database for all browsing tasks except the advanced search. This insures fast data searching, whatever the database size.
- The use of an XML exchange format for the answer to queries allows decreasing the size of the data transfers between server and client.
- The use of AJAX techniques (Asynchronous JavaScript and XML) allows further decreasing the size of the traffic between the client (Web browser) and the server (ESPON Web site), by transferring only the parts of query that have changed (in XML) and redisplaying them accordingly on the client (using JavaScript). This allows for load balancing between client and server, as the task of building the presentation from the XML file is performed on the client.
- The dropdown lists used in the interface have been developed as new components in order to match the ESPON look & feel requirements.

The underlying subsections contain a description of the different pages that can be displayed by the application. The order of the pages follows the logical order of the navigation performed by a common user. The starting page of the application is the search page (see section 1.5), but any user willing to use the full functionality of the ESPON DB Application will have to register first (see section 1.1). A registered user can login to the application (see section 1.2), update his registration information (see section 1.3), see the help page (see section 1.4), use the application to search (see section 1.5) and download data (see section 1.6). If users are data providers, they can also upload their own information into the ESPON Database (see section 1.7). Privileged users (administrators of the application) can also change the application settings and manage user registration requests (see section 1.8).

## 1.1 Registration page



Search Basket Log in Register Terms&Conditions

### ESPON 2013 Database Registration Form

Most of the data available in the ESPON database is downloadable without registration. You can directly download datasets by querying the database.

Registration is only needed for downloading data protected by copyrights and is only available for teams involved in ESPON Projects (Priorities 1 to 4). In other terms, it concerns only Lead Partner, Project Partners, ESPON Contact Points and Monitoring Committee.

Thank you for your cooperation.  
The ESPON 2013 Database Project Team

In order to make operational your registration please fill these mandatory fields:

Status

ESPON Project

Organization

Family name

First name

E-mail

Phone

**Figure 2:** The registration page of the ESPON DB Web Application.

The registration form contains a small number of fields required for establishing the identity of the user (first name, family name and organization) their contact coordinates (telephone number and email address) and their relationships with ESPON (the ESPON Project which they are members of and their function within this project).

Once the users submit the form, they receive a confirmation email. In the same time, the database administrator is notified for the registration request. The administrator can check the information entered by the user and validate or not the registration request, via the administration interface (see section 1.8 for the administration page). Once the administrator has validated the user registration, the user receives via email Administration page (administrators only) their login and password for the application.

## 1.2 Login page

When typing the URL of the ESPON 2013 Database Web application in his/her browser address bar, any user is first invited to choose between both following types of login:

- *anonymous* browsing;
- *registered account* login.



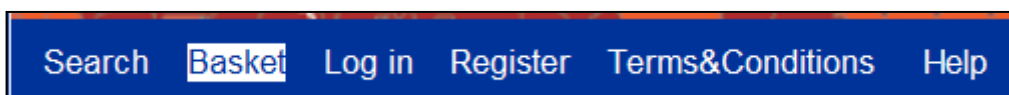
**Figure 3:** The login page of the ESPON DB Web Application.

The differences between these two ways of browsing essentially consist in the features the user will be offered by the application. Roughly speaking, a registered user will be allowed to access the whole set of pages that can be delivered by the application; an anonymous user will be allowed to search and download results from only a subset of available data. Also, anonymous users cannot upload data or edit metadata. Besides the login fields (*Login* and *Password*), **Figure 3** shows that the login page displays a link in order to ask for a new login or password, if the users have lost the old ones.

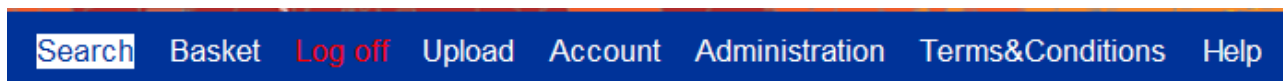
On the login page, there is a link that redirects the user towards a registration form. The same form can be accessed by clicking on the menu item "Register", which is only visible to unregistered users. Upon clicking on one of these links, the user arrives on the registration form page.

Although the layout of the pages is identical for both types of session (logged in or anonymous), the displayed menu bar on the header of pages is different.

**Figure 4** shows the menu bar for an anonymous session, **Figure 5** shows the menu bar for a registered user. The main difference between both menus is the availability of the upload page for a registered user (described in section 1.7) and of the account page (described in section 1.3).



*Figure 4: Menu bar for an anonymous session.*



*Figure 5: Menu bar for a registered user.*

For registered users, the authentication leads them to the ESPON Database Web Interface Home Page. Figure 5 shows that this menu bar gives shortcuts the available features of the Web application:

- the home page;
- a page to update one's account;
- the search page;
- the basket page;
- the upload page;
- a link to quit the session;
- the on-line help.

### **1.3 Account page (registered users only)**

The account page allows users to update and complete their personal information and to change their password for accessing the application. In order to update any profile element, the users are required to enter their old password.

**Your user profile data**

**Congratulations!**  
Your profile has been updated successfully.

Your registration data

Your registration date: 2010-07-20  
Your status : Administrator

[You have not accepted the Terms&Conditions Agreement yet.](#)

Your personal data

First name:

Last name:

Email:

Phone number:  ▲

Your ESPON projects affiliation

ESPON project:  ▲

Lead partner:  ▲

Organization:  ▲

Your application identifiers

Login:  ▲

Current password:  ▲

New password:  ▲

New password (repeat):  ▲

**Figure 6:** The account page of the ESPON DB Web Application.

From the account page, users can also see if they have agreed with the terms and conditions of the ESPON Database Application. As they are already registered, they only need to agree with the T&C once. Unregistered users have to accept the T&C at each session.

## 1.4 Help page

Help contents

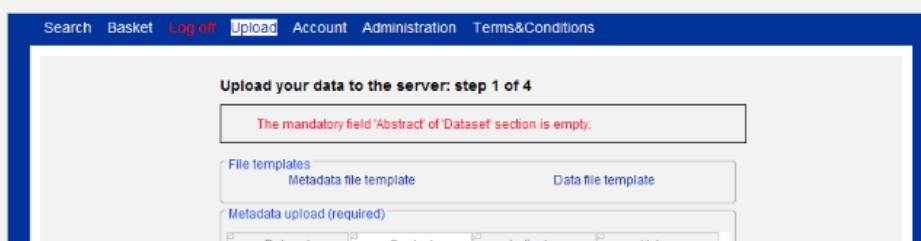
**The Uploads page (registered users only)**

The upload page aims at allowing ESPON project members to upload their data deliveries in order to fill the ESPON 2013 Database. The ESPON 2013 Database Application allows the users to upload 3 types of files to the ESPON server:

- metadata files;
- data files;
- additional documentation files (technical reports, etc., describing in more detail the different methodologies used for calculating indicators or for estimating missing data).

This is typically done in a 4 step process.

In the first step users are invited to upload a metadata file on the server. However, the step can be much more complex. When loading a metadata file within the application, the users are prompted the existing errors in their document. Errors can be either syntactical errors, or simply missing information for mandatory fields. Fields with errors are highlighted in red and a textual description of the error is also given in the header of the metadata editor (see an example in Figure 1).



**Figure 7:** The help page of the ESPON DB Web Application.

The help page of the ESPON 2013 Database Application is aimed at helping users navigate easier through the pages of the ESPON 2013 Database Application (see **Figure 7**). The help contains mainly the same *vademecum* as the first chapter of the present document, that is, the logical succession of pages (with their explanations) navigated in a typical user session.

An index allows the users to visit the six categories of the online help:

1. Registration page
2. Login page
3. Profile page (registered users)
4. Search page
5. Basket page
6. Upload page (registered users)
7. Help page

From each of these pages, a link to the help content (index) is provided.

## 1.5 Search page

The ESPON 2013 Database Application offers currently two types of searches: a search by project (the data providers) and a search by theme (according to the ESPON 2013 Database thematic classification).

For each type of search, two modes of research can be performed:

- simple search;

- advanced search.

The screenshot shows a search interface with a blue background. At the top right, there is a toggle switch labeled 'Simple search'. Below it, there are five main selection panels:

- Selection by study area:** A list with options: EU15, EU25, EU27, ESPON31, CC, EFTA, and By country. Each option has a red arrow icon to its right.
- Selection by geographic object:** A panel with 'Select all' and 'Unselect' buttons at the top. Below are 'Regional data' (with a red arrow) and 'Local data'.
- Selection by NUTS revision:** A panel with 'Select all' and 'Unselect' buttons at the top. Below is a list of years: 1980, 1988, 1995, 1999, 2003, 2006, and 2010.
- Selection by NUTS level:** A panel with 'Select all' and 'Unselect' buttons at the top. Below are: Country level (NUTS0), Regional level (NUTS1), Subregional level (NUTS2), and Departamental level (NUTS3).
- Selection by covered period:** A panel with a red arrow icon at the bottom.
- Selection by publication date:** A panel with a red arrow icon at the bottom.

**Figure 8:** The advanced search criteria.

The user may switch from one mode to the other mode by clicking a button situated in the middle of the screen. In advanced mode (see **Figure 8**), a series of additional criteria can be added to a search in order to narrow down the results:

- study area (which means choosing either one of the groups of countries – EU15, EEFTA, etc., or choosing countries one by one);
- geographic object type (regional data, world data, local data etc.);
- covered period;
- publication date.

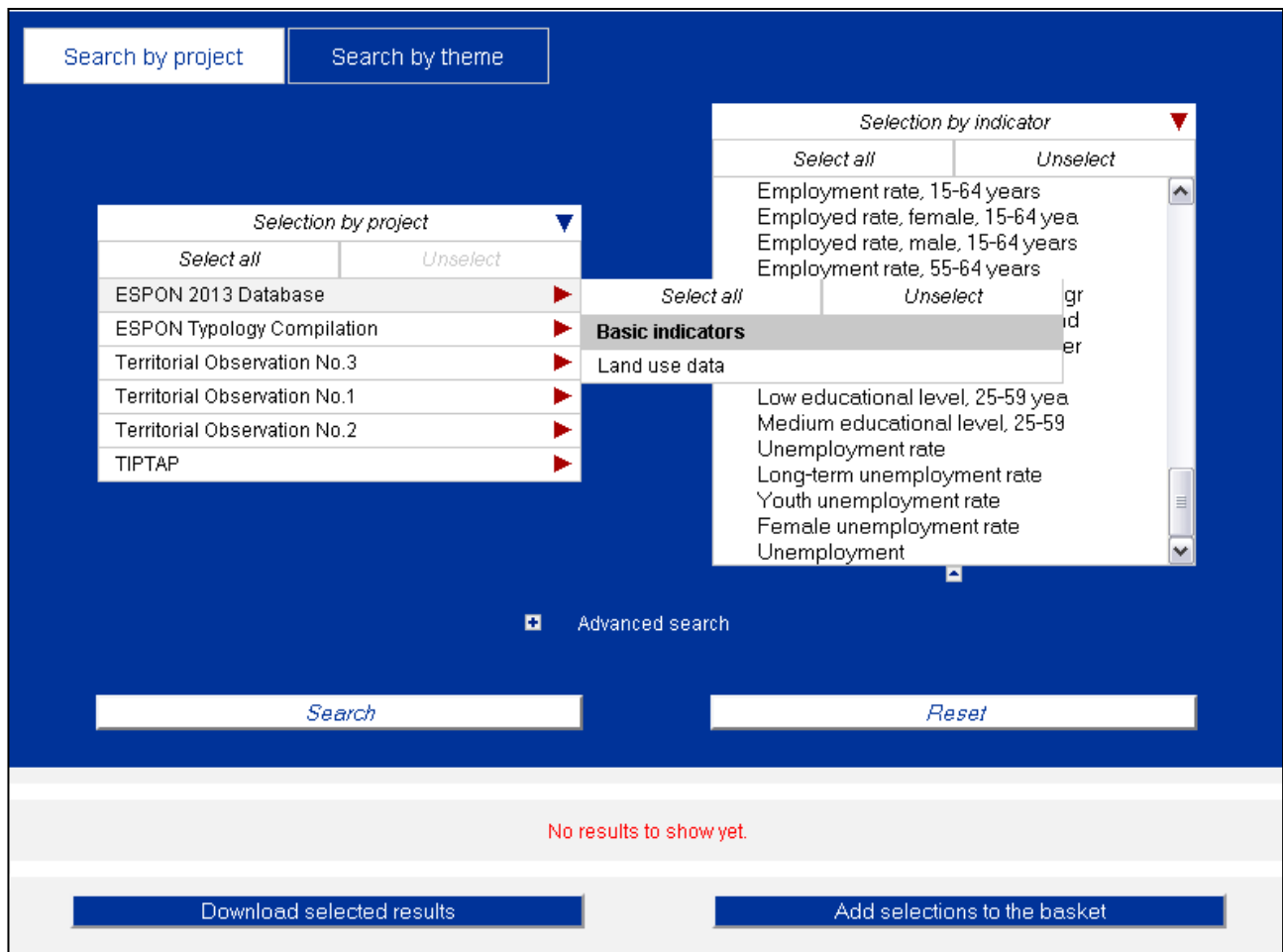
Once the user has chose a type of geographic object, two new search criteria appear:

- nomenclature revision (e.g. NUTS 2003, NUTS 2006, etc.)
- nomenclature level (e.g. NUTS 2, NUTS 3, etc.)

For each type of search, the user must select at least one mandatory field in order for the query to return any result.

For the search by project (see **Figure 9**), the user has to select at least one (possibly all) project in order to perform a query. When moving the mouse over one project, a sublist with all the available datasets for that project becomes visible, allowing users to refine their search. Multiple datasets and/or projects can be selected. Once a dataset is clicked, all the indicators available

in it are displayed in the indicator list displayed on the right. The window with the indicator list can be maximized by clicking on the small button below.



**Figure 9:** The search by project.

For the search by theme (see **Figure 10** **Figure 9**), the user has to select at least one (possibly all) theme in order to perform a query. When hovering over one theme, a sublist with all the available subthemes for that theme becomes visible, allowing users to refine their search. Once a theme or a subtheme is clicked, all the indicators available in it are displayed in the indicator list on the right. Multiple themes and/or subthemes can be selected.



**Figure 10:** The search by theme.

As previously mentioned, the query is performed only if the user has selected at least one criteria (theme, subtheme, project or dataset). If this requirement is not met, clicking the “Search” button will return a warning message. Otherwise, the application performs the query and the results are displayed in the same page, under the form of a result table, as shown on **Figure 11: The result table, displayed by datasets.**

2 items found							
<input type="checkbox"/>	Data set	Project	Geo objects	Study areas	Years	Completeness	Metadata
<input type="checkbox"/>	Basic indicators	ESPON 2013 Database	NUTS and similar 1999 <input checked="" type="checkbox"/> 2003 <input checked="" type="checkbox"/> 2006 <input checked="" type="checkbox"/>	(ESPON31)	1987 more... 2007	86% Show by levels	<input type="button" value="▶"/>
<input type="checkbox"/>	Land use data	ESPON 2013 Database	NUTS and similar	(EU27) more...	2000 <input checked="" type="checkbox"/>	70% Show by levels	<input type="button" value="▶"/>


**Figure 11:** The result table, displayed by datasets.

The screenshot above shows that the query returned two results (two datasets). Each result is displayed on one line. Considering this table of results, the displayed information for each result is respectively made of the following columns:

1. First column (no header title): the checkbox allows to select the current result in order to perform two possible actions:
  - adding this result to the basket;
  - immediately downloading this result.



The basket is roughly speaking a temporary area where the user can save multiple search results while he/she performs various queries. This basket functionality is further described in section 1.6.

2. Dataset column: displays the name of the dataset which is concerned by this result item. Clicking on this name displays all the indicators that have been delivered within this dataset. The indicator names can be selected and unselected for inclusion into the basket or in the direct download. The  icon is displayed if the result concerns data with copyright constraints attached. This case implies both behaviors:
  - if the user is anonymous, the checkbox in the second column is disabled: though he/she can see this result and the associated metadata, he/she is not allowed to download it;
  - if the user has logged in with a registered account, the checkbox is enabled. Nevertheless, he/she is warned about the confidential status of data for this result.
3. Project: this field shows the project that provided the data of this result.
4. Geographic object column: the current version of the application only manages NUTS. Clicking this name displays all the nomenclature versions for which data have been delivered within this dataset.
5. Study area: this field shows the spatial coverage of the current result. Whenever the countries present in the result can be aggregated by groups, they are aggregated to the largest group possible, when aggregation is no longer possible, the remaining countries are aggregated as individual countries (e.g. "EU 15 + IS, NO").
6. Covered period: this field shows the temporal coverage of the current result. When data range over a period of more than 3 years, a clickable "more" label allows the users to see the complete list of years. When there are three years or less, they are all visible by default.
7. Completeness: the percentage of the global completeness of the current result is represented by a colored bar. Long blue bar: high rate; short blue bar: low rate. A clickable "Show by levels" text allows having a more refined view of the completeness, per nomenclature level.

7 items found						
Indicator	Data set	Geo objects	Study areas	Years	Completeness	Metadata
<input type="checkbox"/> Unemployment	Basic indicators	NUTS and similar	(ESPON31)	2000 more... 2007	95% Show by levels	
<input type="checkbox"/> GDP Growth	Basic indicators	NUTS and similar	(EU15) more...	1995 more... 2004	76% Show by levels	
<input type="checkbox"/> Youth unemployment rate	Basic indicators	NUTS and similar	(EU27)	1999 <input checked="" type="checkbox"/> 2002 <input checked="" type="checkbox"/> 2005 <input checked="" type="checkbox"/>	73% Show by levels	
<input type="checkbox"/> Unemployment rate	Basic indicators	NUTS and similar	(EU27)	1989 more... 2005	72% Show by levels	
<input type="checkbox"/> Female unemployment rate	Basic indicators	NUTS and similar	(EU27)	1999 <input checked="" type="checkbox"/> 2002 <input checked="" type="checkbox"/> 2005 <input checked="" type="checkbox"/>	71% Show by levels	
<input type="checkbox"/> Long-term unemployment rate	Basic indicators	NUTS and similar	(EU27)	1999 <input checked="" type="checkbox"/> 2002 <input checked="" type="checkbox"/> 2005 <input checked="" type="checkbox"/>	71% Show by levels	
<input type="checkbox"/> Water courses	Land use data	NUTS and similar	(EU15) more...	2000 <input checked="" type="checkbox"/>	61% Show by levels	

Figure 12: The result table, displayed by indicators.

Dataset **Indicators** Sources Completeness

### Indicators information

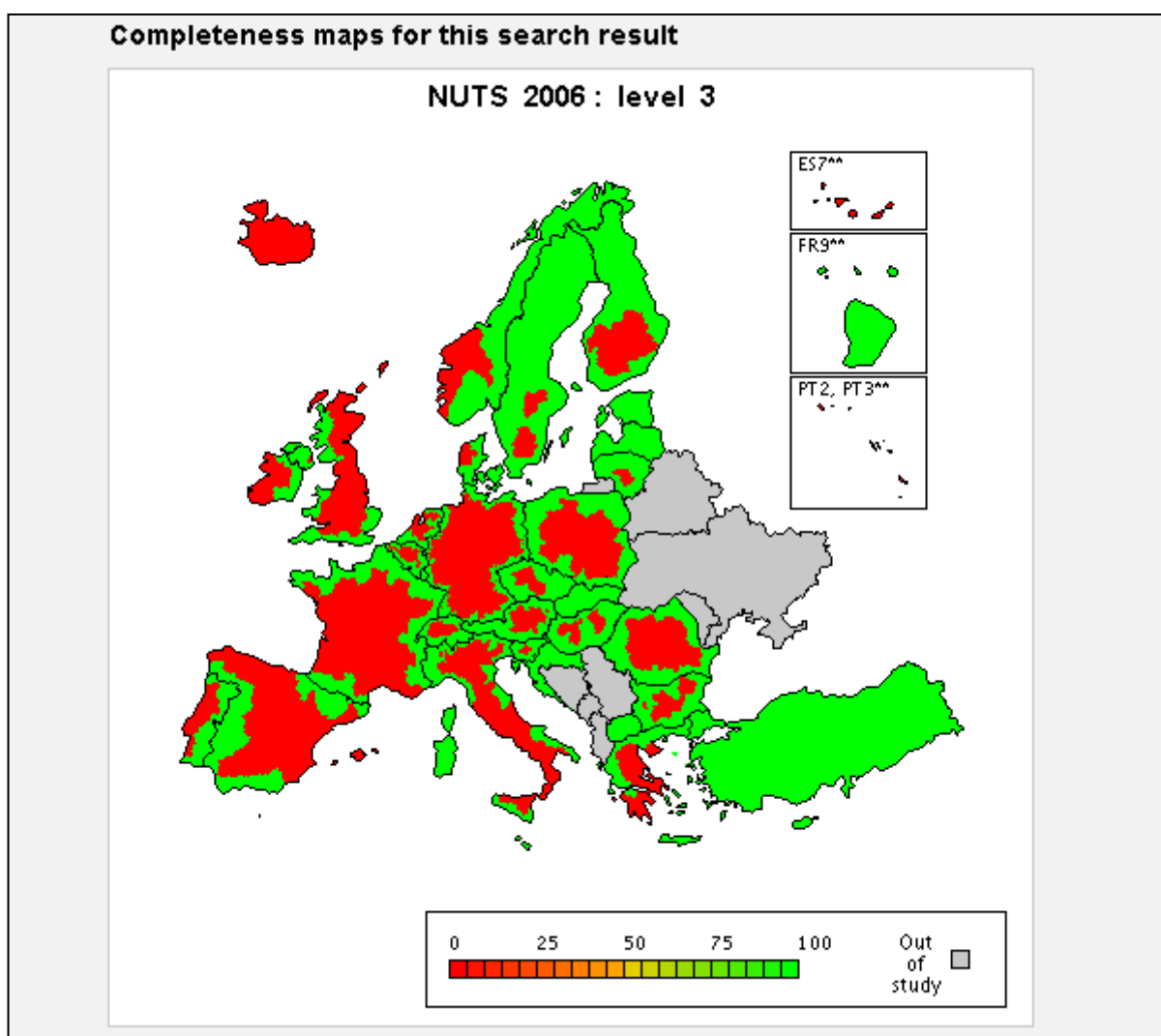
Indicator name:	<b>Border regions</b>		
Abstract:	Typology on border regions		
Code:	<b>bc_border</b>	Unit of measure:	none
Topic(s):	NO TOPICS ASSOCIATED WITH THIS INDICATOR		
Keyword(s):	NO KEYWORDS ASSOCIATED WITH THIS INDICATOR		
Methodology:	Description of the categories of the typology - 1 European integration regions: regions with an internal border, major border characteristics and a high density of border crossings - 2 Internal fringe regions: regions with an internal border, major border characteristics and a low density of border crossing or a maritime border - 3 EU/EFTA entrance regions: regional with an external border, major border characteristics and a high density of border crossings - 4 External fringe regions: regions with an external border, major border characteristics and a low density of border crossing or a maritime border - 0 Areas not covered by classification: regions with no borders to a foreign country... <a href="#">less</a>		
Indicator name:	<b>Costal regions</b>		
Indicator name:	<b>Island regions</b>		
Indicator name:	<b>Mountainous regions</b>		
Indicator name:	<b>Sparsey populated regions</b>		
Indicator name:	<b>Urban and metropolitan regions</b>		

Figure 13: The detailed metadata page.

Metadata: by clicking the icon, the application opens a popup window where the user can consult further details about the metadata of this current result (see **Figure 13**). Four different types of information are available in this metadata page:

- identification of the dataset;
- the list of indicators for this dataset;
- the data lineage (the data sources for all the values) of this dataset;
- the map of Europe showing the rate of completeness of data for each country which is concerned by the dataset. Figure 9 describes this feature.

A cartographic representation of the level of completeness of the data for the chosen dataset is also displayed in the metadata section. The completeness described the result as a whole. If more indicators or more years or more NUTS revisions have been chosen from the search criteria, then the displayed completeness is averaged for all these criteria. The data completeness can be seen at every available nomenclature level.



**Figure 14:** The data completeness metadata screen.

Finally, from this search result table, the user can perform the two following actions on selected items (by checking checkboxes on each line):

- adding selected results to his/her basket;
- downloading selected items.

The “basket” feature allows to temporary save search result items. Thus, the users can keep their attention on their queries, the basket allows them to drill their research process.

However, the basket is not a mandatory step for downloading results: if the user has found the expected results, they may directly use the “Direct download” functionality. This “direct download” can be considered as a short-cut to the basket service, which is further described in the following section.

## 1.6 Basket page

**Figure 15** shows the basket page: excepting the project name (which disappears) and the dataset name (which is renamed to “Item”), the displayed information is similar to the search results table on the search page: the basket is filled with search result items. Once the user completed his/her research, he/she can now refine the selection of search results items that he/she wants to download.

The screenshot shows a web interface for a basket of items. At the top, it says "Your basket : 3 item(s) collected." Below this is a table with the following columns: Item, Geo objects, Study areas, Years, Completeness, and Metadata. There are three rows of items, each with a checkbox on the left. Below the table are two buttons: "Download selected results" and "Remove selected".

<input type="checkbox"/>	Item	Geo objects	Study areas	Years	Completeness	Metadata
<input type="checkbox"/>	Typology compilation	NUTS and similar	(CC) (ESPON31)	2009 <input checked="" type="checkbox"/>	NUTS3 90%	
<input type="checkbox"/>	Lande use data	NUTS and similar	(EU27) more...	2000 <input checked="" type="checkbox"/>	86% Show by levels	
<input type="checkbox"/>	Lisbon strategy performance	NUTS and similar	(ESPON31)	2000 <input checked="" type="checkbox"/> 2006 <input checked="" type="checkbox"/>	NUTS2 81%	

**Figure 15:** The basket page.

Through the checkboxes on the left side of each line, the user is invited to select items on which he/she can perform both following actions:

- a deletion;
- a download.

The deletion consists in removing the selected items from the basket. Note that in such a case, the user is asked to confirm before processing.

The download action triggers the following set of tasks on the server-side: a Microsoft Excel file is built for each selected item which is expected to be downloaded. This spreadsheet is composed of at least four sheets:

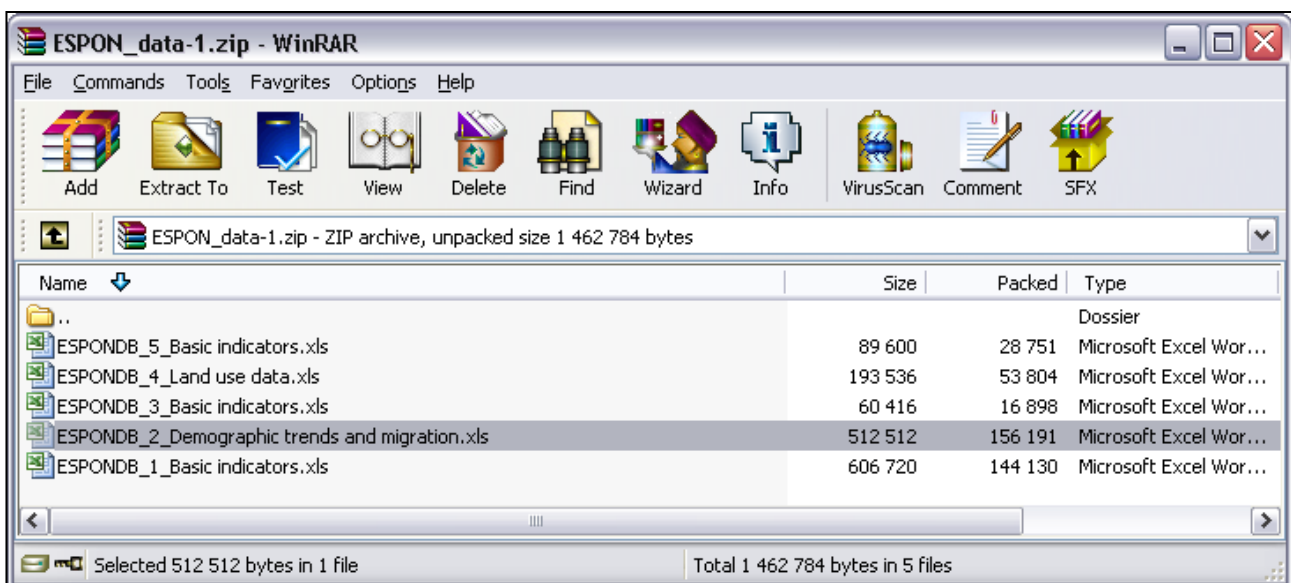
1. the first sheet is untitled "Dataset", it provides general information about the current dataset (name, contact, etc.);
2. the second sheet is untitled "Indicators", it provides metadata information about included indicators in this dataset: a description, the unit, the methodology, etc.;
3. the third sheet is untitled "Lineage", it provides metadata information about the lineage of the dataset: validity start, methodology, etc.
4. the following sheets display the values of the dataset, broken down by nomenclature revision and nomenclature level, there will be one sheet for each revision/level combination existing in the dataset (e.g. one sheet for NUTS0 2006, one for NUTS1 2006, one for NUTS0 2003, etc.).

If the user has selected several items to be downloaded, a zip archive is built, gathering the built xls files (one file per result item) by the previous step. The application finally returns the built file to the user as an attachment (an xls file or a zip archive, depending on the number of selected items).

On the client-side, the users are prompted by their browser to open or to save the built file on their disk.

In the case of one selected item, the proposed filename for the downloadable file will be **ESPON\_data.xls**. In the case of several selected items to be downloaded, the default filename for the proposed downloadable file will be **ESPON\_data.zip**.

Once downloaded on their local disk, the content of the zip archive can be extracted with a standard archiving tool, the included files are simply named **ESPON\_number of result\_dataset title.xls**, and there will be as many files as there are selected items in the search results or in the basket (see *Figure 16*).



**Figure 16:** An example of downloaded zip file with 5 search results.

Caution: note that the basket is emptied when the user logs out. This is the reason why the user is asked to confirm his/her wish when he/she clicks the “Log out” button on the menu bar.

## **1.7 Upload page (registered users only)**

The upload page aims at allowing ESPON project members to upload their data deliveries in order to fill the ESPON 2013 Database. The ESPON 2013 Database Application allows the users to upload 3 types of files to the ESPON server:

- metadata files;
- data files;
- additional documentation files (technical reports, etc., describing in more detail the different methodologies used for calculating indicators or for estimating missing data).

This is typically done in a 4 step process.

In the first step users are invited to upload a metadata file on the server. However, the step can be much more complex. When loading a metadata file within the application, the users are prompted the existing errors in their document. Errors can be either syntactical errors, or simply missing information for mandatory fields. Fields with errors are highlighted in red and a textual description of the error is also given in the header of the metadata editor (see an example in *Figure 17*).

Search Basket Log off Upload Account Administration Terms&Conditions

**Upload your data to the server: step 1 of 4**

The mandatory field 'Abstract' of 'Dataset' section is empty.

File templates  
 Metadata file template Data file template

Metadata upload (required)

Dataset	Contact	Indicator	Value
Contact 1 of 1			
From my profile? <input type="checkbox"/>			
Name?	Johanna Roto		
Organization?	Nordregio		
Function?	Data integrator		
E-mail?	johanna.rote@nordregio.se		
Phone?			
Role?	Originator		

Summary Load XML/XLS Save as XML Save as XLS

Go to next step Reset the form

**Figure 17:** The upload page, step 1.

The metadata profile of the ESPON 2013 Database has been previously defined as an Excel template. The Web editor contains largely the same fields as the Excel template, but allows filling them in a more user friendly manner. The interface defines the same three levels for filling the metadata: description of the dataset, of the indicator and of the record (for lineage). It includes all the information required for compliance with the INSPIRE directives. However, in order to simplify the task of the data providers, the application fills in automatically all the data that can be derived from the dataset. For instance, the data providers are not required to describe the spatial or the temporal extent of their dataset, this information will be derived by the ESPON 2013 Database Application and filled in automatically.

The first level defines the main characteristics of the whole dataset. It includes all the common points for all the elements inside the dataset and, also, the metadata description, like the name of the delivery, the authors, etc.

Information	Description
Dataset name	Name of the data delivery
Abstract	Short description of the dataset content
Date	Date of creation of data and metadata
Contact Point	Name, email, etc of the author of this metadata

The second level of metadata information is based on the characteristics of each socio-economic indicator. Inside a dataset, each indicator has its particular specifications and characteristics. The selection of appropriate themes and keywords is a key point, at this level, to obtain a good description of the resource.

In order to ease the edition of indicators and to harmonize definitions and codes inside the database, the editor provides a list of well known indicators (that are already present in the database).

Information	Description
Code	Code of the indicator (harmonized with existing nomenclature)
Name	Name of the indicator
Abstract	Full text description of the indicator
Unit of measure	Can be meter / square, thousands, etc: free text
Classification (themes, keywords)	Various themes and keywords found into thesauri to describe the resource. Multiple themes and keywords can be attached to one indicator, if needed.

The value level is based on the characteristics and description of the indicator values and it is linked with the "category column" of the data files. This is the most specific level because it defines the main characteristics of the value of a specific indicator that belongs to one dataset. The lineage and reliability contribute to give information about the quality, the procedures for data collection, the sources and the methods used for delivering and transforming data.

The information for each value is grouped by "scopes". A scope defines a subset of data which have the same lineage. For example, if a dataset for the EU 27 contains official data coming from Eurostat, from the Romanian and from the Bulgarian national statistical institutes, the values metadata will be grouped in three scopes, one for EU 25 values (source EUROSTAT), one for Ro values (source: Romanian national statistical institute) and one for Bg values (source: Bulgarian national statistical institute).

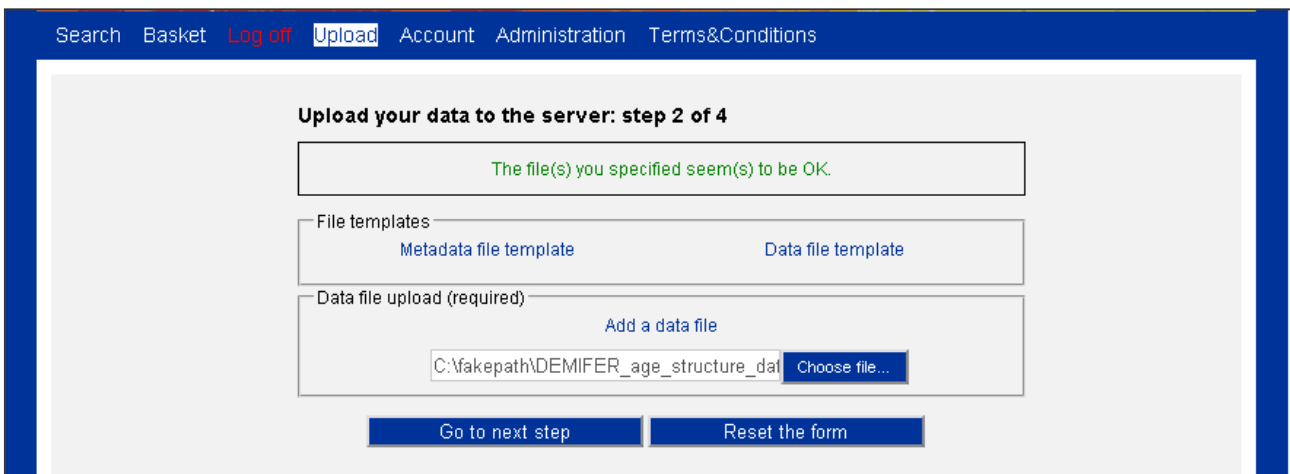
Information	Description
<b>Scope</b>	
Label	Label in front of each value in dataset, that allows to map metadata information with data information
<b>Lineage</b>	
Provider, provider URL	The statistical institute or ESPON project that provides the values and a link to its website
Date	The date of extraction of the data
Source , source URL	The document name (report or article) and its URL, if available on the Web
Methodology	A free text description of the methodology used for estimating, calculating or correcting the data in this scope This is required whenever modifications were applied by the provider to the original data, for instance when data are complex indicators that were built from less complex indicators, when data are result of corrections or estimations of the original data, etc.
<b>Reliability</b>	
Estimation	<b>true</b> if the value comes from an estimation
Quality	A level of confidence to the reliability of the value: <b>low, medium, high.</b>
<b>Constraints</b>	
Public data access	<b>true:</b> the data is public which means all users, without any restriction, can view and download the data <b>false:</b> means that only members of ESPON projects will



		be able to download those datasets
	Public metadata access	If true, users can see what exists in the database, whatever the data access's value.  If false, nobody can see the data stored inside the database. It is a special issue for purchased data that can only be used on the demand of an expert.
	Copyrights	A text area that allows users to communicate who owns the rights for the data within this scope

The users have a lot of flexibility in editing and storing their metadata files. They have the possibility to edit their metadata online, via an interactive process, or to simply edit it offline (in Excel, for instance), and then load it in the metadata editor. They can save it either in Excel or XML format and reload it at a later time. However, an important point to note is that metadata file upload cannot be performed if the user has not corrected all the errors and filled in all the mandatory fields. Once the file is error free, clicking on the "Go to the next step" brings them to the second step.

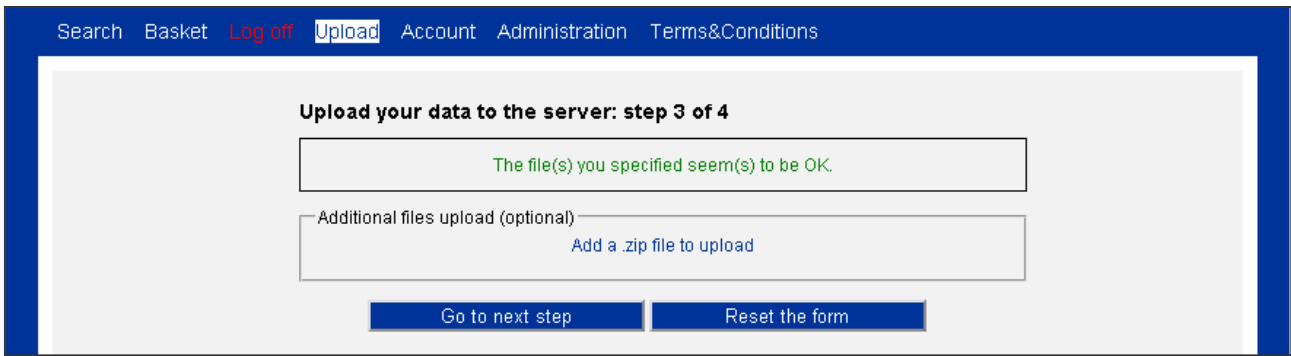
In the second step, the user is required to upload a data file to the ESPON server (see **Figure 18**). The data file must correspond to the previously uploaded metadata file, that is, all the indicators and scopes present in the data must have been described in the metadata file.



**Figure 18:** The upload page, step 2.

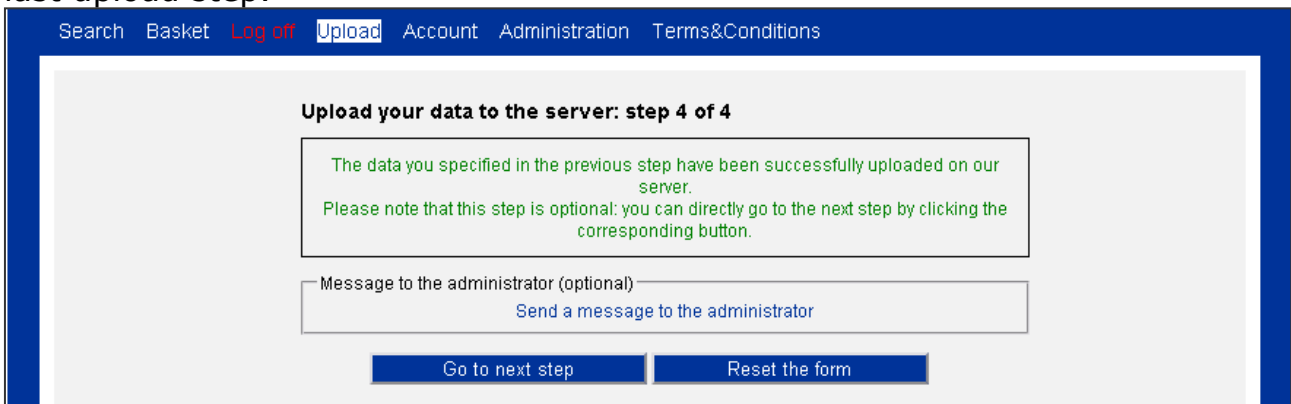
Note that the uploaded files are obviously not directly integrated to the ESPON Database, this upload step has to be followed by several processes: verification, harmonization, etc.

If no correspondence errors are detected, clicking on the "Go to the next step" brings the users to the third upload step.



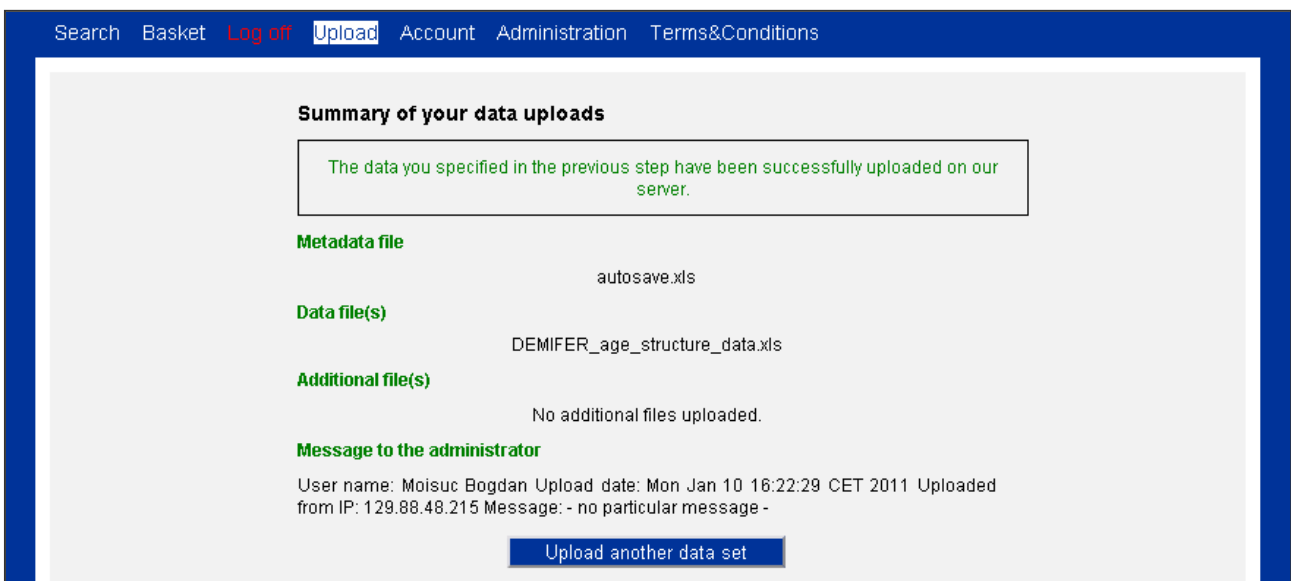
**Figure 19:** The upload page, step 3.

In this step one or more reports can be inserted, including intermediate data in Excel sheets in order to illustrate more clearly the methodology. All the files must be archived together in one .zip file. This step is optional, so, if there are no additional documentation files, the user can go directly to the fourth and last upload step.



**Figure 20:** The upload page, step 4.

In this step, users can send a message to the application administrator. If an error was committed, the upload process can be restarted by clicking the button "Reset this form".



**Figure 21:** Successful upload message.

In this step, users can send a message to the application administrator. If an error was committed, the upload process can be restarted by clicking the button "Reset this form".

## **1.8 Administration page (administrators only)**

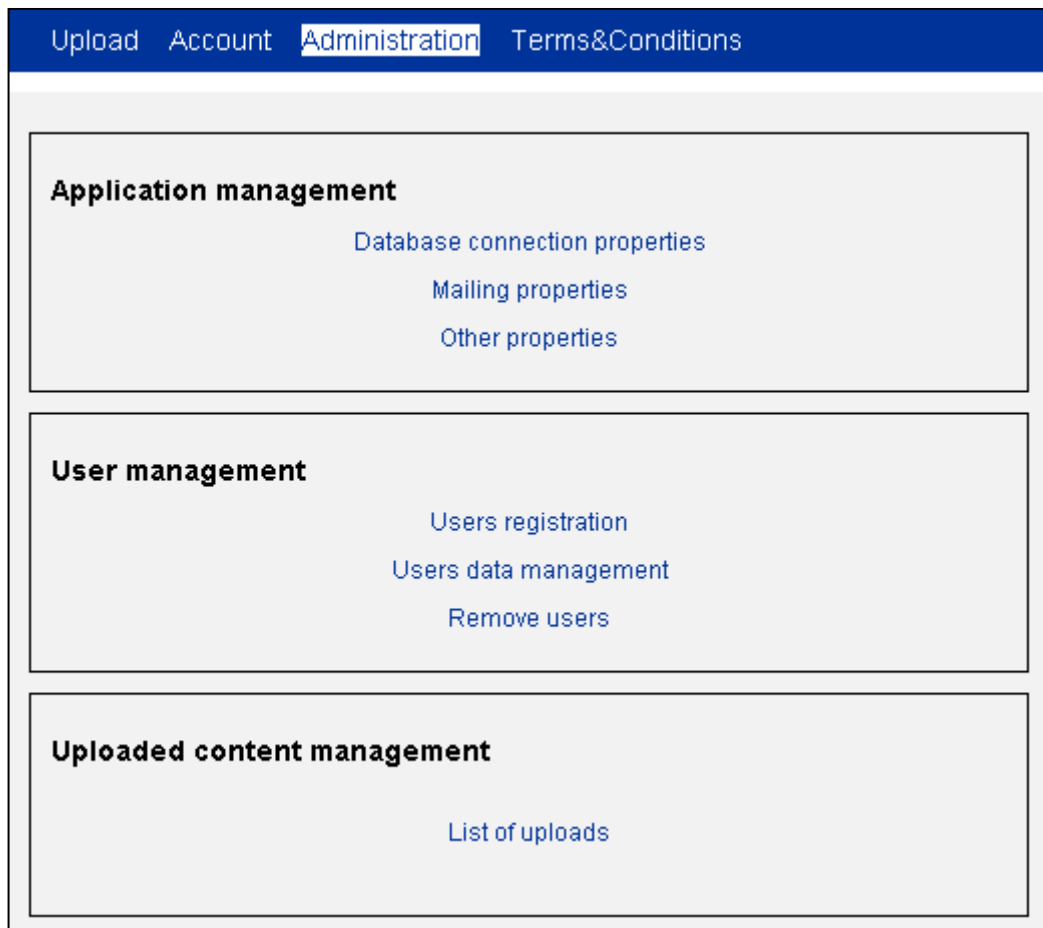
The administration page is aimed at helping ESPON Application administrators with the database maintenance and update tasks. There are three main categories of tasks in the administration interface: application management, users management and upload (data) management.

Application management tasks are related to the well-functioning of the application as a software. This comprises 3 categories:

- database connection properties, containing all the descriptors necessary for the application in order to have physical access to the ESPON Database;
- mailing properties, containing all the notification properties for the application (who gets notified when new uploads occur or when new users register);
- other properties include the formats and the templates for data and metadata.

User management tasks are related to the maintenance a list of application users with their respective privileges. They are also divided into 3 categories:

- users registration management allows administrators to have a fast overview of new registration requests and to validate or invalidate them;
- users data management allows updating information about registered users in a quick manner;
- user removal, for users who are no longer active members of the ESPON Database community.



**Figure 22:** *The administration page.*

The third type of tasks is aimed at dealing with new data uploads performed by ESPON Database data providers. The interface allows administrators to see what new uploads exist and to retrieve the files in order to perform data checking and acquisition into the ESPON Database.

## 2 The Databases

The ESPON DB Application uses two databases for the long term storage of statistical data. The separation is done in order to obtain an application optimized for two different (and conflicting) purposes:

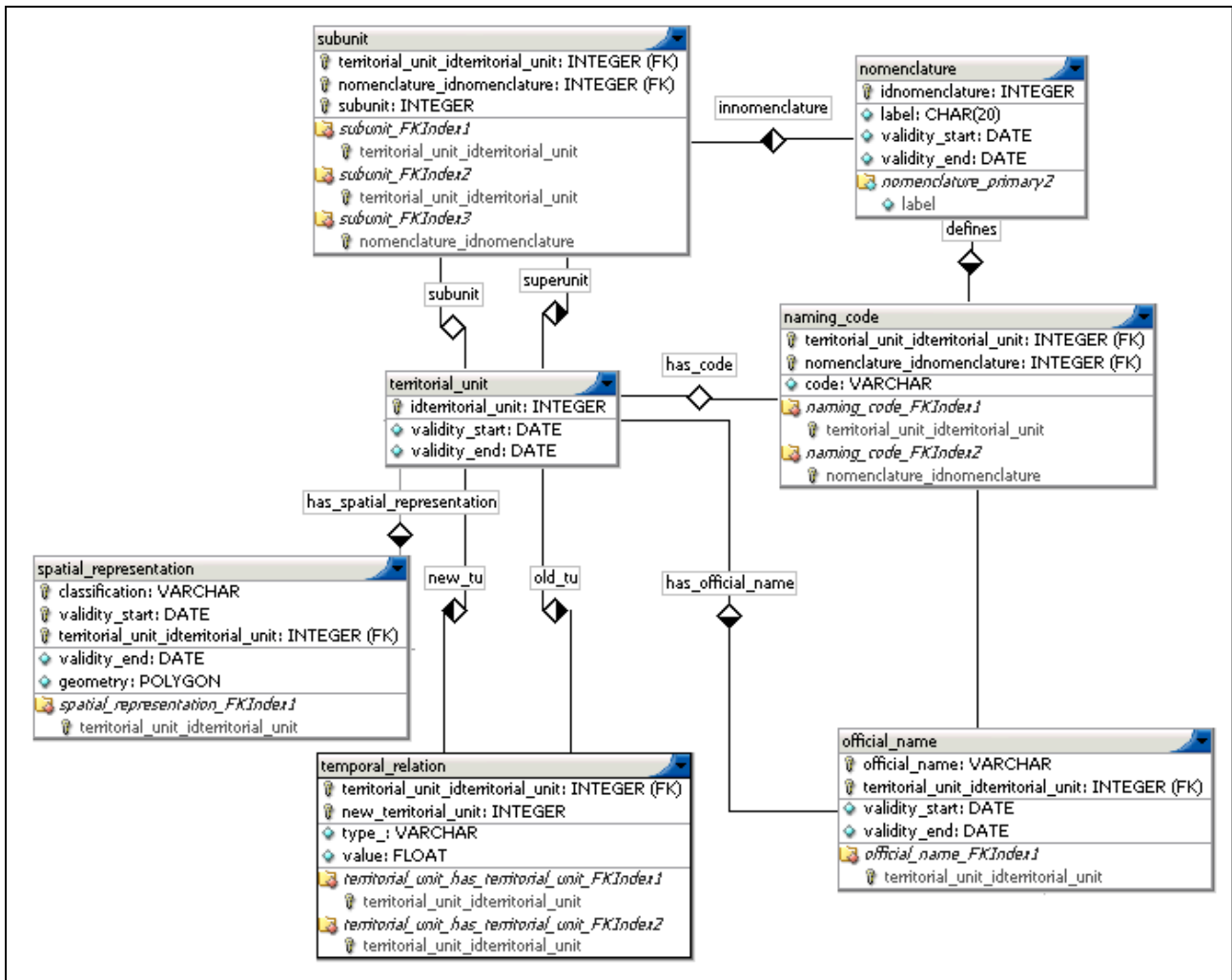
- The ontology database is based on a conceptual schema optimized for data harmonization. This conceptual schema imposes more separation between entities, and separation implies more effort at query time (thus, query processing performance is decreased).
- The ESPON DB is based on a snapshot schema optimized for query performance in the Web interface. The data are structured in such a way that fast query answer is privileged.

The ESPON DB Application also integrates a standalone Java application that allows inserting the content of paired data and metadata files into the ontology database.

### 2.1 The ontology database

Due to the size of the ontology database model, we split its presentation into several parts. Each of these parts is equivalent to one dimension of the data. From a conceptual point of view, the ontology database can be viewed as a hypercube with 5 dimensions to any statistical data:

1. The spatial dimension identifies precisely the spatial unit described by the indicator value;
2. The thematic dimension identifies precisely what type of statistical indicator is described by a given indicator value;
3. The lineage dimension identifies precisely the source of the indicator value (the database or the organization that published the value), as well as the transformations (corrections, estimations, etc.) that were applied to the value until its actual state;
4. The temporal dimension defines the moment in time or the period described by the indicator value;
5. The dataset dimension describes the publication (name, author, etc.) of which the indicator value is part of.



**Figure 23:** The spatial dimension in the ontology base.

The spatial description of the data is given via a spatiotemporal ontology. This ontology describes a complete list of the administrative units, stored in the `territorial_unit` table.

Territorial units are described as parts of a nomenclature. A nomenclature (see table `nomenclature` in **Figure 23**) defines two main properties for territorial units: their code, which is an abstract, standard and unambiguous name based on conventions (see table `naming_code` in **Figure 23**), and their hierarchical inclusion relations (see table `subunit` in **Figure 23**). Nomenclatures can evolve in time (see attributes `validity_start` and `validity_end` in the table `nomenclature`), and, as such, units codes can also change.

Territorial units also have official names that can change in time (see table `official_name` in **Figure 23**).

The temporal relations between territorial units are depicted via the table `temporal_relation`, which contains an attribute (see `type_` attribute in **Figure 23**) that allows describing the specific type of relation linking two territorial units.

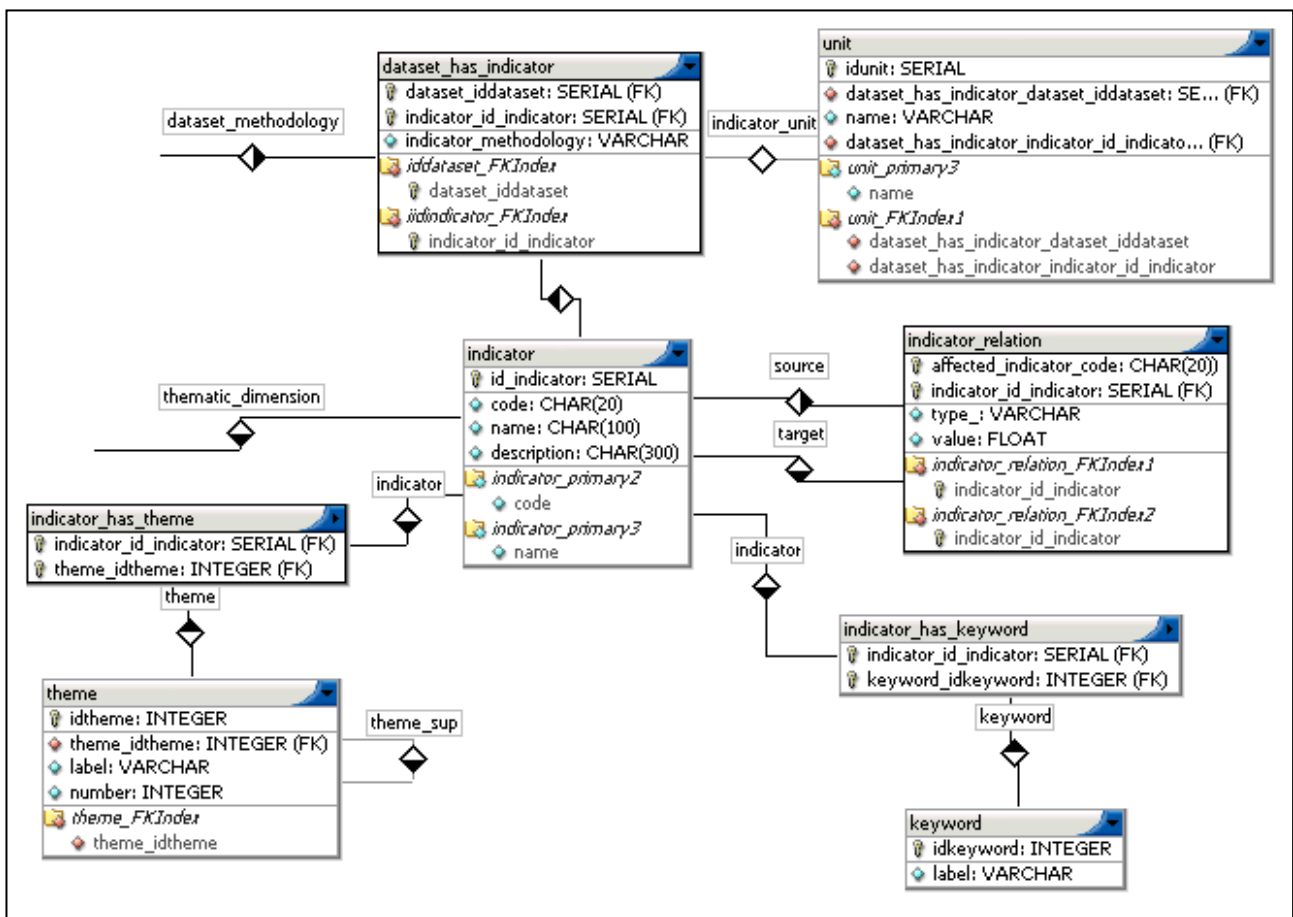
Each territorial unit can also have several spatial representations (see table `spatial_representation` in **Figure 23**), different geometries being fit for different purposes (e.g. more detailed geometries for applying geometrical operations and simplified geometries for display on screen). The spatial representation of a territorial unit can also evolve in time.

The thematic description of the indicator values is given via a thematic ontology. The table *theme* stores the ontology of themes and subthemes used in order to class the indicators in the ESPON Database (see **Figure 24**). Each indicator must belong to one or more themes and subthemes.

Keywords (see table *keyword* in **Figure 24**) are used in order to add more thematic information to indicators. Although keywords are not used by the current version of the ESPON 2013 Database Application, future versions may use keywords in order to make faster and more precise querying possible

The table *units* is required in order to keep track of the different measure units for the same indicator. This covers conversions between different measure systems and also multiples. For instance, the same indicator, total population, can be expressed in one dataset in thousands of inhabitants and, in another dataset, in inhabitants.

The table *indicator\_relation* is a generic table (can be used for a specific type of relation by changing the type attribute) designed to store vertical and horizontal relations between indicators.



**Figure 24:** The thematic dimension in the ontology base.

The description of the lineage of an indicator value is given in the *lineage* table (see **Figure 25**). The lineage of an indicator value is described by three elements: the original data provider (table *provider*), the methodology for the transformations applied to the data (table *methodology*), and the copyright constraints attached to the indicator value (table *constraints*).

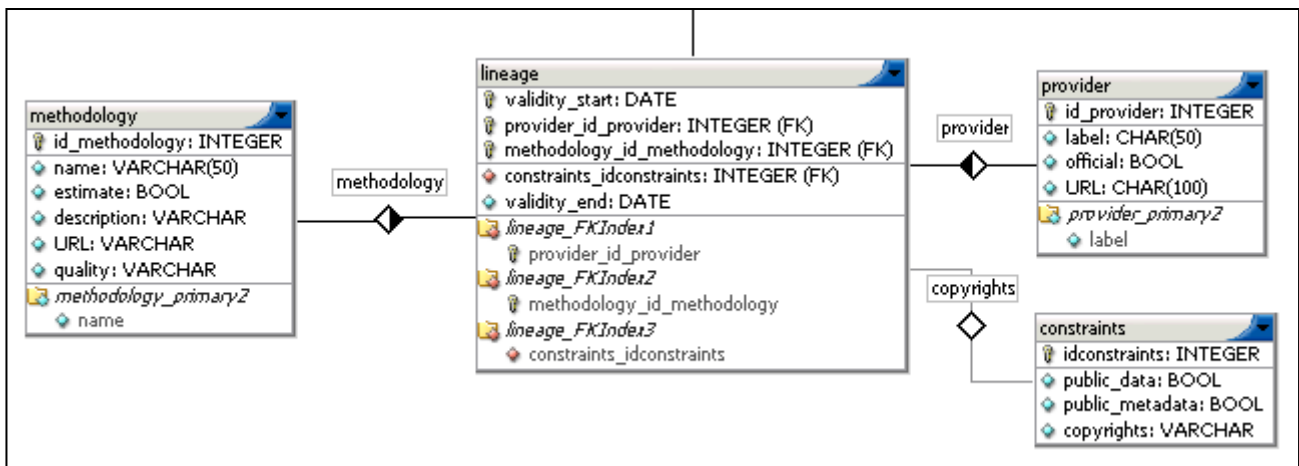


Figure 25: The lineage dimension in the ontology base.

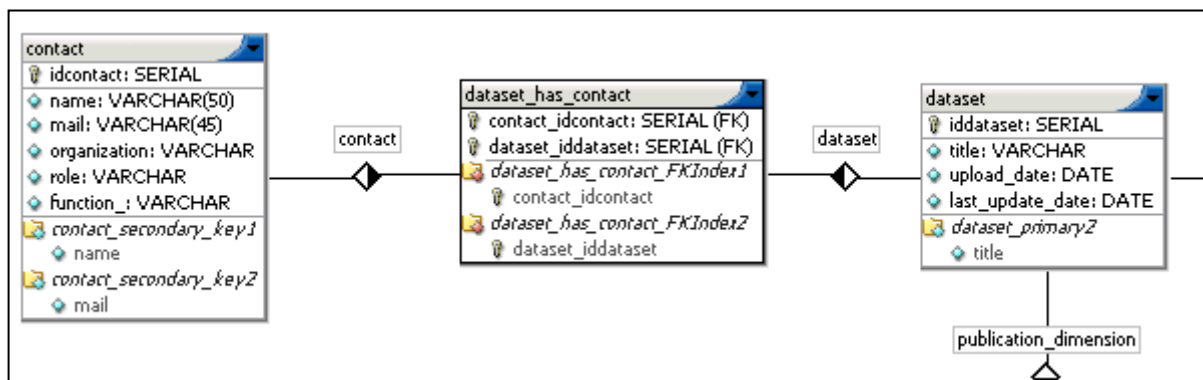


Figure 26: The publication dimension in the ontology base.

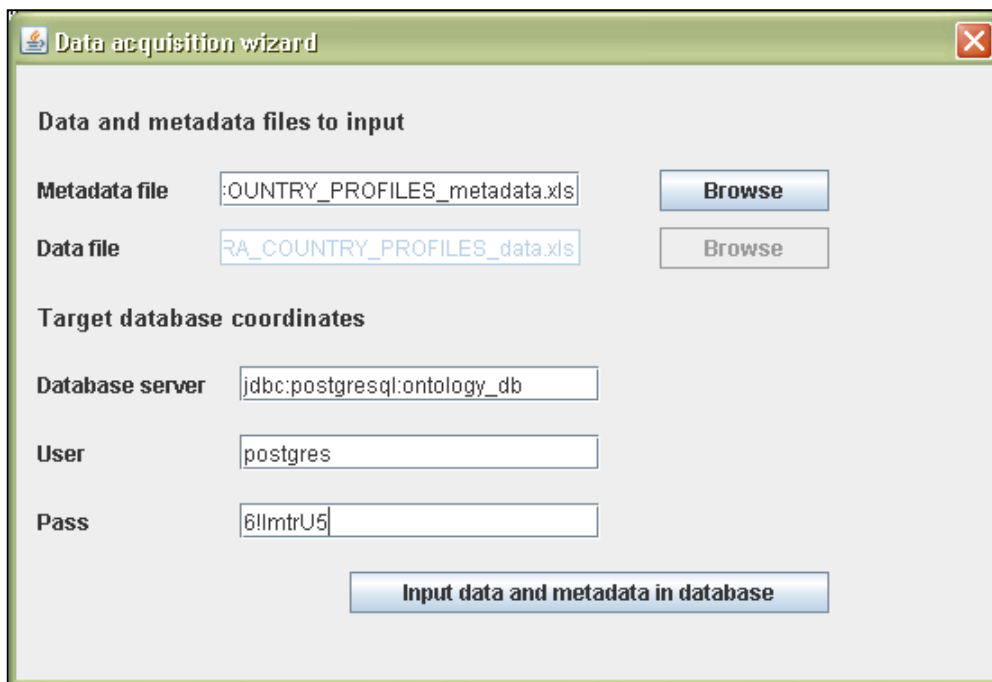
A publication is described in the ontology database through the table *dataset* (see **Figure 26**). Another type of information about the dataset maintained in the ontology database is about the person or persons who submitted the publication (see table *contact*). For one dataset, there can be more than one author, and the same author can work on several datasets. Besides keeping a trace of data deliveries from ESPON projects, the coordinates of the contact person from projects are kept with a double purpose, in order to help with the fast filling of the metadata in the Web metadata editor (where, through a connection to the ontology database a list of contacts will be displayed) and in order to manage user access to the ESPON database (as most of the data providers will be members of ESPON projects that will also need privileged access to the ESPON database).

In order to fill the ontology base with data from the data and metadata files uploaded by data providers, a custom data acquisition software has been developed. The main concept behind this software is that of mapping. A generic, open source software package allows describing mappings between documents and databases. This allows taking values described in text documents and inserting them in a database schema (here, in the ontology database), at the right place. However, before actually inserting the data in the database, a preformatting of the excel file is needed (this is done directly in the Excel files, via *Visual Basic* scripts).

After preformatting, metadata and data files are input in the Java desktop application and checked for errors. There are three categories of errors to be corrected: metadata errors, data errors and data/metadata correspondence



errors (e.g. when indicators described in the metadata are not present in the data or *viceversa*). The errors are pointed out to the application administrator, who must correct them before entering the data and metadata in the ontology database.



**Figure 27:** The data acquisition application.

Once data are in the ontology base, they are ready to be processed and enriched via automatic harmonization processes. Although automatic harmonization methods were not developed yet, in future phases of the ESPON 2013 Database Application they are expected to be.

After the harmonization process, the data and metadata are ready to be transferred to the ESPON Database, in order to be queried and downloaded by the final users. The transfer between the ontology database and the ESPON Database is done via an SQL script, which is used in order to make the conversion between the two schemas.

## 2.2 The Espon database

The ESPON 2013 Database uses a complex data structure representation that is based on the principles of reflecting the main features of the ESPON metadata scheme and introducing some conceptual entities that allow to optimize data access in a Web application context.

The present paper describes the core elements of the ESPON 2013 database and introduces the algorithm that is used to search statistical data in the database and to produce the information about the completeness of the existing data.

### **2.2.1 Core data structures in ESPON 2013 Database**

Most of the elements of the essential ESPON metadata structure are reflected directly in the ESPON 2013 database. For example, a dataset, the main element of the metadata structure, is represented in the database as an entry in the table called DATASET. The columns of this table reference the mandatory attributes of a dataset in terms of the metadata: the name of the dataset, its creation (upload) date and its abstract. Some additional attributes have been also created: each dataset table entry has a serial ID that facilitates search processing, and a short version of the name, that is used in the Web application user interface to harmonize the view aspect of the search form items. This principle of almost direct reflection of the metadata structure concerns multiple entities used to describe the ESPON statistical data: contacts, indicators, thematic classification and keywords, lineages and other elements are represented using this principle.

Another side of the ESPON 2013 database scheme is the representation of territorial units used for statistics. They are not described by the ESPON metadata, but are directly taken from the official documentation on territorial units for statistics. The datasets already processed for use in the Web application reference different versions of the NUTS nomenclature. So, the part of the ESPON database describing territorial (geographic) units contains a library of nomenclature and geographic units data for the NUTS of versions 1995, 1999, 2003 and 2006. This library was created during the database building process and is based on official documents fully covering the corresponding nomenclature versions.

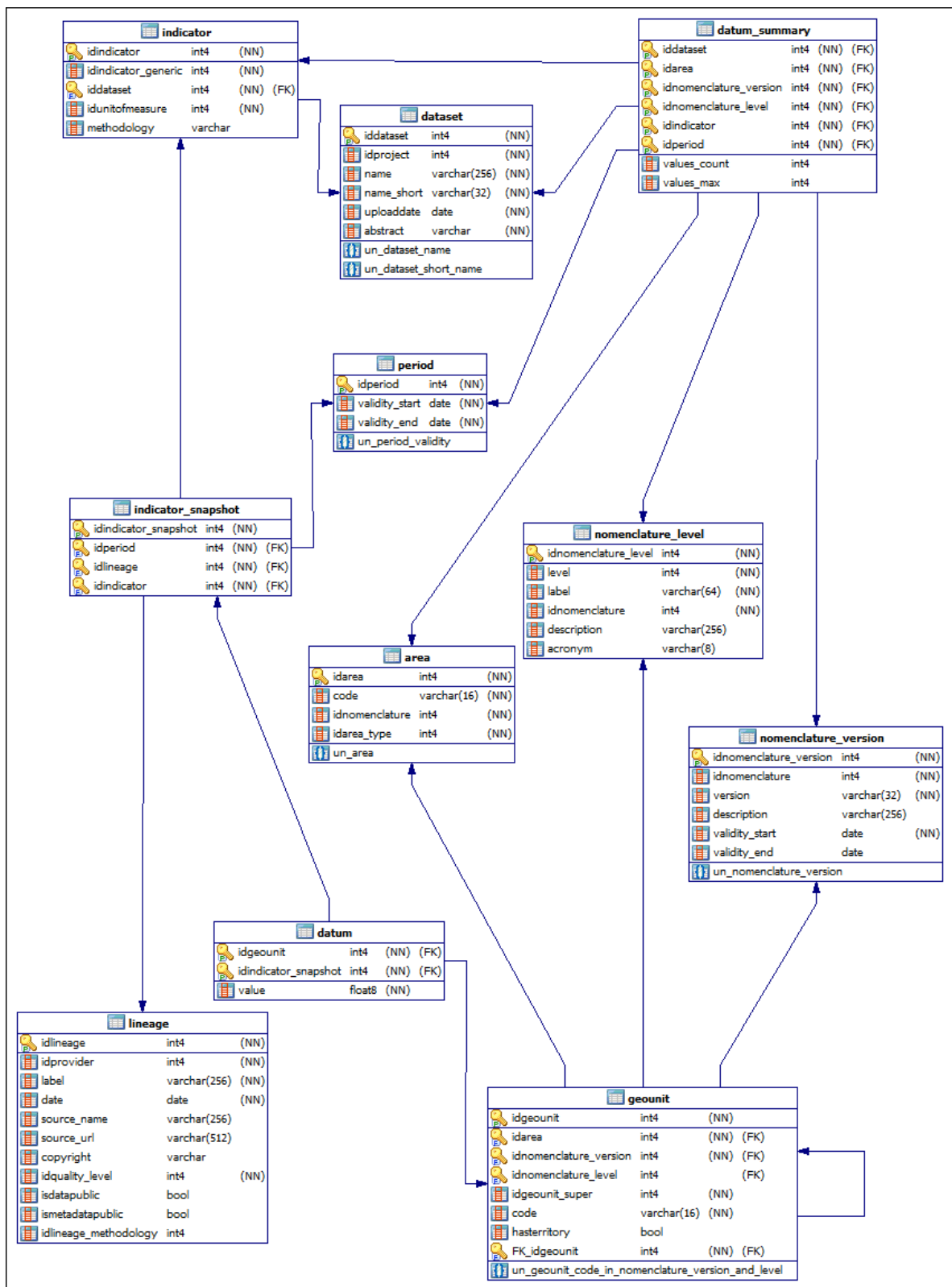
The ESPON statistical data represent a link between metadata information and a territorial unit. For example, the total population (an ESPON indicator) is measured for different territorial units. The link between them is an indicator value stored in the table called DATUM. In the context of the ESPON metadata structure, this association becomes complex because ESPON statistical indicators have two dimensions. The first one is the period of validity of the indicator value. The total population indicator changes in time, it varies according to the demographic, social and economic trends characterizing the region to which belongs the territorial unit. This means that the database scheme must take into account the fact that there may exist as many associations between the indicator and the territorial unit, as there are time periods for which the value was calculated. The second dimension is the source reference for the indicator value, called lineage. The same indicator for the same period of validity may be calculated by different statistical organizations or research actors. Each indicator value must be associated with its source of production in order to be valid.

The amount of statistical data produced by ESPON projects is very important: the smallest datasets already processed contain thousands of indicator values; the hugest ones can contain hundreds of indicators, coupled with periods and lineages references they give hundreds of thousands of values registered. These quantities of data are difficult to store and manipulate in a complex table, so, while designing the database structure, we decided to make the DATUM table as simple as possible. To obtain a simplified data structure, we

introduced the notion of the indicator snapshot, which associates an indicator with its time and source dimensions. These snapshots are stored in the table called `INDICATOR_SNAPSHOT`. Thus, the `DATUM` table can contain no more than three columns: a territorial unit reference, an indicator snapshot reference and a numeric value corresponding to the association between them. This approach facilitates the data search and significantly optimizes the access to the database table.

Even if the main statistical data table can be considered as simplified and optimized for access, it remains very inefficient for data search when general criteria are applied, because of the quantities of data present in it. In fact, all the search requests that can be made through the Web application interface concern higher-level entities and never descend to indicator values level or to small territorial units: the search can be made by dataset, theme, country, nomenclature level or version and covered period. This important detail gave us the idea to design a particular table in the database, that contains a generalized view of the low-level data registered. That table was called `DATUM_SUMMARY`; it groups the existing statistical values by the main search criteria of the application and contains about ten times less entries than the `DATUM` table. It does not provide the statistical values, but the information about the values that exist. Thus, it contains six foreign-key columns that reference the dataset, the geographic area, the nomenclature version and level, the indicator and the period. Two other columns contain the number of indicator values existing in the database for the combination of the foreign keys and the maximum number of these values possible for the area and the level referenced by the table entry. The `DATUM_SUMMARY` table allows to process search requests without accessing the `DATUM` table and to extract all the necessary information to produce search results. This table ensures the rapidity of search in the Web application context. It is also necessary to add that `DATUM_SUMMARY` table is maintained automatically, without any interference from the application administrator's side: the entries of this table are calculated and stored by internal database requests when the application is executing its initialization sequence.

The following diagram presents the relationships between the core tables of the ESPON 2013 database.



### 2.2.2 Data search algorithm

The search algorithm used in the ESPON 2013 Web application consists of two phases.

During the first phase, the search is done using the simple search criteria of the search form (search by dataset, theme or indicators). This search is done independently of the type of search requested by the user. If he requests a simple search, the result of this phase is already the final search result that will appear in the search results table. If the user requests an advanced search, the simple search algorithm is executed with the aim of restricting the amount of data to search in, because it helps to optimize the advanced search algorithm and to make it more rapid.

The design of the application takes into account that the objects retrievable from the database are relatively constant, they do not evolve after being placed into the database unless corrections are made by the application maintainers in the data. That is why we found that one of the ways to improve the application performance is to define a set of objects that are cached in the application memory and that can be directly accessed without the necessity to query the database. This collection of objects is built upon the application start up or when the database connection parameters have changed. Objects like datasets, indicators, thematic data and some others are stored in this collection. The simple search algorithm finds the objects cached in this collection without querying the database and produces the search result. For example, if the user selects a project in the simple search form and submits a search request, the search algorithm simply collects all the datasets that correspond to the selected project and produces the search result. The same sequence is applied for the other simple criteria - topics and indicators selections.

The second phase of the search algorithm is executed only in case of an advanced search request. If the user has made at least one selection in the simple search form, the simple search algorithm is applied before this phase. If there were no selections in the simple search form, but only in the advanced form, an additional database request is made to retrieve candidate dataset and indicator identifiers that can restrict the field of the main advanced search request. That search for candidates is made on the DATUM\_SUMMARY without significant joins on other tables and is usually very quick. If the simple search made previously or the candidate search give at least one result, we may conclude that there is a high probability of finding data for the final result and we build a collection of preliminary results using the data retrieved. If both the results are empty, the advanced search algorithm does not continue and the application outputs an empty search results table.

The next step of the advanced search phase is to filter the candidate results according to the search criteria imposed by the user and to retrieve all the necessary information to fill the search results table. The dynamic SQL request in this case is more complex than in the candidates search step, it uses the DATUM\_SUMMARY table joined with other tables that contain data for building the results: NOMENCALTURE\_VERSION, PERIOD and DATASET. It is parametrized using the indicator values of the candidate search result and thus executes very quickly. If the response for this database request contains at least one row of data, that means that the information found corresponds to the search criteria specified by the user and can be transformed into a search result object. For each search result object, the algorithm defines the nomenclature versions and levels, the periods and the geographic areas found.

It also calculates the priority of the search result: how many criteria specified by the user were found for this object. The priority value is used to sort the entries in the search results table. Finally, after executing all the advanced search results, the remaining part of the algorithm calculates the completeness values of each result object.

The complete search algorithm used by the application can be summarized in the following pseudo-code listing:

```
Find the cached objects that correspond to the user's search
criteria.
```

```
Produce simple search results (collection CS).
```

```
If (it's a simple search)
    output the search results table with CS and exit the
algorithm.
```

```
If (CS is empty)
    Build the candidates search query and execute it to produce CS
with elements.
```

```
If (CS is still empty)
    Output an empty search results table and exit the algorithm.
```

```
Build the advanced search query and execute it with the data of
each element of CS (collection CA).
```

```
For each element in CA
    calculate the completeness of the data found.
```

```
Sort by priority the elements of CA, if any.
```

```
Output the search results table with the elements of CA, if any.
```

```
End of the algorithm.
```

### **2.2.3 Data completeness calculation**

In the ESPON 2013 Database Web application context, the completeness of data can be defined as the percentage of effectively present indicator values against the maximum number of values possible for the given indicators, time periods and study areas. A dataset has the 100% completeness if all its indicators for all time periods have a value associated with each territorial unit, on any level of nomenclature hierarchy. The completeness value is a constant for a dataset taken entirely, but it can vary significantly between dataset subsets of indicators, time periods and territorial units. The completeness does not show the quality of data; it means only how much data is really present in

the selected range of criteria against the maximum amount of data that could be present for this selection if all combinations of spatial, temporal and thematic criteria had a value for these criteria.

During the application development, we considered two ways of calculating the completeness value. The first one was to calculate the completeness on-the-fly, counting each value present and each value absent, then producing the percentage. This is a very exact, but at the same time a very expensive method of calculation.

The way that we finally chose to produce completeness values for search results is to use the DATUM\_SUMMARY table. The contents of this table is always up to date with the one of the DATUM table. Two columns of DATUM\_SUMMARY, values\_count and values\_max, contain information about the effective and maximum number of indicator values present for the given indicator, time period, study area, nomenclature version and level in the DATUM table.

The calculation of completeness values is made when search result objects are already produced. They contain information about the indicators, periods, study areas and nomenclature details found. A simple SQL query allows to summarize the number of existing values for each search result and the maximum number of values possible. This sum is calculated for each level of nomenclature separately: that allows to show the completeness details per levels in the search results table. The sum of the completeness values by levels give the general completeness of the search result. The percentage value is calculated using the very simple formula:

$$\text{Completeness} = \text{values\_count} * 100 / \text{values\_max}$$

The SQL query built to retrieve the completeness data may be illustrated by the following example:

```
SELECT idnomenclature_level, SUM(values_count), SUM(values_max)
FROM datum_summary
WHERE idarea IN ({list or area IDs})
AND idnomenclature_version IN ({list of nomenclature versions
IDs})
AND idnomenclature_level IN ({list of nomenclature levels IDs})
AND idindicator IN ({list of indicator IDs})
AND idperiod IN ({list of period IDs})
GROUP BY idnomenclature_level;
```

The result of this query is produced in a very short time and lets to calculate the exact values of search result data completeness without demanding additional algorithmic procedures.

## Conclusion and future works

The specific achievements for the current phase of the ESPON 2013 Database Project from an application point of view can be divided into three categories: metadata related achievements, data related achievements and interface related achievements.

From a metadata point of view, the outputs of the ESPON 2013 Database Project include:

- a metadata profile specifically dedicated to statistical territorial data, compliant with existing standards (ISO 19115, the INSPIRE directive);
- an Excel template allowing ESPON 2013 DB data providers to fill compliant metadata offline;
- a Web application allowing ESPON 2013 DB data providers to fill compliant metadata online, to import metadata files in XLS and XML formats, to correct syntactical errors and to export metadata, again, in either XLS or XML format. The Web application is completely integrated within the ESPON DB data flow and within the ESPON 2013 DB Web Application.

From a data point of view, the outputs of the ESPON 2013 Database Project include:

- two operational database schemas optimized for data harmonization and, respectively, for fast querying;
- a software suite allowing importing data from XLS files into the ontology database and transferring them to the ESPON database, after harmonization;
- a complete ontology of territorial changes for the NUTS nomenclature between 1995-2006;
- a coding scheme and a thematic ontology for indicators;
- an operational relational database filled with regional (NUTS) data supplied by ESPON Projects, complying with the ESPON DB format (over 1 million records);
- an operational file database filled with all the remaining data delivered by ESPON projects that do not fit in the ESPON database yet (local data, cities data, world data, environmental data, etc.).

From an interface point of view, the outputs of the ESPON 2013 Database Project include:

- an operational Web application, allowing to fulfill the complete ESPON DB data flow, from user registration, data upload to data search and download;
- a database administration interface, integrated in the same Web application, allowing administrators to reconfigure the application on the fly.



Future improvements to be brought to the ESPON 2013 Database Application can also be divided into the same three categories.

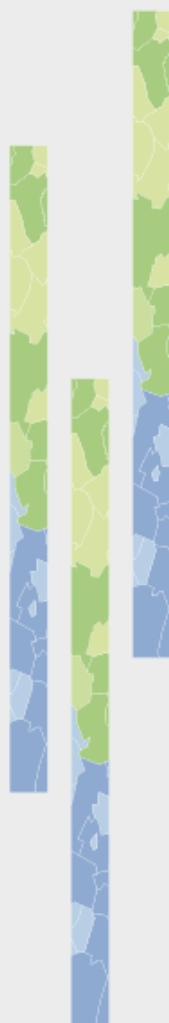
From a metadata point of view, in future phases of the ESPON 2013 Database Project the TPG will have to ensure that the metadata profile is coherent with the data formats used for including new types of data in the ESPON database, i.e. continuous spatial data (usually referred to as environmental data) and flow data (i.e. statistical information describing not one territorial unit, but a relation between two territorial units).

From a data point of view, future outputs of the ESPON 2013 Database Project should include:

- extending the ESPON DB schema in order to accommodate discrete flow data (i.e. occurring between territorial units, like regions or cities);
- designing and implementing some automatic data harmonization and estimation methods;
- designing and implementing a solution for storing spatially continuous data (environmental data) and for easy conversion of this data towards a discrete format (i.e., the actual model of the Espon database);
- building spatial ontologies for world territorial units and, as much as possible, for local units and cities (completeness is hard to achieve within this scope, due to constant changes);
- continuously filling the ESPON DB with spatially discrete statistical data (regional, local, world and cities) data, flow data and spatially continuous data.

From an interface point of view, future outputs of the ESPON 2013 Database Project should include:

- extending the ESPON DB Web Application with new types of search (i.e. keyword search);
- extending the ESPON DB Web Application with new search criteria (i.e. spatial typology criteria);
- making the search interface more flexible (i.e. allowing spatial selection of other types of units than countries).



## A two-step approach to structure the ESPON 2013 DB by themes and sub-themes

### MAIN RESULTS

- Database structures are used as source of information to determine, by means of visual grouping techniques, homogenous clusters of words
- We apply text mining methods to derive sub-themes. Here, ESPON evidence is used as textual data to extract potential keywords
- Two distinct parameters of data co-occurrence are employed to understand relational similarities
- We evaluate the explanatory power of words with an algorithm that weights the importance of each word in large corpora of textual data
- Mapping techniques of multidimensional scaling are used to depict similarities and ease interpretation
- Our results suggest that the ESPON 2013 DB should be structured in 7+1 themes and 27 sub-themes

**ESPON 2013 DATABASE**

**MARCH 2011**



## **LIST OF AUTHORS**

Nuno Madeira, University of Luxembourg

Geoffrey Caruso, University of Luxembourg

### **Contact**

E-mail: [nuno.madeira@uni.lu](mailto:nuno.madeira@uni.lu); Tel. +352 46 66 44 9691

E-mail: [geoffrey.caruso@uni.lu](mailto:geoffrey.caruso@uni.lu); Tel. +352 46 66 44 6625

## Table of contents

Introduction .....	4
Research background and methodology .....	4
Research background .....	4
Methodology .....	5
Matrix visualisation techniques for cluster analysis .....	7
Discussion of preliminary results.....	7
Towards a first set of themes .....	10
Text mining methods and visualisation tools .....	11
Methods .....	11
Results .....	13
Agriculture & Fisheries .....	13
Demography .....	14
Transport .....	16
Energy & Environment .....	17
Land Use .....	18
Social Affairs.....	19
Economy .....	19
Cross-Thematic and Non-Thematic Data.....	21
Conclusions and future work.....	21
References .....	23
Appendices.....	24

## 1. Introduction

The ESPON 2013 DB aims to improve the access to regional and spatial information. This process has been initiated by the previous ESPON Programme in order to increase the number of variables that may positively support the analysis of spatial structures and trends across European cities and regions (see, for instance, ESPON project 4.1.3).

The goal of this technical report is to structure the ESPON 2013 DB by themes. Besides, it complements the technical report entitled "*Towards an ESPON thesaurus? Some preliminary considerations for the thematic structuring of the ESPON database*" that seeks to derive themes and sub-themes from a corpus of textual data. Here, we argue that database structures, nomenclatures and taxonomies developed by other organisations should be taken into account when structuring the ESPON 2013 DB. The reason for this lies of the fact that many of these structures have established common themes that often aggregate similar data.

First, we focus themes and use that information to analyse similarities between the different database classifications. In addition, we employ matrix visualisation techniques to make the description more comprehensive.

The results should be used to further progress on the user interface prototype and hopefully constitute a robust basis to improve the performance of text mining methods (see previous technical report). Arguably, it is worth mentioning that methods employed in this report will only take into account statistical and geographical sources that have been used to develop indicators by ongoing ESPON projects. In other words, only indicators delivered up to mid-September 2010 will be considered in this analysis.

As a second step, we propose to link each indicator to a theme and sub-theme. Eventually, this process will facilitate harmonisation of codes – variable names – defined in an uncoordinated fashion by TPGs involved in ESPON projects. This is significant because it would offer some consistency to the entire database and assist other research projects when naming variables developed by ESPON to evaluate territorial trends, structures and policy impacts in Europe.

## 2. Research background and methodology

In this section we provide some research background and describe the methodology applied to determine themes that will eventually constitute the backbone structure of the ESPON 2013 DB.

### 2.1. Research background

As a first approach we assembled a list of first-level themes established by organisations from which ESPON projects normally obtain raw data, namely UNEP, EEA, EUROSTAT, OECD, UNESCO, WDI, and ILO<sup>1</sup>. This is meaningful because most of these database structures have provided and will continue to provide raw data both in terms of environmental and socio-economic topics to develop ESPON indicators and indices. With this regard, each word or expression used as a theme has been listed, evaluated in terms of similarity, and ultimately aggregated into similar themes. However, we must point out that the aggregation of words into thematic clusters has been purely inductive and based

---

<sup>1</sup> For more detailed information on each database classification, please visit the following Internet sites: UNESCO (<http://stats.uis.unesco.org>); ILO (<http://laborsta.ilo.org>); EUROSTAT (<http://epp.eurostat.ec.europa.eu>), OECD (<http://www.oecd.org/statsportal>), EEA (<http://themes.eea.europa.eu>), UNEP (<http://geodata.grid.unep.ch>), WDI (<http://ddp-ext.worldbank.org>).

on the semantic value of each theme. For detailed information, please see Appendix 4 and 5 to this report.

The dataset consists of 85 words or expressions taken from the 7 database structures. Each database has in average 18 first-level themes. Both UNEP and World Bank share the largest structure with 26 themes whereas UNESCO has structured its database with only 6 themes.

A prior step in this analysis is data preparation. The input data matrix is described by a binary (presence/absence) relationship model. That is, all values range between convergent (1) and divergent (0). Table 2.1 lists some of the words (rows) and database classifications (columns) employed in this analysis. If we take the first example, we would be able to understand that 'Tourism' is considered as a first-level theme by UNEP and EEA while other databases do not devote the same attention to such topic. On the other hand, 'Unemployment' has only been labelled as a first-level theme by ILO. This is reasonable due to the purposes of each database.

	UNEP	EEA	EUROSTAT	OECD	UNESCO	ILO	WPI
(...)							
Tourism	1	1	0	0	0	0	0
Trade	1	0	1	1	0	0	1
Transport	1	1	1	0	0	0	0
Unemployment	0	0	0	0	0	1	0
(...)							

**Table 2.1** Short example of data input for analysis

In order to understand the structure generated by this binary matrix some graphical techniques have been applied to determine clusters, identify blocks within the matrix and increase visual perception of commonly used themes. Following the well-known methods developed by Bertin (1967), we explore the concept of matrix visualization and cluster analysis offered by generalized association plots, or GAP (Chen, 2002; Wu et al., 2008). This open source tool can be understood as recordable matrix to communicate data structures and patterns. Basically it offers the possibility to visualise raw data and display tabular quantities and relationships by means of colour-based representation. The output of such experiment is displayed in a rather natural, inductive perspective but sufficiently helpful to identify proximities between subjects and variables.

The proximity measure one can employ to relate objects in such an experiment depends on the data type (i.e. binary, nominal, ordinal, etc). Within this context, the choice of proximity measure has an effect on the association patterns which directly influences the visual representation of the interaction structure (Wu et al., 2008). GAP offers some specific measurements for asymmetric information. As our matrix corresponds to a binary data type (presence or absence of a theme in a given classification) we have applied Jaccard's coefficient.

## 2.2. Methodology

The choice of each database derives from the fact that ESPON evidence is strongly based on raw data provided by those above mentioned institutions. As a consequence, it seems appropriate to consider each database classification and validate by means of generalized association plots the degree of similarity and dissimilarity. The usefulness of such approach is to harmonise words and/or expressions used by some of the most prominent statistical databases and therefore provide end-users a common language of understanding.

The matrix visualization is illustrated by a series of images that explore correlation between themes (subjects) and databases (variables). In order to capture potential differences among those databases we decided to include the ESPON 2006 DB structure of first-level

themes and identify specific features that could validate or refute our cluster analysis (see Appendix 1, 2, and 3). To this end, each exercise is illustrated by two matrices as an attempt to reveal potential difference. Clearly, some patterns can be discerned from those matrices. Next, we will explore and understand the structure embedded in each data matrix and determine a hierarchy of themes that could support the thematic structuring of the ESPON 2013 DB.

As a first step, we added to our correlation matrix the structure used by the previous ESPON database (ESPON, 2005). Surprisingly, some of the results indicate a weak correlation between ESPON 2006 DB and other structures. Even though, EUROSTAT has the strongest similarity whereas UNEP and UNESCO reveal less significant correlation coefficients. Somehow this reflects how crucial it would be for ESPON to be in accordance with main data providers.

In order to demonstrate the existing similarities between different classifications we employ a simple correlation analysis. The different goals defined for each database led to low correlation values. However, some interesting results emerged from this exercise. For instance, it is clear from the correlation matrix that EUROSTAT and OECD share the strongest correlation value (0.50). One reason that could be claimed to justify the degree of resemblance between these two structures is the nature of its content. Indeed, the fact that EUROSTAT and OECD collect and disseminate similar data for similar audiences has produced an impact on the classification of both databases.

The opposite scenario, i.e. weak correlation values, is rather frequent and little interpretation can be discerned. Still some assumptions must be done due to the degree of clearness, particularly among environmental databases. More precisely, the fact that those sources are committed to cover specific issues such as *environmental hazards, marine and coastal areas, or air pollution* (see Appendix 2 and 3) that often require detailed data also intensifies the number of discrepancies in most of the themes or categories adopted by each organisation. Preliminary results of this analysis are illustrated by Table 2.2.

	UNEP	EEA	EUROSTAT	OECD	UNESCO	ILO	WPI	ESPON 2006
UNEP	1.00							
EEA	0.22	1.00						
EUROSTAT	0.12	0.27	1.00					
OECD	-0.09	-0.12	0.50	1.00				
UNESCO	-0.10	-0.09	0.14	0.22	1.00			
ILO	-0.20	-0.11	-0.07	-0.10	-0.11	1.00		
WPI	-0.12	-0.15	0.22	0.24	-0.02	-0.07	1.00	
ESPON 2006	0.10	0.18	0.32	0.20	0.07	0.13	0.18	1.00

**Table 2.2** Correlation matrix of the database classifications employed in this experiment.

Interesting enough in this analysis is the fact that environmental databases tend to be more detailed when compared with socio-economic databases. To a certain extent, this ensures a high level of accuracy and promotes its utility for large audiences. However, there is an enormous discrepancy on the content provided by each database on environmental topics. On the contrary, both EUROSTAT and OECD have defined a broad list of categories to search and retrieve socio-economic data. As a consequence, semantic similarities are higher and the degree of resemblance between these two entities is much stronger.

Despite the purpose and content of each database it is obvious that organisations do not give much importance to labelling harmonisation. Given the role of the ESPON 2013 Programme for policy advice and development, the ESPON 2013 DB project constitutes a major opportunity to demonstrate the advantages in establishing a harmonised thematic structuring that could rely on classifications defined by the main data providers but also

taking into consideration the INSPIRE initiative for the creation of an European Spatial Data Infrastructure.

### **3. Matrix visualisation techniques for cluster analysis**

The methodology applied in this section is partly based on visual grouping techniques to extract words from prominent database structures. We assume that such methods disclose an important capacity to define standards that eventually can lead to improved harmonisation and coherence of spatial concepts and therefore offer the possibility to organise knowledge for information retrieval by end-users. Next we discuss the results obtained with GAP to determine clusters and identify blocks.

#### **3.1. Discussion of preliminary results**

The figures presented in Table 3.1 demonstrate a clustering of words. The ranking of words is defined by GAP and includes the structure used in the previous ESPON Programme. The ranking has no absolute meaning but the relative position of words is useful to interpret. GAP ranking is actually the result of a permutation of words so that words that share a similar pattern of presence/absence within the different classifications are positioned in neighbouring rows. Here, we used the single linkage algorithm to obtain the block structure of rows from the permutation. Appendices 1 and 2 to this report display some of the techniques to help identify blocks. Despite its value in terms of matrix visualisation we will give a primary focus to Appendix 1.

Our initial assumption is that GAP offers very helpful features to interpret data matrix association, patterns and ultimately behaviours. This helped to identify some of the key ideas underlying matrix visualisation needs, namely in terms of adopting a practical solution to display matrices. In fact, the main advantage of such tool corresponds to what Wilkinson & Friendly (2009) designated by *cluster heat maps*. The expression itself is very fortunate because it gives the idea of clusters by shading association. That is, data matrices structured by similarity and/or dissimilarity to facilitate analysis and interpretation.

In this section we report our results with GAP (Wu et al., 2008). The goal is to illustrate by means of correlation matrices relevant patterns that could easily be interpreted and communicated. More precisely, we propose to find relatively homogeneous clusters of themes. In order to enrich our analysis the number of citations by theme will also be taken into account. Then, we discuss the results from this experiment to propose a first set of themes. Ultimately, these results are compared and clusters interpreted with respect to the indicators delivered up to date (i.e. mid-September 2010).

The preliminary results have provided substantial information to comprehend our data collection. According to Appendix 1 it became clear that certain themes are very representative to the different databases while others are less visible. For instance, if we consider the bottom right hand corner of the first figure on Appendix 1 we observe that the correlation of certain themes (subjects) is very strong among the different database structures used in this experiment. Themes such as *Agriculture*, *Population*, *Transport* or *Energy* are exceptionally transversal and consequently among the most-cited categories established by certain database classifications. This is significant and somehow justifies the need for adopting such themes within the ESPON 2013 DB.

On the contrary, the second figure on Appendix 1 does not include any reference to the ESPON 2006 DB structure. This was intentional as explained above. Indeed, after computing data the association matrix has slightly changed its appearance. Some of the themes have gained more visibility while others expressed a reverse tendency. However, it should be stressed that the primary group of four themes identified in the previous matrix has been kept very alike (i.e. *Energy*, *Transport*, *Population*, and *Agriculture*). Similarly, we



have identified a less prominent group of themes, mostly clustered on environmental topics, but totally disconnected from the above mentioned cluster. Themes such as *Tourism, Land Use, Climate, Resources* or *Health* lose their importance if not included in the same matrix as ESPON 2006 DB.

Surprisingly enough in this experiment is the fact in both matrices the number of citations is fairly similar, respectively 25% and 28.6%. Two themes, however, react in a different way and demonstrate common behaviours. Both *Tourism* and *Land Use* assume different ranking positions when GAP is employed and somehow the percentage of citations reflects that situation. This is extremely relevant because it justifies the ranking of each theme. Ultimately, it confirms that *Tourism* and *Land Use*, two themes credited to the previous ESPON database, are not so important when considering the entire group of words or expressions used in this experiment. An opposite dynamic is observed with *Trade* and *Environment*. Both themes are cited as much as those observed in the first cluster but apparently emerge too disconnected from the structure if the ESPON 2006 DB is considered. Despite this situation, it is clear that such themes should be aggregated to the first set of themes for the ESPON 2013 DB. Besides, it would compensate some of the environmental-oriented themes identified previously (i.e. *Water, Climate, Consumption, Resources*).

The results summarized by Table 3.1 reveal as well other groups of themes that may require further attention. The main feature of the fourth group is related with the predominant focus on socio-economic issues. Themes such as *Finance, Development, Science, Infrastructure* or *Education* assume greater importance within this cluster. On one hand, this is essentially due to the ranking defined by GAP when grouping themes that intersect both OECD and EUROSTAT database structures. On the other, it justifies the fact that most of these themes are linked to economic, social and development-oriented data. Nevertheless, it is also clear from Table 3.1 that an independent subgroup emerges within this primary group of themes. Indeed, it seems that the choice of computing a correlation matrix without including the ESPON 2006 DB structure has produced some significant impacts on the permutation result, especially on the position of *Technology, Fisheries* and *Industry*. Our interpretation is that those themes are strongly linked with the classification adopted by EEA and the motivation for this behaviour seems to be related to the fact that ESPON has not been considered in one of those occasions.

From this point onwards the structure is much more balanced both in terms of ranking and number of citations. This means that little interpretation can be discerned if the ESPON 2006 DB structure is employed by one of the correlation matrices. Next, we argue that those less prominent themes should be included or grouped within bigger groups since most of them are often related to a specific theme. This process has been developed in a rather inductive way and merely based on the semantic value or weight attributed to each theme. That is, the meaning of a given word (or expression) will define its value or weight when compared with themes and therefore determine the level of closeness.

As stated initially, this section justifies the choice of aggregating some themes that otherwise would be completely disconnected from our analysis. Consequently, we should stress that this experiment has to a considerable extent been influenced by the level of semantic closeness to other major themes previously identified. Against this background, it seems obvious that an important set of less prominent terms (or expressions) should be treated as environmental-oriented issues. A strong argument to support this view is related to the fact that most of those themes derive from environmental database structures such as EEA or UNEP. Thus, it is not surprising that our aggregation method considered domains on *Biodiversity, Waste, Elevation, or Slopes* as traditional environmental issues. The same applies to socio-economic issues largely labelled as integrative components. For instance, we noticed that *Taxation, Market Regulation, Employment, Labour, or Wages* can be understood as basic socio-economic themes that characterize the diversity of data published and disseminated by OECD or ILO on their own websites.

**Table 3.1** GAP ranking of words or expressions used as a theme

Themes	GAP Ranking (including ESPON 2006)	GAP Ranking (excluding ESPON 2006)	Number of citations, including ESPON 2006 (%)	Number of citations, excluding ESPON 2006 (%)	Groups
Agriculture	1	1	62.5	57.1	
Population	2	2	75.0	71.4	
Transport	3	5	50.0	42.9	
Energy	4	6	50.0	42.9	(1)
Tourism	5	17	37.5	28.6	
Land use	6	19	37.5	28.6	
Climate	7	14	25.0	28.6	
Water	8	13	25.0	28.6	
Urban	9	15	25.0	28.6	
Consumption	10	16	25.0	28.6	
Resources	11	18	25.0	28.6	
Health	12	20	25.0	28.6	(2)
Trade	13	4	50.0	57.1	
Environment	14	3	62.5	71.4	(3)
Finance	15	11	37.5	42.9	
Development	16	22	37.5	28.6	
Social	17	10	50.0	42.9	
Regional	18	26	25.0	28.6	
Science	19	12	37.5	42.9	
Technology	20	9	62.5	57.1	
Fisheries	21	8	37.5	42.9	
Industry	22	7	37.5	42.9	
Communication	23	21	37.5	28.6	
Infrastructure	24	25	37.5	28.6	
Economy	25	24	25.0	28.6	
Education	26	23	37.5	42.9	(4)
Air	27	27	12.5	14.3	
Biodiversity	28	28	12.5	14.3	
Chemicals	29	29	12.5	14.3	
Coastals	30	31	12.5	14.3	
Waste	31	30	12.5	14.3	
Soil	32	32	12.5	14.3	
Seas	33	33	12.5	14.3	
Scenarios	34	34	12.5	14.3	
Pollution	35	35	12.5	14.3	
Noise	36	36	12.5	14.3	
Welfare	37	60	12.5	14.3	
Demography	38	61	12.5	14.3	
Taxation	39	62	12.5	14.3	
Services	40	63	12.5	14.3	
Productivity	41	64	12.5	14.3	
Patents	42	65	12.5	14.3	
Market regulation	43	66	12.5	14.3	
Globalisation	44	68	12.5	14.3	
Information	45	67	12.5	14.3	
Boundaries	46	49	12.5	14.3	
Vegetation	47	50	12.5	14.3	
Elevation	48	52	12.5	14.3	
Threatened (species)	49	51	12.5	14.3	
Slopes	50	53	12.5	14.3	
Fertilizer	51	57	12.5	14.3	
Food (supply)	52	59	12.5	14.3	
Pesticides	53	54	12.5	14.3	
Marine	54	55	12.5	14.3	
Land cover	55	56	12.5	14.3	
Hazards	56	58	12.5	14.3	
Employment	57	44	25.0	14.3	
Labour	58	48	37.5	28.6	
Household	59	39	37.5	28.6	
Wages	60	40	12.5	14.3	
Consumer price (indices)	61	42	12.5	14.3	
Unemployment	62	41	12.5	14.3	
Strikes & lockouts	63	43	12.5	14.3	
Occupational (injuries)	64	45	12.5	14.3	
International labour migration	65	46	12.5	14.3	
Hours of work	66	47	12.5	14.3	
Wealth	67	-	12.5	-	
Spatial typologies	68	-	12.5	-	
Research	69	-	12.5	-	
Public sector	70	-	12.5	-	
Culture	71	37	12.5	14.3	
Literacy	72	38	12.5	14.3	
Balance of payments	73	69	12.5	14.3	
Exchange rates & prices	74	70	12.5	14.3	
External debt	75	72	12.5	14.3	
Governance	76	73	12.5	14.3	(5)

For those terms (or expressions) where uncertainties arise we adopted a more pragmatic solution. Themes like *Globalisation*, *Governance*, or *Welfare* which may be interpreted as very general concepts with meanings that often gravitate between different subjects, we decided to analyse what type of data was being labelled as such. Here, we noticed that such themes have not been equally considered by the database structures employed in this experiment. Somehow, this explains the singularity and different purposes attached to each database.

### 3.2. Towards a first set of themes

The thematic structure of the ESPON 2013 DB should not be seen as a normative approach, but rather as a descriptive one. However, the choice of themes itself is very crucial for the success of the ESPON 2013 Programme because it offers the possibility to support policy developments which, in turn, can and will be used by different target groups who wish to promote policy documents, technical reports, or academic studies. Moreover, data publically available for retrieval can be used as a source for developing trends and scenarios.

This represents a significant gaining for policy development on European spatial planning but most likely is subject of criticism. Indeed, one could ask if this theme or that were emphasized more, or if an attempt was made to add one theme or another. We believe that our preliminary results should be seen as images of the future or as elements that correspond to the needs of a particular moment. We listed below a first set of themes to help end-users to understand the structure we propose for the ESPON 2013 DB. Taking into consideration the methodology applied in this experiment, we labelled themes as follows:

<p><b>01. Agriculture and Fisheries</b></p> <p><b>02. Demography</b></p> <p><b>03. Transport</b></p> <p><b>04. Energy and Environment</b></p> <p><b>05. Land Use</b></p> <p><b>06. Social Affairs</b></p> <p><b>07. Economy</b></p> <p><b>99. Cross-Thematic and Non-Thematic Data</b></p> <p><i>9901 Integrative indices, typologies and scenarios</i></p> <p><i>9999 Geographical objects</i></p>
--

**Table 3.2** Themes proposed to structure the ESPON 2013 DB

This list assembles themes used by some of the main data providers. Occasionally, the meaning of the word derives from similar terms or expressions. This has been the case for *Social Affairs* that often recalls societal-related issues that have great effects on many members of those societies and, for that reason, considered to be problems (e.g. poverty, unemployment) or subjects that need further improvement (e.g. healthcare, education). We also add a group to cover cross-thematic and non-thematic data. A first sub-theme includes variables that mix themes on purpose (e.g. integrative indicators, complex typologies, scenarios). The second sub-theme refers to base maps (administrative units) and other geographical objects (e.g. grids, cities, networks) or spatial delineations (e.g. morphological zones, functional areas).

Those themes that have not been mentioned in this list should be considered as less interesting for the moment, although this assumption should not be taken as granted.

Besides, it is not feasible to address all the relevant political, environmental or social aspects. Nevertheless, we can still consider different approaches to conjecture about the degree to which different topics will develop and gain more or less visibility. Here, we argue that text mining methods and tools have the capacity to identify key words on documents that both employ and communicate ESPON evidence and results. We assume that such approach would contribute to a comprehensive thematic structuring of the ESPON 2013 DB. For the moment, it is not obvious that this analysis will introduce new themes or sub-themes within the predefined structure. The emphasis on a particular theme also depends on other variables such as data deliveries (i.e. indicators, indices, typologies), but also the demand from users and potential users, or even EU policy agenda. Whether this occurs or not, many other themes and sub-themes are likely to be added to the ESPON 2013 DB.

## **4. Text mining methods and visualisation tools**

In the previous section of this report we assume that database structures adopted by international organisations represent an important source of information to extract themes. For this purpose, we apply a visual grouping technique to illustrate, by means of correlation matrices, homogenous clusters of words. The rationale defined for sub-themes is slightly different. Here, we use a large collection of textual data to extract potential keywords to label sub-themes. More concretely, for each of the themes that emerged from our experiment with GAP we employ text mining methods and tools on documents related to each of those themes.

The goal of text mining is to find patterns across textual data and, therefore, derive new information. Such methods enable users to identify keywords that, inductively, create thematic overviews of large text collections. Against this background, we argue that text mining methods may positively support the thematic structuring of the ESPON 2013 DB. It is accepted that ESPON introduced new vocabulary of spatial concepts which, in turn, have definitely influenced the terminology adopted by EU institutions. We make use of this evidence to extract keywords from qualitative and unstructured data, in particular ESPON scientific reports but also texts delivered by EU institutions that use or make reference to ESPON evidence. In order to further enrich this approach we depict the results with mapping techniques of multidimensional scaling.

### **4.1. Methods**

The goal of our investigation is to identify keywords on textual data according to their co-occurrence and use that information to conveniently structure the ESPON 2013 DB. As explained in the introductory part of this report, our contribution will only focus on keywords that can be used as sub-themes. For this purpose, we employ the findings of the previous section where we propose a list of 7+1 first-level themes. Primarily, it is important to identify documents that potentially address each of the themes proposed in the last section of this report and, secondly, ensure that we integrate ESPON reports with evidence-based knowledge on European territorial potentials and dynamics.

The most challenging task before applying text mining tools is data preparation (Berry, 2004; Weiss et al. 2005). Due to the fact that textual data is unstructured and often arranged inconveniently it is necessary to follow certain procedures to ensure some consistency to the overall process. The first step is obviously to collect data. In our case this represents any document, study or policy note addressing ESPON evidence and results.

For this purpose, we have initially identified 27 final project reports delivered by the ESPON 2006 Programme. Our desk research expanded then to reports published in the current ESPON Programme and documents released by other sources that offer a wide range of perspectives to ESPON knowledge (e.g. European Parliament, European Commission). In total, we have collected 53 documents (see Annexes 6-12). Altogether, these documents

constitute a large corpus of textual data that needs to be structured as efficiently as possible before applying any methodological approach.

Similarly, we have to bear in mind that textual data is a complex conjunction of words and phrases that frequently need to be considered as a whole. There is a quite huge amount of dependency that should not be ignored. Moreover, it is also important to overcome word and semantic ambiguities that may adversely influence our analysis. To this end, the usability offered by the software WordStat is quite straightforward and no additional expertise is needed (Lewis, 1999; Davi et al., 2005). The pre-processing of our text collection took into account some of the features offered by this text mining module, particularly with regard to stop-word lists and lower case conversion.

One of the most interesting features provided by this software is the compilation of non information-bearing words that basically exclude terms without any predictive capability, such as articles, pronouns or prepositions. These words are often characterised as noise data and hardly add new information. Besides, it is also possible to add more words to this dictionary of stop-words and improve the accuracy of the corpus for analysis. It should also be mentioned that WordStat merely records the number of times a word appears within a text regardless the content of a sentence or paragraph. After computing data there is a wide variety of ways in which the result can be displayed. The most basic output offered by this application is the word frequency distribution. This knowledge will constitute the basis to explain our results.

As a first step, we apply a pre-defined list on non-information bearing words with no semantic value. Next, we make use of word lemmatization to reduce inflectional form of words to a common root word and ultimately exclude words based upon a frequency criterion. With this regard, we suggest that words below 100 occurrences should be ignored from our analysis. This option exposed a significant number of words that allowed us to further analyse the knowledge structure in text collections.

However, several authors state that words with high frequency distribution do not offer a solid basis for analysis. With this regard, Luhn (cited by Blanchard, 2007: 309) says that 'mid-frequency words in the distribution have both high significance and high discriminating power'. This means that words above an upper cut-off and below a lower cut-off should be removed and, as a consequence, more effort should be added to those that have a mid-frequency (see Figure 1). We took into consideration these aspects and defined a threshold from the estimated densities in order to make emerge other words with explanatory power. The threshold for words with high explanatory power is based on the difference between term frequencies. In other words, the highest gap determines the cut-off.

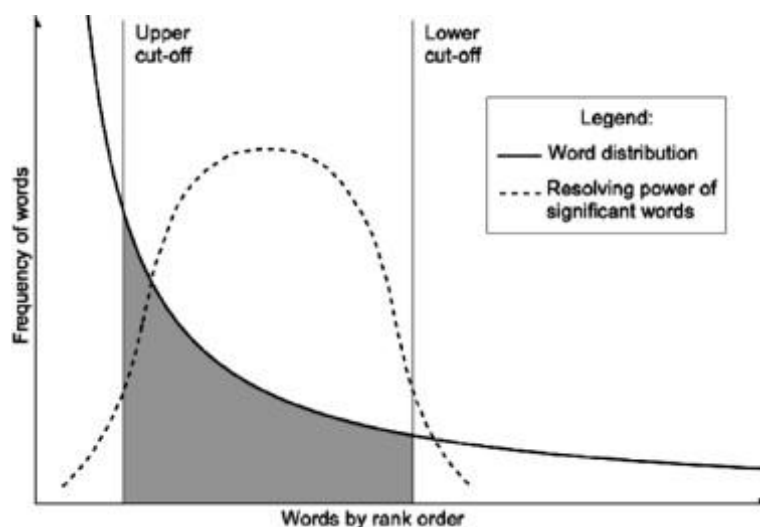


Figure 1: Illustrative plot of word distribution. The power of significant words in higher between the two cut-offs (in grey). Source: Luhn (cited in Blanchard, 2007: 310).

In order to further evaluate the explanatory power of terms in large corpora of textual data, we decided to use TF\*IDF (i.e. Term Frequency \* Inverse Document Frequency). Basically, this scoring algorithm is a weight that measures the importance of each term in a collection of documents. As a result, the importance increases proportionally to the number of times a term appears in a specific document but is counterbalanced by the frequency of the term in the same corpus or text collection. To a certain extent, TF\*IDF algorithm computes the relevance of a document with respect to a particular term. Here we limited our analysis to the 100 most significant terms.

In addition, other features have been applied. Our initial assumption also questioned the ability to use such tools to determine co-occurrence analysis on data derived from a large textual corpus. This function is available on WordStat and it can be used in a variety of ways. For this experiment we have considered two distinct parameters of data co-occurrence – paragraph and group of 5 words. The output file is square matrix where the rows and columns refer to the words and the value of each cell corresponds to the number of co-occurrences. However, it is important to point out that the methods described in this part of the report are solely applied for each of the seven themes that came out from our analysis with database structures. That is, we selected a set of documents for each theme and, based on these documents, prepared data for empirical analysis.

Ultimately, we employ mapping techniques to depict similarities. According to Yang et al (2008), one of the crucial needs in text mining is the ability to visualize relationships between words extracted from large text collections. In order to simplify the analysis of data co-occurrence for words with both explanatory power and non-explanatory power we apply a mapping technique of multidimensional scaling to display information, i.e. VOSviewer (Eck & Waltman, 2007; Eck & Waltman, 2008; Eck et al, 2010). The comparison of these two approaches is also subject to discussion.

## **4.2. Results**

In this section we describe the results obtained for the seven themes that came out from our analysis with database structures. In addition, we review the rationale defined for the theme on both cross-thematic and non-thematic data. For each theme and sub-theme we provide an ordinal code of, respectively, two and four digits to simplify membership. In addition, we coded the last theme with the number 99 so that more themes can be added to the ESPON 2013 DB. Apart from the theme on cross-thematic and non-thematic data, the interpretation of the other themes solely refers to the annexes present in this report.

### **4.2.1. Agriculture and Fisheries**

The maps illustrated in Figure 1a and 1b correspond to the density view of data co-occurrence in a multidimensional scaling. For data on 'Agriculture and Fisheries', we selected 7 texts that combine ESPON evidence and results on issues related to the above mentioned topic (see Appendix 6).

The comparison of these maps suggests that words in Figure 1a are less concentrated than Figure 1b. Somehow this can be explained by the dimension of each parameter for data co-occurrence. Despite the relevance of this aspect what is interesting here is that, in both cases, several densities can be depicted. The main densities aggregate more or less the same number of words and therefore similarities are very alike. This is quite evident for inflectional variants of words, such as 'territorial', 'develop', and 'impact', or 'structur', 'farm', and 'agricultur'.

As stated in the methodology part of this report, the threshold for words with high explanatory power is based on the difference between word frequencies. As it can be seen in Figure 1b, such approach increases the importance of other words, such as 'environment', 'farm' and 'product'. Despite the significance of this adjustment to identify

an explanatory power on words with mid-frequency distribution, the exclusion of anchor words produces an effect that suggests the concentration of many words in just one individual density. This is particularly clear on both parameters.

The ambiguity of the results observed in the previous maps is now reduced to clear and visible densities of words. In addition, it also indicates that terms with high frequency distribution may also disclose similar TF\*IDF importance. That seems to be the case for 'territori' and 'impact'<sup>2</sup>. If we look to the results in more detail, it also becomes evident that rural development and the Common Agricultural Policy receive similar attention.

The number of sub-themes as well as the labelling to describe the content of each sub-theme is therefore based on textual data co-occurrence by means of visualisation techniques in VOSviewer. The sub-themes related to 'Agricultural and Fisheries' looks as follows:

#### 01 AGRICULTURE AND FISHERIES

0101 Farm structure (e.g. farm type, income from farming, organic farming)

0102 Livestock (e.g. livestock output)

0103 Aquaculture and sea fisheries (e.g. aquaculture resources in coastal areas)

0104 Forestry (e.g. production, consumption, and import/export products)

0105 Rural characteristics (e.g. rural employment, rural access to services)

The first sub-theme is designed to incorporate data on farm type, size of farms, income from farming and other related components, such as organic farming. The decision to set out this sub-theme derives from the relatedness between 'farm', 'agricultur', and 'structur' (see Figure 1a and 1b). The third sub-theme is meant to integrate data on 'aquaculture and sea fisheries' and its presence is justified by the high TF\*IDF distribution observed for the word 'aquacultur'.

The same applies for 'Forestry' and 'Rural characteristics'. Here, the sub-theme on 'Forestry' is meant to cover data on production, consumption, and trade, while the other reflects the increasing diversity of rural areas in Europe that demand a better understanding of the development opportunities and challenges for the countryside. Therefore, the goal is to address the differences and similarities in regional economic structures through indicators and indices on diversity patterns, including access to services, rural employment, and core-periphery structures.

#### 4.2.2. Demography

The capacity of visualizing and analysing data co-occurrence matrices with VOSviewer reduces the complexity inherent to the content of texts. The collection of documents in the field of 'Demography' (see Appendix 7) indicates a strong emphasis on subjects associated with the following root words: 'popul', 'migrat', and 'labour'. The co-occurrence of such words, regardless the parameter used in this analysis, is quite evident on Figure 2a. Hence, one could assume that the presence of such words is highly important for analysis and eventually contribute to the definition of sub-themes. This assumption is not completely accurate and, as explained previously, some adjustments should be added to our research by withdrawing those words from our corpus and therefore identify other potential words with explanatory power. If we look to both maps in Figure 2b, we recognize inflectional

---

<sup>2</sup> Of course, documents chosen to integrate the corpus for analysis might have influenced the results. But, in any case, this can be explained by the strong emphasis given to ESPON projects on tools and methodologies to evaluate the effects of agricultural policies on the EU territory.

forms of specific words strongly related to the theme on 'Demography'. This is the case for 'fertile', 'develop', 'scenario', 'trend', and 'rural'.

But what would happen to our dataset if we just consider words that emerge frequently in a document but do not have the same frequency in the remainder of the corpus? In other words, what would be the result of our analysis if we just keep a maximum of 100 words based on TF\*IDF distribution? With this regard, some of the major words identified previously can still be seen in Figures 2c and 2d. Despite the usefulness of this approach, Figure 2c does not improve the distribution observed in Figure 2d. By excluding words that demonstrate high TF\*IDF distribution, the density view in both maps is rather similar. Somehow this feature suggests that our corpus is not sensitive to such boundaries and words with high frequency distribution tend to be dispersed more homogeneously across documents (e.g. 'labour', 'immigrate', 'depopula').

Having said that, it is important to interpret the results and derive sub-themes related to this specific theme. Some of the terms are implicitly associated with migration (e.g. immigration, migration replacement) or natural change in population (i.e. fertility, mortality, life expectancy) while others demonstrate close links to demographic ratios, trends, and scenarios in rural, urban and peripheral areas. Surprisingly enough we observe a constant correlation between 'labour' and 'forc'. The fact that these two inflectional forms of words disclose a strong correlation is normal, what is not normal is to see these words isolated from other words equally relevant, such as 'ageing' and 'decline'. The topic itself has been on the EU agenda for some years now and many policy measures have been implemented to address the issue of ageing labour force and ageing in general.

Despite the feasibility of text mining techniques in analysing documents related to specific themes or subjects, we would expect to find other words with high explanatory power. This seems to be the case for 'household', 'health', or 'dependency'. Nevertheless, we believe the information collected for analysis is solid and robust to extract sub-themes. The structure of second-level themes related to 'Demography' looks as follows:

## 02 DEMOGRAPHY

0201 Population structure (e.g. age distribution by group and gender)

0202 Natural changes (e.g. fertility, mortality, life expectancy)

0203 Households (e.g. number of households)

0204 Migrations (e.g. migration replacement, high-skilled labour migration)

The sub-theme related to 'population structure' is meant to understand how EU cities, regions and other territories are made up of people of different ages, and of males and females. Another sub-theme that emerged from our analysis concerns 'natural changes'. This topic is widely covered by statistical databases and is intended to describe the difference between the number of births and deaths. On the contrary, 'households' refers to an individual or a group of individuals who occupy the same dwelling. Here, it is suggested that TPGs should integrate data on persons per households, expenditures per household, but also median income per household member. Finally, the sub-theme 'migrations' is expected to cover data related to international migrations, including migration replacement and high-skilled labour migration. Alongside natural changes, the balance of international migrations is an important component of population growth. In some cases, negative natural population growth can be fully compensated by migration surplus.

The analysis performed on documents related to 'Demography' sets out four distinct sub-themes. These sub-themes try to cover the most significant topics present in the corpus for our initial experiments. It is worth mentioning, however, that this proposal only takes into consideration textual data delivered up to now on ESPON evidence and results, including inception, interim and final reports of both ESPON 2006 and ESPON 2013 Programmes. In



the future, more documents should enrich the corpus of this theme and eventually extract new terms and label other potential sub-themes.

### 4.2.3. Transport

Transport-related issues have always represented a major issue for analysis and discussion within the ESPON Programme. Our corpus features 6 documents that explore subjects related to the transport sector in Europe (see Appendix 8). The major topics under discussion stem from the need to monitor the EU transport policies, such as mobility, accessibility, sustainable transport policies, impacts on the environment, competitiveness of the economy, or leverage effects on EU territorial development. With this regard, a series of applied research projects commissioned by the ESPON 2006 Programme dedicated a large attention to the Trans-European Transport Network (TEN-T) Initiative. Its importance for competitiveness and growth has been recognized as one of the most fundamental initiatives to set out a proper EU transport policy. The findings of these studies enriched many policy discussions on the developments of the TEN-T with new data on transport networks and traffic flows, including performance indicators, typologies, and scenario-based projections.

The mapping perspective of data co-occurrence in Figure 3a varies according to the parameter. To a large extent, it is correct to say that we observe the same densities. For instance, 'transport', 'access', 'impact', or 'network' confirm their importance. But when we exclude these terms from our dataset the result conveys new densities or clusters of words. This is the case for 'infrastructur', 'develop', 'model', 'scenario', 'road', and 'rail'. If we now focus our analysis on terms with high TF\*IDF distribution, we see that some of the terms identified previously become less visible or, inclusively, unobserved. Due to the indexing rationale behind TF\*IDF measure, words like 'flow', 'model', 'ten', or 'gdp' gain more visibility. To some extent, this situation increases the robustness of the dataset by focusing on terms with high levels of explanatory power. As it can be seen in Figure 3c and 3d, the best terms to label sub-themes correspond to 'model', 'scenario', and 'ten'.

Clearly, one of the best densities that can be identified in both Figure 3c and 3d correspond to a group of words that include 'air', 'rail', 'maritim', 'traffic', and 'flow'. Somehow this shows a strong focus of our corpus on transportation systems. Similarly, it seems that 'scenario', 'model', and 'forecast' suggest a distinct density. In this case, one could say that the impact of EU transport policy compromises medium and long-term scenarios. Also important, but less merged with other words, is 'ten'. Here the word corresponds to Trans-European Transport Network (TEN-T) and most of the maps depict data without making any sort of similarity linkage to other potential words. The only exception are the maps of data co-occurrence based on TF\*IDF distribution.

In sum, we believe our analysis discloses some hidden knowledge on a dataset of text information related to 'transport' and, by doing so, facilitates the decision-making process by reducing uncertainty and doubt on words to label sub-themes. Against this background, we suggest that the above-mentioned theme of the ESPON 2013 DB should integrate the following sub-themes:

#### 03 TRANSPORT

0301 Accessibility (e.g. performance indicators, multimodal accessibility)

0302 Flows (e.g. vehicles, passengers, goods, freight)

0303 Infrastructure (e.g. transportation systems, railways, airports, harbours)

To a certain extent, the proposal of sub-themes to allocate data in the field of 'Transport' is similar to the one suggested in the ESPON 2006 Database (ESPON, 2005). Nevertheless,

we believe that such proposal combines a comprehensive structure of transport information, ranging from data on flows of vehicles, passengers and goods to infrastructures, safety, and investments in the transport sector. It considers as well data that could be delivered on sustainable development, modal split, and environmental impact indicators, including the contribution of each mode of transport, used alone and in combination with others.

#### **4.2.4. Energy and Environment**

We assume that energy and environment are complementary and, in many ways, essential for sustainable development. Several policy documents delivered by EU institutions state that sustainable development corresponds to the improvement of citizen's quality of life while reducing the use of natural resources and pressures on the environment (CEC, 2001; EEA, 2002). However, the quality of life is enhanced by costly energy services. The main question, according to these institutions, is how to make use of available energy resources without preventing the needs of future generations. In order to meet the right balance it is necessary to consider other aspects, such as climate change, loss of biodiversity, or ozone layer depletion.

The empirical comparison of data co-occurrence for the two parameters (i.e. paragraph and group of 5 words) is far too similar. The densities observed in both maps only give prominence to 'energy' (Figure 4a). This result was expected, in part, due to the high presence of inflectional forms of words related to this specific word. However, when we add this word to an exclusion list the picture obtained is rather different as other densities emerge. This seems to be the case for 'climat' and 'chang', 'environment' and 'sustain', or 'urban', 'transport' and 'demand'. It is also interesting to note the presence of a density related to industry, biomass and fossil fuels. Here we noticed that replacing fossil fuels by sustainable-produced biomass is seen has a safe method to reduce CO<sub>2</sub> emissions to the atmosphere and therefore the negative impact on the environment (Gustavsson, 1995; Forsberg, 2002).

Let us now consider the results of data co-occurrence based on TF\*IDF distribution in somewhat more detail. The density view of both maps is meaningful because it shows the presence of some terms identified previously and related to nuclear, fossil, and renewable energy sources (e.g. 'oil', 'coal', 'wind', 'solar', 'nuclear', 'thermal'). The difference is that now we just consider terms with high explanatory power and try to depict the information using distinct parameters of linguistic discourse. Here, it is visible the presence of terms that express concern about the subject in analysis, such as 'sensit' or 'vulnerab', or even terms that call for adaption, such as 'adapt'. This is even more evident for densities that combine 'household', 'gdp', and 'employ' to illustrate some of the possible effects emanating from climate change. In a way, the combination of these terms is understandable, especially if we consider that low income households tend to live in areas with low GDP growth, high unemployment rates and therefore more likely to be affected by climate change, and have far less ability to move or make the necessary adjustments to their living conditions.

After conducting the co-occurrence analysis in our corpus, we propose two sub-themes to structure data delivered by TPGs in the field of 'Energy & Environment'. The sub-theme structure and inheritance is the following:

#### **04 ENERGY AND ENVIRONMENT**

0401 Energy and resources (e.g. renewable, nuclear, and fossil energies)

0402 Environmental facets of climate change (e.g. GHG emissions, air pollution)

The first sub-theme is intended to include data on energy and resources, including renewable, nuclear, and fossil energies. The second sub-theme should incorporate distinct features or elements that actively contribute to climate change (e.g. air and soil pollution, biodiversity loss, water management, greenhouse gas emissions).

#### **4.2.5. Land Use**

Land use refers how the earth's surface is used, including the location, type and design of human development. As a result, land use patterns have diverse economic, social and environmental impacts. In the previous ESPON Database, 'Land Use' is defined as a first-level theme but its inheritance is somehow vague in the description<sup>3</sup>.

The collection of documents used for analysis is not very substantial. Nevertheless, we have managed to gather some documents based on research activities conducted by EU institutions or commissioned to universities and research institutions on behalf of the funding entity (see Appendix 10). The visualization of similarities of terms extracted from such context provided a better understanding of data co-occurrence. In Figure 5a we can see the pattern of similarities between terms with high frequency distribution for documents related to 'Land Use'. Clearly, 'urban', 'model', 'land', and 'area' are among those. However, the knowledge obtained from the exclusion of these terms conveys other relationships. This seems to be the case for 'chang', 'impact', 'environment' and 'develop'. The same applies for 'agricultur', 'produc', and 'cropland'. Less visible, but still important, is relationship between 'transport', 'sprawl', and 'energi'. The term 'scenario', itself very important in land use discussions, appears completely isolated from the other main densities. Overall, the maps presented in Figure 5b disclose relevant information for analysis.

Among the visible interactions established by our corpus in the field of 'Land Use', it is possible to identify a sequence of terms that correspond to changes in land-use for both rural and urban settings. This facet seems to be evident for terms like 'chang', 'impact', 'rural' and 'urban'. In fact, most of the impacts related to land-use have an effect in rural and urban contexts. This also holds true for socio and economic factors. In addition to these, the term 'scenario' also discloses some significance. The ability to forecast land-use scenarios is essential to better understand dynamic processes which are determined by a range of driving forces, including demographic, socio-economic, and environmental change.

If we consider the strong emphasis given to these terms and its similarity, one could assume that the focus of our corpus is oriented to land-use changes, impacts and scenarios. As a consequence, we propose the following structure of sub-themes on 'Land Use' data:

#### 05 LAND USE

0501 Land use and land cover types (e.g. CORINE Land Cover, GMES)

0502 Urban land use attributes and changes (e.g. LUZ, Urban Atlas)

0503 Rural land use attributes and changes (e.g. Natura 2000)

The first sub-theme has not been defined with text mining tools. However, its purpose is to integrate data related to CORINE Land Cover (CLC), Natura 2000, and the Urban Atlas Initiative. The other two sub-themes derive from a qualitative description of term similarity maps within a corpus of documents in the field of 'Land Use' and are meant to integrate data on changes, including indicators and indices on land use changes and impacts.

---

<sup>3</sup> Two sub-themes have been defined to structure data in the field of 'Land Use'. These are: '111 Natural resources' and 'Land use' (ESPON, 2005).

#### **4.2.6. Social Affairs**

The theme related to 'Social Affairs' is meant to cover data on social, economic and cultural issues with an emphasis on employment, labour market, income, living conditions, and poverty. Our collection of documents is based on ESPON evidence and results and, alternately, findings from other sources used by TPGs while undertaking research in this sort of topics. In total, we collected 12 documents from different sources, ranging from studies, policy notes and technical reports (see Appendix 11).

Unfortunately, in this case, the density maps of data co-occurrence do not add any relevant information. This means that the maps on both parameters are dominated by terms with little explanatory power and therefore terms with the highest frequency distribution should be added to an exclusion list. This seems to be the example for 'polici' and 'social' (see Figure 6a). However, if we remove these terms from our analysis the density map will depict other similarities offering a better understanding of how inflectional forms of words are associated with each other. Here, a special mention should be made for 'labour' and 'employ', 'econom' and 'develop', and 'indic' and 'cultur' (see Figure 6b).

The capacity of visualising similarities in a multidimensional scaling dramatically increases with the TF\*IDF scoring algorithm. As it can be seen in Figure 6c and 6d, the knowledge structure is relatively easy to comprehend. Somehow, these maps suggest that TF\*IDF measure reinforces the importance of terms less visible within the corpus. Besides, it clearly differentiates the major similarities. For instance, 'cultur' and 'heritag' reveal a distinct similarity. The same applies for 'labour' and 'job', 'household' and 'health', or 'incom' and 'famili'. Equally relevant is the presence of 'poverti'. Most of these similarities underline the rationale behind the theme in the field of 'Social Affairs' and, as a consequence, facilitate the definition of sub-themes.

The thematic structure designed for the ESPON 2006 DB suggests that data on employment and labour market should be disconnected from social exclusion (e.g. poverty) (ESPON, 2007). However, the results of this experiment indicate the opposite meaning that both should be integrated and, if possible, include data on similar issues, such as living conditions and health systems.

#### **06 SOCIAL AFFAIRS**

0601 Education (e.g. training, lifelong learning)

0602 Labour market (e.g. labour force, labour costs, economic inactivity, earnings)

0603 Living conditions (e.g. poverty, social exclusion, health systems)

0604 Culture (e.g. socio-cultural activities, cultural consumption)

The sub-theme on 'education' is designed to integrate data on training and lifelong learning while 'labour market' is more focused on economic inactivity, average earnings, and productivity. Within the same structure we suggest a second-level theme related to 'living conditions'. Ideally, this sub-theme will serve the purpose of integrating data on poverty, social exclusion as well as other types of living conditions, including data on health systems. Finally, our experiment with text mining also suggests a strong emphasis on issues related with culture and heritage. With this respect, we propose a sub-theme to allocate data on socio-cultural activities, cultural consumption and participation.

#### **4.2.7. Economy**

The economic analysis of EU strengths and weaknesses is of great importance to understand the policy designed by the Lisbon Strategy. According to the European Council its goal was to make the EU 'the most competitive and dynamic knowledge-based economy

in the world capable of sustainable economy growth with more and better jobs and greater social cohesion' (CEC, 2000).

Despite joint efforts to achieve these goals only a small number of actions could have been fully implemented. One of the recent drawbacks to justify the moderate success of this initiative is the serious economic crisis that hit Europe and its citizens. As a response to such event, the European Commission has launched the 'Europe 2020 Strategy' in order to re-adapt the Lisbon Strategy to new challenges (CEC, 2010).

Access to accurate data is therefore of crucial importance to comprehend, for instance, the role of R&D, innovation, and patents to boost competitiveness in Europe. With this regard, the ESPON 2013 Database has the opportunity to structure both statistical and geographical data related to these topics.

The analysis undertaken to extract terms and label sub-themes in the field of 'Economy' is based on a dataset of 10 documents that combine ESPON evidence and results (see Appendix 12). The visualisation of knowledge structure created by these documents in a VOSviewer environment is meant to capture the similarity degree of terms with both explanatory power and non-explanatory power. As it can be seen in Figure 7a and 7b, the keyword co-occurrence of inflectional words is quite obvious for 'develop' and 'econom'. The same applies for 'innov' and 'research', or 'fund' and 'structur'.

Surprisingly enough, the exclusion of terms with high frequency distribution (i.e. 'develop' and 'econom') does not demonstrate the expected impact on the density view of data co-occurrence and, therefore, its structure neither changes nor generate unseen similarities. In this sense, the major subgroups identified previously can still be seen in the VOSviewer maps of Figure 7b. It is particularly interesting to observe the strong correlation between 'servic', 'industri', and 'capit', the continuous emphasis on 'knowledge', 'innov', and research', or the association established between 'territorial' and 'structur'.

The application of the TD\*IDF measure also plays a decisive role in this analysis. The VOS maps presented in Figure 7c depict the knowledge structure of data co-occurrence based on TF\*IDF distribution and, Figure 7d, illustrates what would happen if we added the most significant TF\*IDF terms to an exclusion list. In general, our results do not add more information than what we have so far exposed. Therefore, the relationship established by terms with high TF\*IDF distribution did not extract unknown and potential information. The terms identified by the preceding experiment confirm this evidence. Hence, we propose the following sub-themes in the field of 'Economy':

## 07 ECONOMY

0701 Aggregated accounts (e.g. GDP, purchasing power parities, balance of payments)

0702 Employment (e.g. employment, unemployment, long-term unemployment)

0703 Production and costs per sector (e.g. production of manufactured goods)

0704 Research and innovation (e.g. R&D expenditure, ICT research, patents)

The knowledge structure of our corpus suggests the presence of four sub-themes. The first sub-theme is the less visible in our maps. Still, we decided to introduce it so that TPGs involved in ESPON projects may allocate data on GDP and its main components. The other three sub-themes emerged more implicitly due to the high similarity or relatedness of terms and its clear organisation in clusters. Therefore, the sub-theme on 'employment' should allocate data on employment and unemployment. The sub-theme related to 'production and costs per sector' is meant to incorporate data on production of manufactured goods. Finally, for 'research and innovation' TPGs are encouraged to integrate data on R&D expenditure, ICT research, patents, and public investments.

#### **4.2.8. Cross-Thematic and Non-Thematic Data**

This theme is meant to cover both cross-thematic and non-thematic data. A first sub-theme should include variables that mix themes (e.g. integrative indicators; complex typologies; trends and impacts on both CAP and TEN-T; scenario-based projections on urban development; environmental, social, and economic concerns associated with land-use changes). In addition, it should integrate indicators and indices aimed at evaluating the sensitivity and vulnerability impacts. The second subset refers to base maps (i.e. administrative units) and other geographical objects (e.g. grids, cities, networks) or spatial delineations (e.g. morphological zones, functional areas). Its current structure looks as follows:

99 CROSS-THEMATIC AND NON-THEMATIC DATA

9901 Integrative indices, indicators and scenarios (e.g. typologies, scenarios)

9999 Geographical objects (e.g. administrative units, grids, networks)

We should stress that this proposal is open to more sub-themes. This assumption is also valid for the themes described previously. However, its expansion should depend on datasets delivered by future ESPON projects.

## **5. Conclusions and future work**

In this technical report, we have proposed a two-step approach to structure the ESPON 2013 DB by themes and sub-themes. First, we assume that database structures constitute an important resource of information. To a certain extent, this knowledge reflects the structure of the previous database and will certainly influence the current developments. For this purpose, we apply a visual grouping technique to illustrate, by means of correlation matrices, homogeneous clusters of words. The findings of our experiment constitute the basis to derive a first set of themes and eventually facilitate data allocation.

In the second part of the report we define sub-themes. Here, we believe the demand from the ESPON 2013 DB end-users will correspond to immediate, easy and practical access to datasets. In order to meet this request, we explore the potentialities offered by text mining methods and tools. The rationale of this approach is to find patterns across textual data and generate simple overviews of large text collections.

Our collection of data is based on ESPON evidence and for each theme we demonstrate that it is possible to shed some light on ways to further progress in this field. The definition of sub-themes has been data-driven. This assumption has greatly benefited from mapping techniques of multidimensional scaling to ease the interpretation of relational similarities that came out from data co-occurrence with both explanatory power and non-explanatory power. The identification of these patterns suggests that the ESPON 2013 DB should be structured in 27 sub-themes unveiling the inheritance of 7+1 themes meaning that we would add a last theme to cover cross-thematic and non-thematic data (see Appendix 13).

Ultimately, it is necessary to allocate data into themes and sub-themes. For this purpose, we have considered data from of the ESPON 2006 Programme and data delivered up to date, i.e. mid-September 2010. During the course of this analysis we also suggest a potential second theme and sub-theme that somehow could improve classification and data retrieval. If some doubts subsist in our evaluation we propose other words to describe data. Hopefully this rather inductive process will rationalize the ability to restrict a search query when looking for specific datasets and consequently allow end-users to achieve greater level of precision and recall.

Our findings provide a new approach to data structure but also point out to considerable future work, of both empirical and conceptual nature. At the empirical level, it is clear that we need to refine our understanding of what is being measured to better allocate each variable to a specific theme and sub-theme. The quality of metadata is of course crucial in that regard. But, perhaps more fundamentally, there are some open questions at the conceptual level. Here, future work should validate the usability of this method. Taking into consideration that ESPON will launch new calls for proposals it is important to understand what kind of data future projects will deliver. If there's no appropriate theme and sub-theme, new analysis should be performed on text collections to improve the quality of the ESPON 2013 DB thematic structure. Hence, some of the difficulties that might emerge in the future should be further investigated by means of text mining methods.

## References

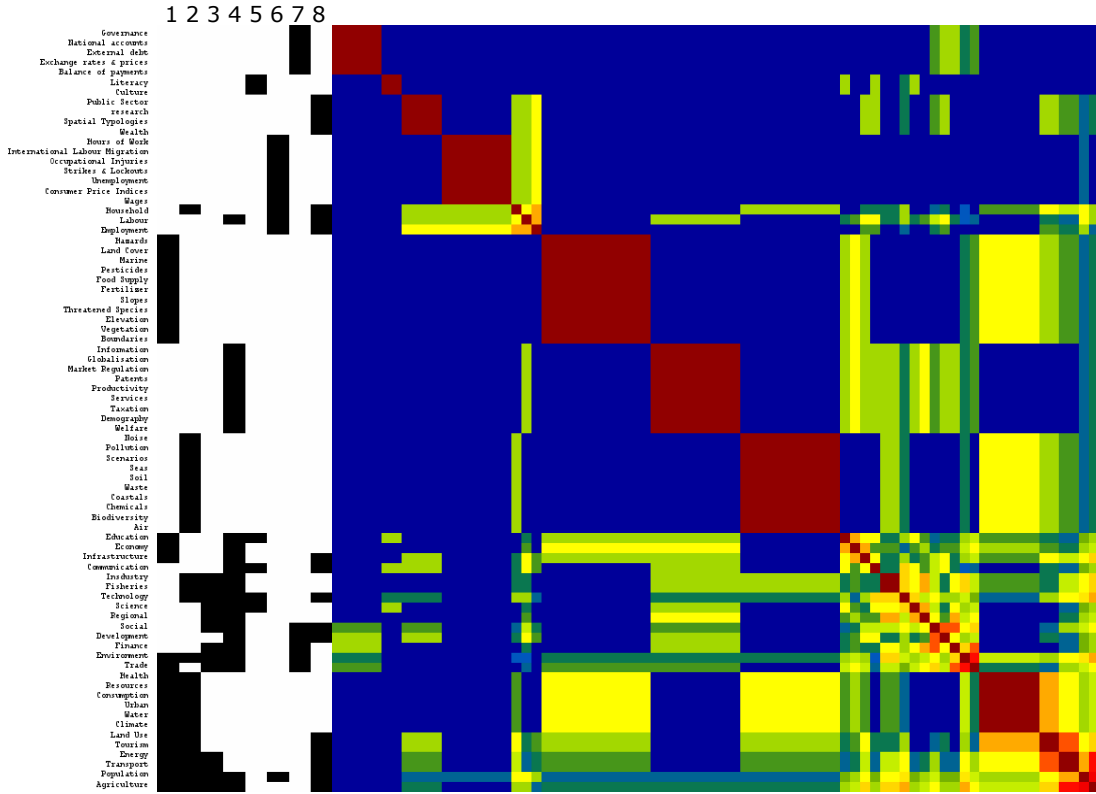
- Berry, M. [ed.] (2004). Survey of text mining. Clustering, classification, and retrieval. Springer: New York.
- Blanchard, A. (2007) Understanding and customizing stopword lists for enhanced patent mapping. *World Patent Information*, 29 (1), 308-316.
- CEC (2001) Environment 2010: Our future, our choice. Communication from the Commission to the Council. Commission of the European Communities: Brussels.
- CEC (2010) Europe 2020. A European Strategy for Smart, Sustainable and Inclusive Growth. Commission of the European Communities.
- Davi, A.; Haughton, D.; Nasr, N. et al (2005). A review of two text mining packages: SAS TextMining and WordStat. *American Statistician*, 59(1), 89-103.
- Eck, N. & L. Waltman (2007). VOS: a new method for visualizing similarities between objects. In H.-J. Lenz and R. Decker [ed.]. *Advances in Data analysis: Proceedings of the 30th Annual conference of the German Classification Society. Studies in Classification, Data Analysis, and Knowledge Organization*. Springer: Heidelberg, 299-306.
- Eck, N. & L. Waltman (2009). A computer program for bibliometric mapping. In B. Larsen and J. Leta [eds.]. *Proceedings of the XII International Conference on Scientometrics and Informetrics*, 886-897.
- Eck, N. et al (2010) A comparison of two techniques for bibliographic mapping: Multidimensional scaling and VOS. *Journal of the American Society for Information Science and Technology*, 61(12), 2405-2416.
- EEA (2002) Energy and environment in the European Union. Environmental Issue Report No 31. European Environment Agency: Copenhagen.
- ESPO (2005). Integrated Tools for European Spatial Development. ESPON 3.1 Project. Luxembourg: ESPON Coordination Unit, pp. 141-174.
- ESPO (2006) Feasibility study on monitoring territorial monitoring territorial development based on ESPON key indicators. ESPON Project 4.1.3 Final Report. BBR: Bonn.
- European Council (2000) Presidency conclusions. Lisbon European Council, 23-24 March 2000. European Council: Brussels (available for download on [www.consilium.europa.eu](http://www.consilium.europa.eu)).
- Forsberg, C. (2008) Sustainability by combining nuclear, fossil, and renewable energy sources. *Progress in Nuclear Energy*, 1-9.
- Gustavsson, L. et al (1995) Reducing CO<sub>2</sub> emissions by substituting biomass for fossil fuels. *Energy* 20(11), 1097-1113.
- Lewis, B. (1999). Simstat with Wordstat: A Comprehensive Statistical Package with a Content Analysis Module. *Field Methods*, 11(2), 166-179.
- Weiss, S.; Indurkha, N.; Zhang, T. and F. Damerau (2005). Text Mining. Predictive Methods for Analysing Unstructured Information. Springer: New York.
- Chen, C.-H. (2002). Generalized Association Plots: Information Visualization via Iteratively Generated Correlation Matrices. *Statistica Sinica*, 12(1), 7-29.
- Wu, H.-M., et al. (2008). GAP: A graphical environment for matrix visualization and cluster analysis. *Computational Statistics & Analysis*, in press.
- Yang, Y.; Akers, L.; Klose, T. and Yang, C. (2008) Text mining and visualization tools. Impressions of emerging capabilities. *World Patent Information*, 30 (1), 280-293.



## **Appendix 1**

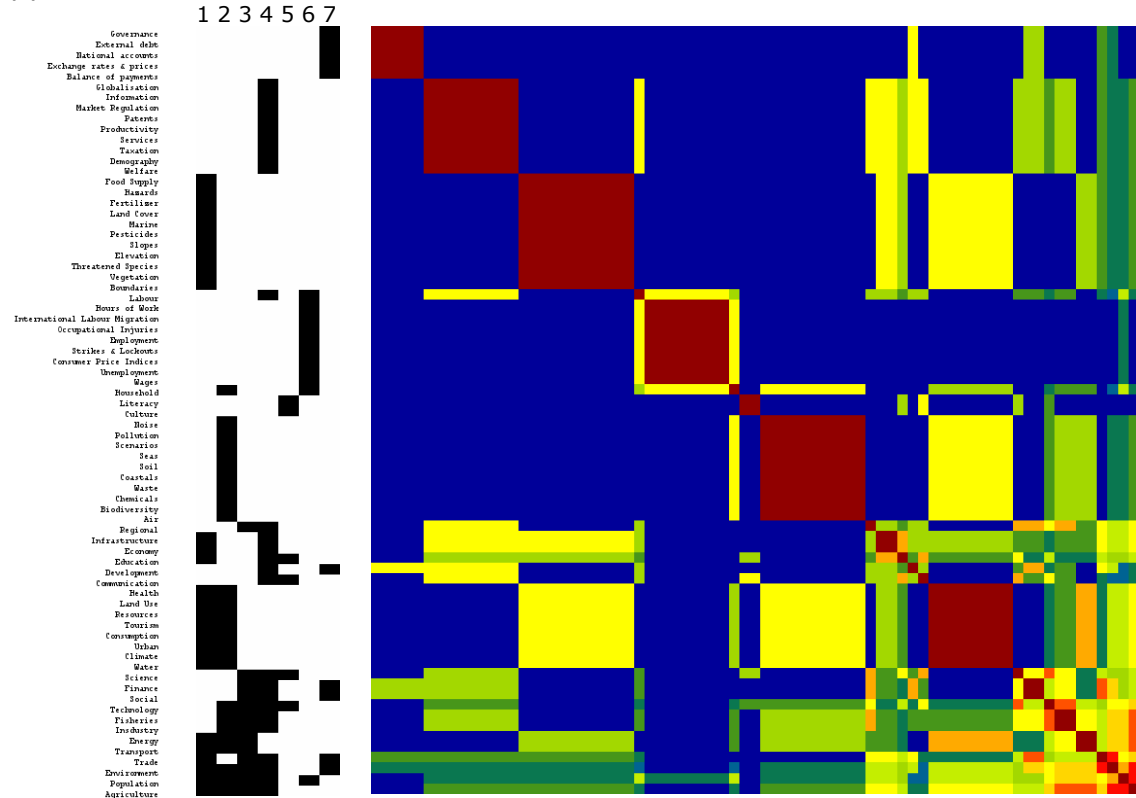
**Matrix visualisation in GAP environment of words used by statistical databases to label themes. Matrix (a) includes the structure of the ESPON 2006 DB while (b) ignores it**

(a)



Minimum Maximum

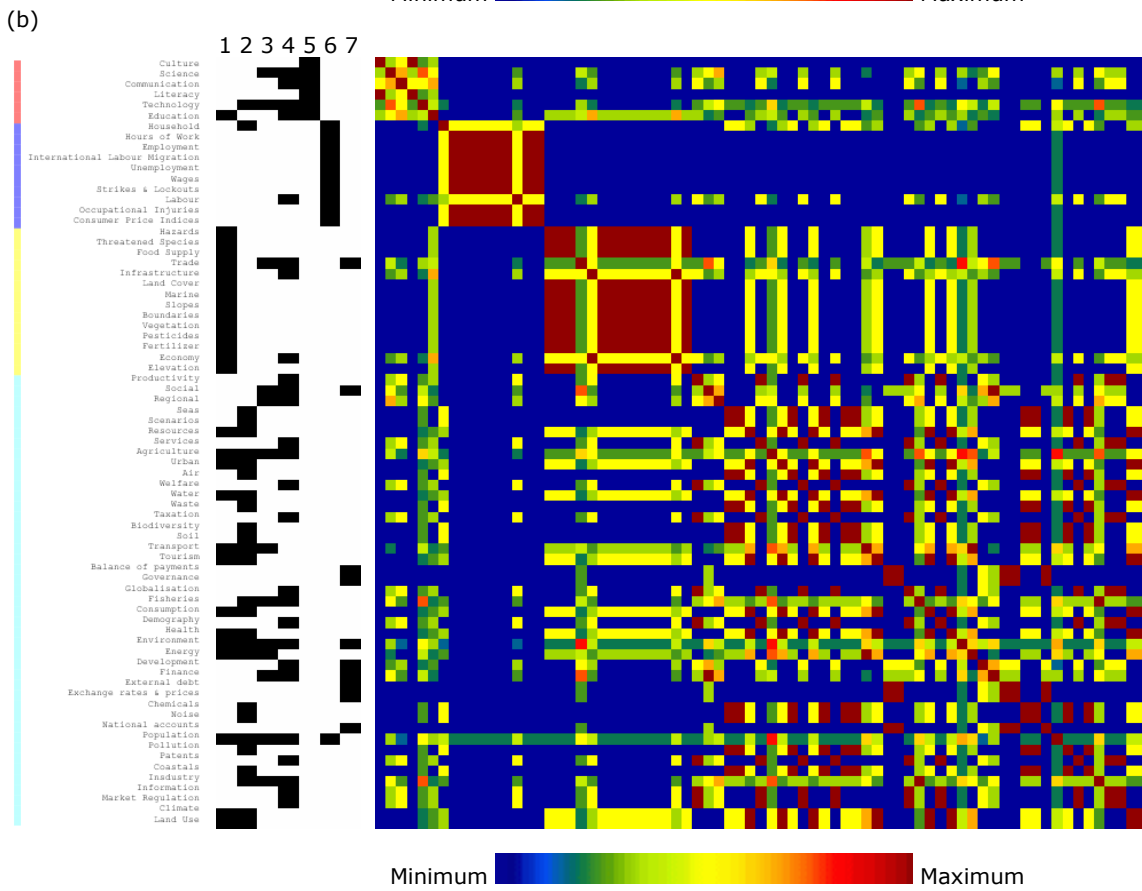
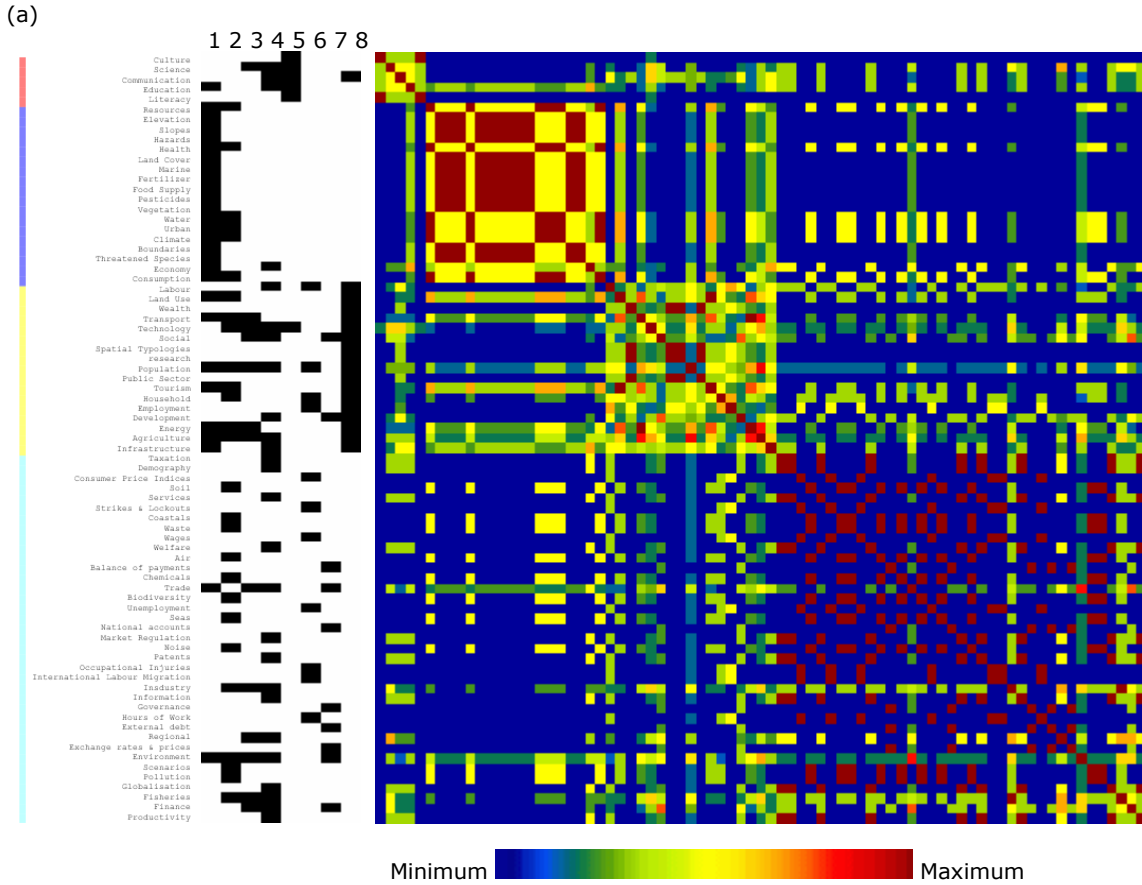
(b)



Note: (1) UNEP, (2) EEA, (3) EUROSTAT, (4) OECD, (5), UNESCO, (6) ILO, (7) WPI, (8) ESPON 2006.

## **Appendix 2**

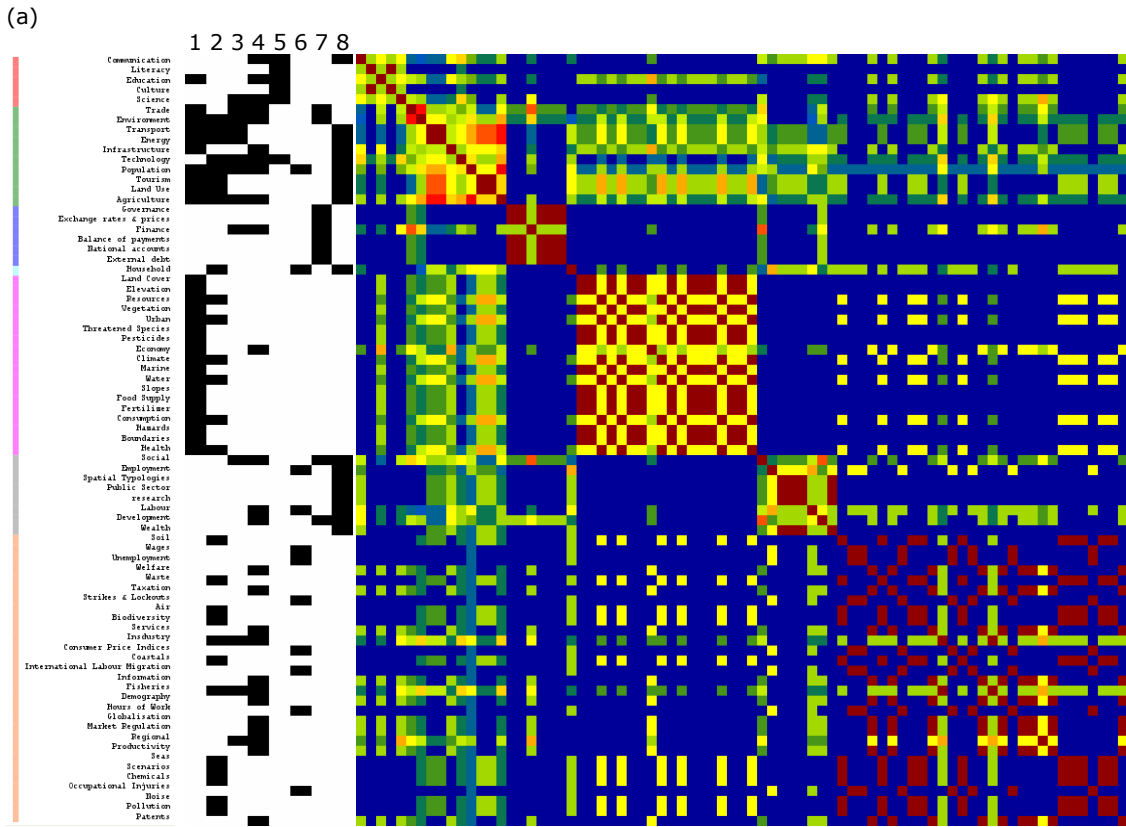
**Cluster analysis based on Jaccard's coefficient in GAP environment. Matrix (a) includes the structure of the ESPON 2006 DB while (b) ignores it**



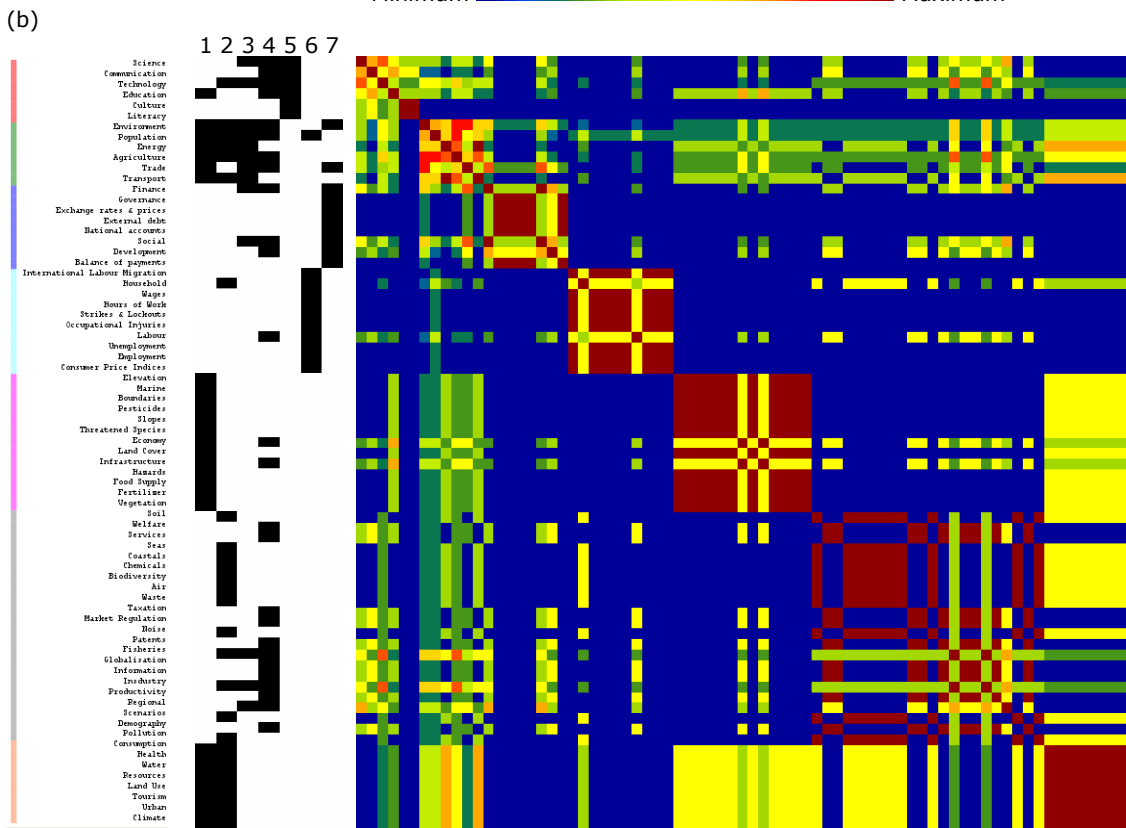
Note: (1) UNEP, (2) EEA, (3) EUROSTAT, (4) OECD, (5), UNESCO, (6) ILO, (7) WPI, (8) ESPON 2006.

## **Appendix 3**

**Cluster analysis based on Jaccard's coefficient in GAP environment. Matrix (a) includes the structure of the ESPON 2006 DB while (b) ignores it**



Minimum Maximum



Minimum Maximum

Note: (1) UNEP, (2) EEA, (3) EUROSTAT, (4) OECD, (5), UNESCO, (6) ILO, (7) WPI, (8) ESPON 2006.

## Appendix 4

### Database structures ordered by themes

#### UNESCO

- 1 Education
- 2 Science & Technology
- 3 Culture & Communication
- 4 Literacy

#### ILO

- 1 Economically Active Population
- 2 Employment
- 3 Unemployment
- 4 Hours of Work
- 5 Wages
- 6 Labour Cost
- 7 Consumer Price Indices
- 8 Occupational Injuries
- 9 Strikes and Lockouts
- 10 Household Income and Expenditure
- 11 International Labour Migration

#### EUROSTAT

- 1 General and Regional Statistics
- 2 Economy and Finance
- 3 Population and Social Conditions
- 4 Industry, Trade and Fisheries
- 5 External Trade
- 6 Transport
- 7 Environment and Energy
- 8 Science and Technology

#### OECD

- 1 General Statistics
- 2 Agriculture and Fisheries
- 3 Demography and Population
- 4 Development
- 5 Economic Projections
- 6 Education and Training
- 7 Environment
- 8 Finance
- 9 Globalisation
- 10 Health
- 11 Industry and Services
- 12 Information and Communication Technology
- 13 International Trade and Balance of Payments
- 14 Labour
- 15 Monthly Economic Indicators
- 16 National Accounts
- 17 Prices and Purchasing Power Parities
- 18 Productivity
- 19 Public Sector, Taxation and Market Regulation
- 20 Regional Statistics
- 21 Science, Technology and Patents
- 22 Social and Welfare Statistics

#### EEA

- 1 Agriculture
- 2 Air
- 3 Biodiversity Change
- 4 Chemicals
- 5 Climate Change
- 6 Coastals and Seas
- 7 Energy
- 8 Environmental Scenarios
- 9 Fisheries
- 10 Households
- 11 Human Health
- 12 Industry
- 13 Natural Resources
- 14 Noise

- 15 Policy Analysis
- 16 Population and Economy
- 17 Regions
- 18 Soil
- 19 Tourism
- 20 Transport
- 21 Urban Environment
- 22 Waste
- 23 Water

#### UNEP

- 1 Agricultural Production
- 2 Boundaries
- 3 Climate
- 4 Economy
- 5 Education
- 6 Elevation and Slopes
- 7 Emissions of GHG and ODS
- 8 Energy Consumption and Production
- 9 Environmental Hazards
- 10 Fertilizer & Pesticides
- 11 Food Supply & Caloric Intake
- 12 Health
- 13 Infrastructure
- 14 Land Use
- 15 Marine and Coastal Areas
- 16 Population
- 17 Private Consumption
- 18 Protected Areas and Environmental Protection
- 19 Technological Hazards
- 20 Total and Threatened Species
- 21 Tourism
- 22 Trade Balances
- 23 Transport
- 24 Urbanisation
- 25 Vegetation and Land Cover
- 26 Water Consumption and resources

#### WORLD BANK

- 1 Agriculture
- 2 Aid
- 3 Childhood Development
- 4 Debt
- 5 Education
- 6 Environment
- 7 Finance
- 8 Gross Domestic Production
- 9 Gender
- 10 Globalisation
- 11 Governance
- 12 Health
- 13 Information Technology
- 14 Infrastructure
- 15 Industry
- 16 Labour & Employment
- 17 Macroeconomics & Growth
- 18 Population
- 19 Poverty
- 20 Purchasing Power Parity
- 21 Private Sector
- 22 Public Sector
- 23 Rural Development
- 24 Social Development
- 25 Trade
- 26 Urban Development

## Appendix 5

### Words employed to label themes in database structures

Words (or expressions)	UNEP	EEA	EUROSTAT	OECD	UNESCO	ILO	WPI	ESPN 2006
Agriculture								
Aid								
Air								
Balance of payments								
Biodiversity								
Boundaries								
Chemicals								
Childhood								
Climate								
Coastals								
Communication								
Consumer Price Indices								
Consumption								
Culture								
Demography								
Development								
Economy								
Education								
Elevation								
Employment								
Energy								
Environment								
Exchange rates & prices								
External debt								
Fertilizer								
Finance								
Fisheries								
Food Supply								
GDP								
Gender								
Globalisation								
Governance								
Hazards								
Health								
Hours of Work								
Household								
Information								
Infrastructure								
Industry								
International Labour Migration								
Labour								
Land Cover								
Land Use								
Literacy								
Macroeconomics								
Marine								
Market Regulation								
National accounts								
Noise								
Occupational Injuries								
Patents								
Pesticides								
Pollution								
Population								
Poverty								
PPP								
Productivity								
Public Sector								
Regional								
Research								
Resources								
Rural								
Scenarios								
Science								
Seas								
Services								
Slopes								
Social								
Soil								
Spatial Typologies								
Strikes & Lockouts								
Taxation								
Technology								
Threatened Species								
Tourism								
Trade								
Transport								
Unemployment								
Urban								
Vegetation								
Wages								
Waste								
Water								
Wealth								
Welfare								



## Appendix 6

### Reports taken into account for text mining purposes on 'Agriculture and Fisheries'

Code	Report
report#01	European Parliament (2007) Regional Dependency on Fisheries. European Parliament: Brussels.
report#02	European Parliament (2007) Reflection on the possibilities for the future development of the CAP. European Parliament: Brussels.
report#03	ESPON (2005) The territorial impact of CAP and Rural Development Policy. ESPON Project 2.1.3 Final Report. Arkleton Institute: Aberdeen.
report#04	ESPON (2005) Territorial Impacts of European Fisheries Policy. ESPON Project 2.1.5 Final Report. NIBR: Oslo.
report#05	ESPON (2010) European Development Opportunities for Rural Areas. EDORA Draft Final Report. UHI Millenium Institute: Inverness.
report#06	ESPON (2010) Territorial Impact for Transport and Agricultural Policies. TIP TAP Final Report A/B. Politecnico di Milano: Milan.
report#07	ESPON (2010) Territorial Impact for Transport and Agricultural Policies. TIP TAP Final Report C. Politecnico di Milano: Milan.

## **Appendix 7**

### **Reports taken into account for text mining purposes on 'Demography'**

Code	Report
report#08	European Parliament (1999) Regional development in less-densely populated regions in the EU. European Parliament: Brussels.
report#09	ESPON (2005) The spatial effects of demographic trends and migration. ESPON Project 1.1.4 Final Report. ITPS: Stockholm.
report#10	ESPON (2010) Demographic and migratory flows affecting European regions and cities. DEMIFER Draft Final Report. NIDI: The Hague.
report#11	CEC (2006) The Demographic Future of Europe: From Challenge to Opportunity. Commission of the European Communities: Brussels.
report#12	CCE (2007) Europe's Demographic Future: Facts and Figures. Commission of the European Communities: Brussels.

## **Appendix 8**

### **Reports taken into account for text mining purposes on 'Transport'**

Code	Report
report#13	European Parliament (2006) The Impact of Trans-European Networks on Cohesion and Employment. European Parliament: Brussels.
report#14	ESPON (2004) Transport services and networks: Territorial trends and basic supply on Infrastructure for Territorial Cohesion. ESPON Project 1.2.1 Final Report. University of Tours: Tours.
report#15	ESPON (2005) Territorial Impact of EU Transport and TEN policies. ESPON Project 2.1.1 Final Report. Spiekermann & Wegener: Dortmund.
report#16	ESPON (2009) Territorial Impact package for Transport and Agricultural Policies. TIPTAP Draft Final Report. Politecnico de Milano: Milan.
report#17	ESPON (2007) Update of selected potential accessibility indicators. Final report. Spiekermann & Wegener: Dortmund.
report#18	CEC (2007) Trans-European Networks: Towards an integrated approach. Commission of the European Communities: Brussels.

## Appendix 9

### Reports taken into account for text mining purposes on 'Energy and Environment'

Code	Report
report#19	ESPON (2005) Territorial Trends and Policy Impacts in the field of EU Environmental Policy. ESPON 2.4.1 Final Report. Geological Survey of Finland: Helsinki.
report#20	ESPON (2005) Territorial Trends of Energy Services and Networks and Territorial Impact of EU Energy Policy. ESPON 2.1.4 Final Report. CEEETA: Lisbon.
report#21	ESPON (2010) Climate Change and Territorial Effects on Regions and Local Economies. ESPON CLIMATE Interim Report. TU Dortmund: Dortmund.
report#22	ESPON (2010) Regions at Risk of Energy Poverty. ReRisk Draft Final Report. Inasmet-Tecnalia: Donostia/San Sebastian.
report#23	European Parliament (2007) Using Sustainable and Renewable Energies in the context of Structural Policy 2007-2013. European Parliament: Brussels.
report#24	European Parliament (2006) Energy and Structural and Cohesion Policies. European Parliament: Brussels.
report#25	European Parliament (1998) Sustainable Development: A Key Principle for European Regional Development. European Parliament: Brussels.
report#26	European Parliament (2003) The Enlargement Process of the EU: Consequences in the Field of Environment. European Parliament: Brussels.
report#27	European Parliament (2008) The Challenge of Climate Change for Structural and Cohesion Policies. European Parliament: Brussels.

## **Appendix 10**

### **Reports taken into account for text mining purposes on 'Land Use'**

Code	Report
report#28	EEA (2006) Urban sprawl in Europe. The ignored challenge. EEA Report No 10/2006. European Environment Agency: Copenhagen.
report#29	EEA (2007) Land-use scenarios for Europe: qualitative and quantitative analysis on a European scale. EEA Technical Report No 9/2007. European Environment Agency: Copenhagen.
report#30	EEA (2009) Ensuring quality of life in Europe's cities and towns. Tackling the environmental challenges driven by European and global change. EEA Report No 5/2009. European Environmental Agency: Copenhagen.
report#31	DG Environment (2008) Modelling of EU Land-use choices and environmental impacts. Scoping study. BIO Intelligence Services: Ivry-sur-Seine.

## Appendix 11

### Reports taken into account for text mining purposes on 'Social Affairs'

Code	Report
report#32	ESPON (2006) The role and spatial effects of cultural heritage and identity. ESPON Project 1.3.3 Final Report. Università degli Studi Ca' Foscari: Venice.
report#33	ESPON (2006) Territorial dimension of the Lisbon-Gothenburg strategy. ESPON Project 3.3 Final Report. Università degli Studi 'Tor Vergata' di Roma. CEIS: Rome.
report#34	European Parliament (2005) Adaption of Cohesion Policy to the enlarged Europe and the Lisbon and Gothenburg objectives. European Parliament: Brussels.
report#35	ESPON (2006) Preparatory study on social aspects of EU territorial development. ESPON Project 1.4.2 Final report. OIR: Vienna.
report#36	European Parliament (2007) Impact of Accession on the Labour Markets of the New Member States. European Parliament: Brussels.
report#37	European Parliament (2007) The role of minimum income for social inclusion in the European Union. European Parliament: Brussels.
report#38	European Parliament (2009) Indicators of Job Quality in the European Union. European Parliament: Brussels.
report#39	European Parliament (2010a) The link between job creation, innovation, education and training: An assessment of policies pursued at EU level. European Parliament: Brussels.
report#40	European Parliament (2010b) Structural and Cohesion Policies following the Treaty of Lisbon. European Parliament: Brussels.
report#41	European Parliament (2010c) Social Policy Agenda. Directorate-General for Internal Policies, European Parliament: Brussels.
report#42	European Parliament (2010d) Mobility and Integration of People with Disabilities into the Labour Market. European Parliament: Brussels.
report#43	European Parliament (2010e) EU Cooperation in the field of Social Inclusion. European parliament: Brussels.

## Appendix 12

### Reports taken into account for text mining purposes on 'Economy'

Code	Report
report#44	ESPON (2005) The Territorial Impact of EU Research and Development Policies. ESPON Project 2.1.2 Final Report. ECOTEC: Birmingham.
report#45	ESPON (2006) Territorial dimension of the Lisbon-Gothenburg strategy. ESPON Project 3.3 Final Report. Universita' degli Studi 'Tor Vergata' di Roma. CEIS: Rome.
report#46	ESPON (2007) Identification of spatially relevant aspects of information society. ESPON Project 1.2.3 Final Report. EUROREG: Warsaw.
report#47	ESPON (2006) Territorial impacts of EU economic policies and location of economic activities. ESPON Project 3.4.2 Final Report. IGEAT: Brussels.
report#48	CEC (2002) First progress report on economic and social cohesion. Commission of the European Communities: Brussels.
report#49	CEC (2003) Second progress report on economic and social cohesion. Commission of the European Communities: Brussels.
report#50	CEC (2005) Third progress report on cohesion: towards a new partnership for growth, jobs and cohesion. Commission of the European Communities: Brussels.
report#51	CEC (2006) The growth and jobs strategy and the reform of European cohesion policy. Fourth progress report on cohesion. Commission of the European Communities: Brussels.
report#52	CEC (2008) Fifth progress report on economic and social cohesion. Growing regions, growing Europe. Commission of the European Communities: Brussels.
report#53	CEC (2009) Sixth progress report on economic and social cohesion. Commission of the European Communities: Brussels.

## **Appendix 13**

### **Overview of the ESPON 2013 DB thematic structure by themes and sub-themes**

#### **01 AGRICULTURE AND FISHERIES**

- 0101 Farm structure (e.g. farm type, size of farms, income from farming, organic farming)
- 0102 Livestock (e.g. livestock output)
- 0103 Aquaculture and sea fisheries (e.g. aquaculture resources in coastal and marine areas)
- 0104 Forestry (e.g. production, consumption, import/export products)
- 0105 Rural characteristics (e.g. rural employment, rural access to services)

#### **02 DEMOGRAPHY**

- 0201 Population structure (e.g. age distribution by group and gender)
- 0202 Natural changes (e.g. fertility, mortality, life expectancy)
- 0203 Households (e.g. number and sizes of households)
- 0204 Migrations (e.g. immigration, migration replacement, high-skilled labour migration)

#### **03 TRANSPORT**

- 0301 Accessibility (e.g. performance indicators, multimodal accessibility)
- 0302 Flows (e.g. vehicles, passengers, goods, freight)
- 0303 Infrastructures (e.g. transportation systems, railways, airports, harbours)

#### **04 ENERGY AND ENVIRONMENT**

- 0401 Energy and resources (e.g. renewable, nuclear, and fossil energies)
- 0402 Climate change (e.g. GHG emissions, air pollution)

#### **05 LAND USE**

- 0501 Land use and land cover types (e.g. CORINE Land Cover, GMES)
- 0502 Urban land use attributes and changes (e.g. LUZ, Urban Atlas)
- 0503 Rural land use attributes and changes (e.g. Natura 2000)

#### **06 SOCIAL AFFAIRS**

- 0601 Education (e.g. training, lifelong learning)
- 0602 Labour market (e.g. labour force, labour costs, economic inactivity, earnings)
- 0603 Living conditions (e.g. poverty, social exclusion, health systems)
- 0604 Culture (e.g. socio-cultural activities, cultural consumption)

#### **07 ECONOMY**

- 0701 Aggregated accounts (e.g. GDP, balance of payments)
- 0702 Employment (e.g. employment, unemployment, long-term unemployment)
- 0703 Production and costs per sector (e.g. production of manufactured goods)
- 0704 Research and innovation (e.g. R&D expenditure, ICT research, patents, investments)

#### **99 CROSS-THEMATIC AND NON-THEMATIC DATA**

- 9901 Integrative indices, indicators and scenarios (e.g. typologies, scenarios)
- 9999 Geographical objects (e.g. administrative units, grids, networks)



Figure 1a: Density view of data co-occurrence based on TF distribution within a paragraph (left) and a window of 5 words (right)

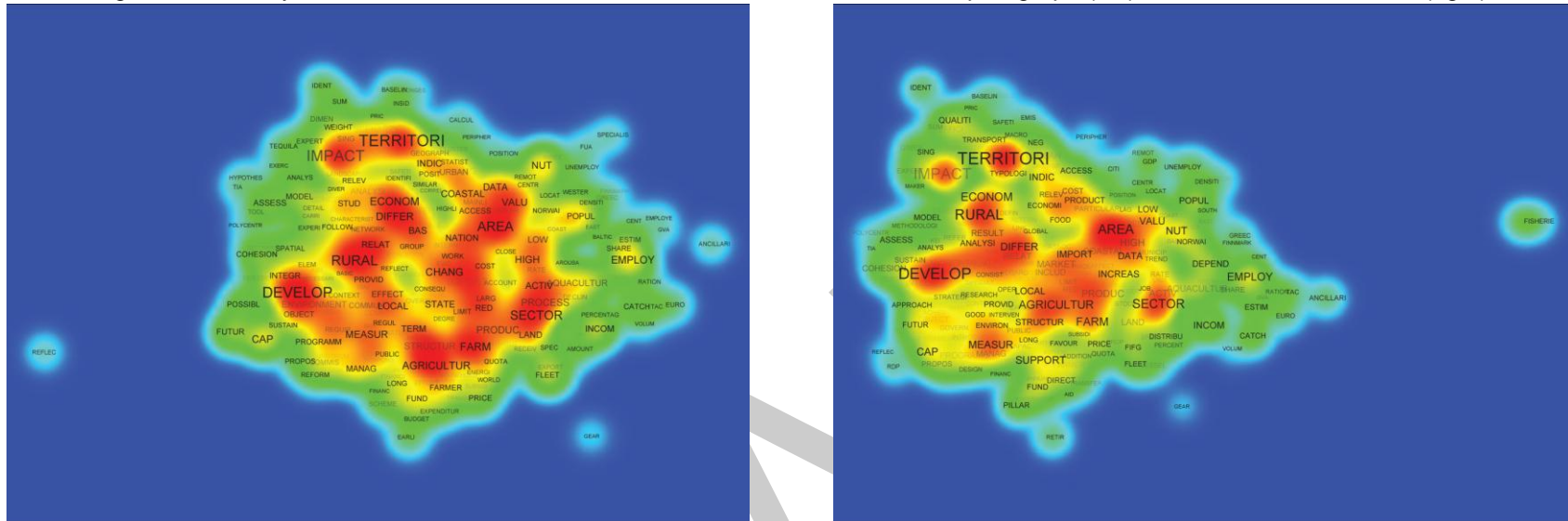
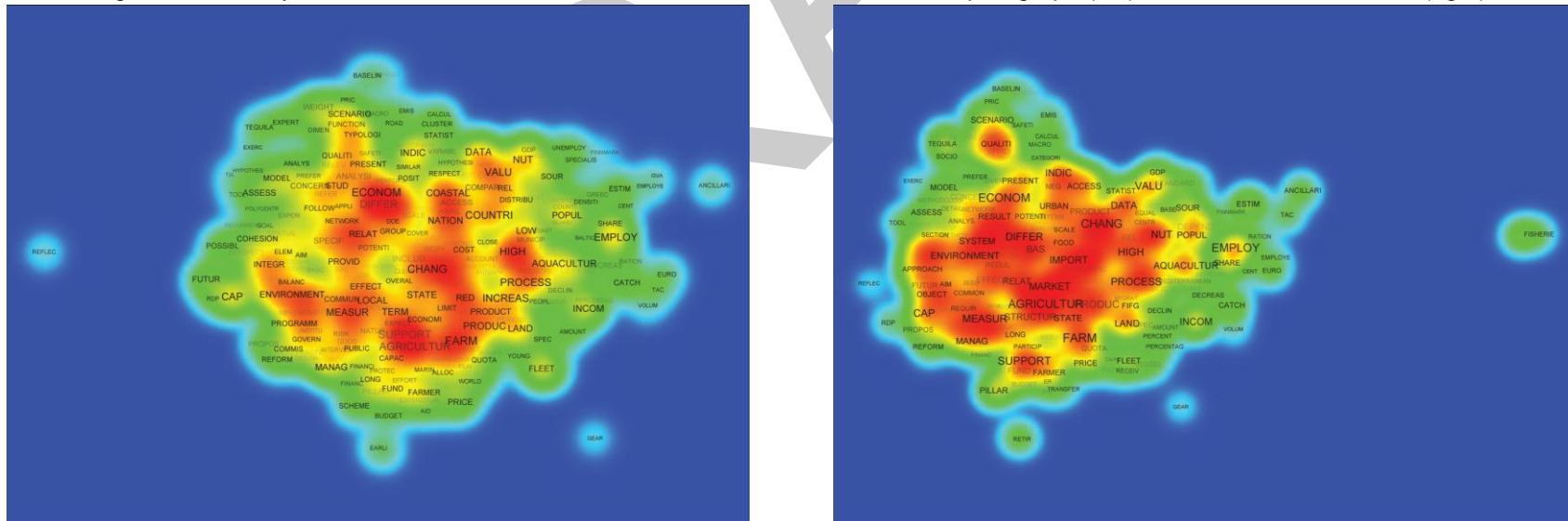


Figure 1b: Density view of data co-occurrence based on TF distribution within a paragraph (left) and a window of 5 words (right)



Note: The words 'territori', 'develop', 'impact', 'area', 'rural', and 'sector', present in (1a), were added to an exclusion list in (1b). TF: Term Frequency.



Figure 2a: Density view of data co-occurrence based on TF distribution within a paragraph (left) and a group of 5 words (right)

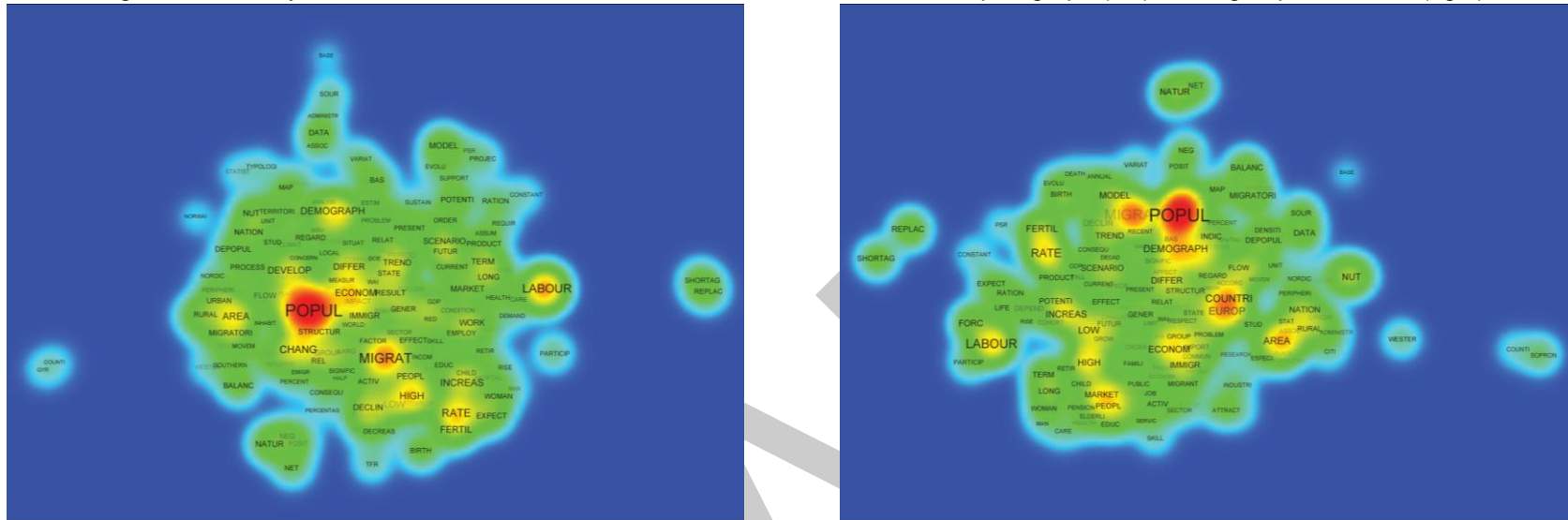
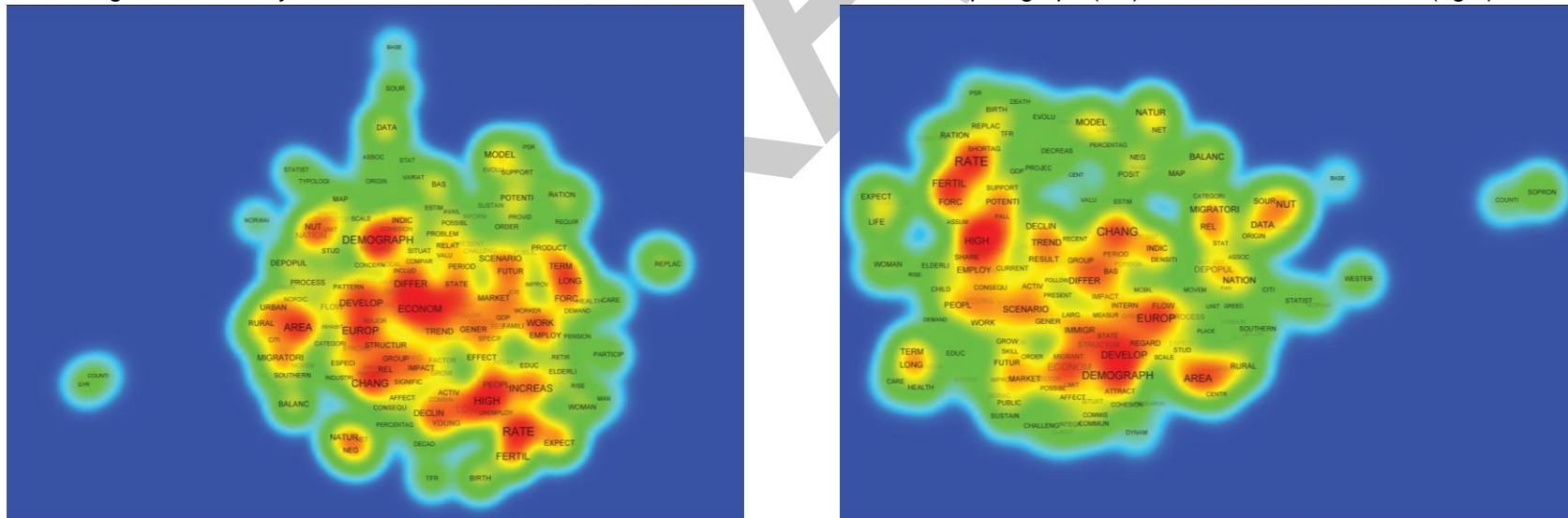


Figure 2b: Density view of data co-occurrence based on TF distribution within a paragraph (left) and a window of 5 words (right)



Note: The words 'popul', 'migrat', 'countri', 'labour', present in (2a), were added to a exclusion list in (2b). TF: Term Frequency.



Figure 3a: Density view of data co-occurrence based on TF distribution within a paragraph (left) and a group of 5 words (right)

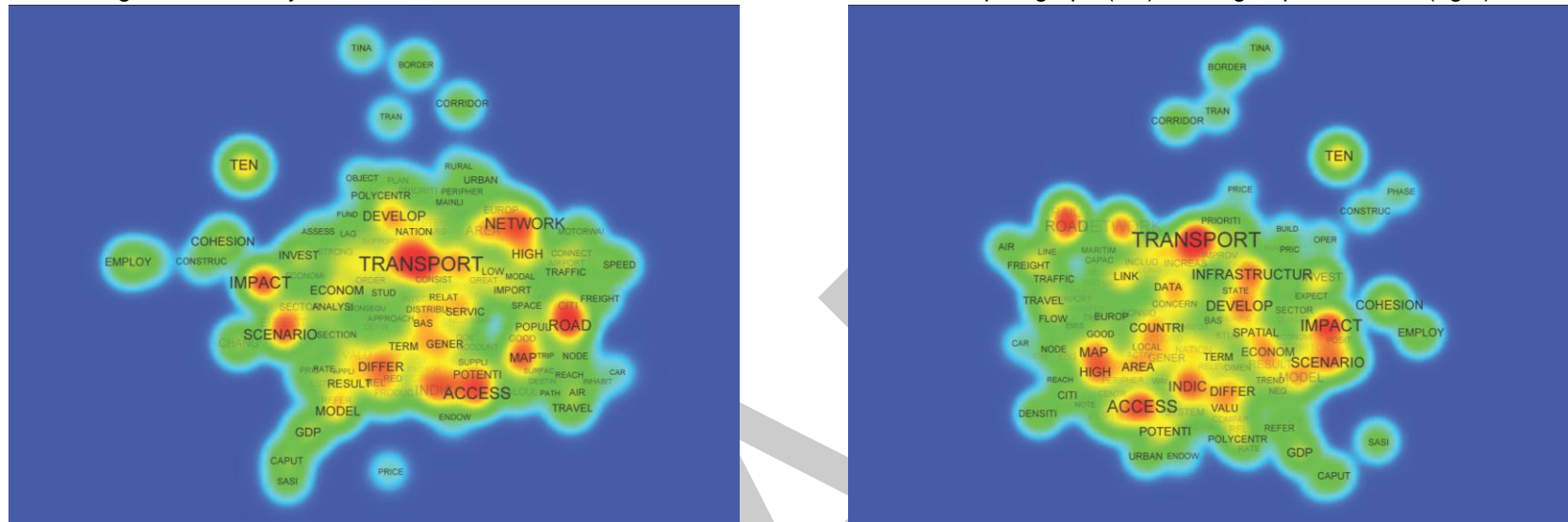
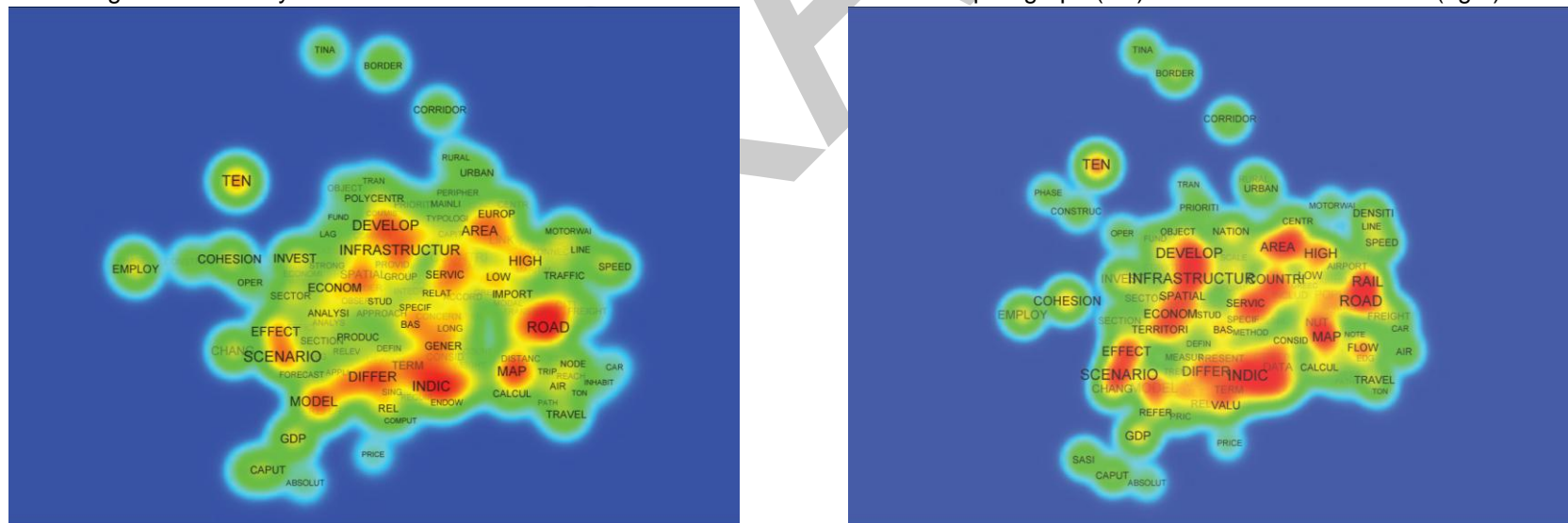


Figure 3b: Density view of data co-occurrence based on TF distribution within a paragraph (left) and a window of 5 words (right)



Note: The words 'transport', 'access', 'network', 'impact', present in (3a), were added to a exclusion list in (3b). TF: term Frequency.

Figure 3c: Density view of data co-occurrence based on TF\*IDF distribution within a paragraph (left) and a group of 5 words (right)

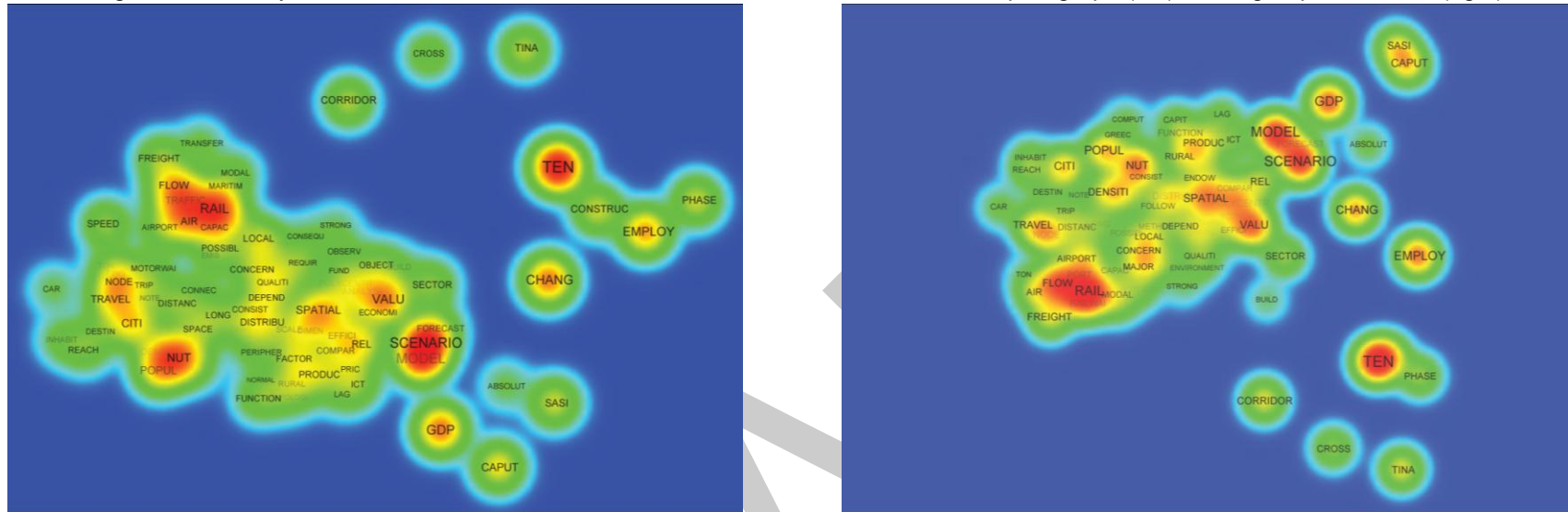
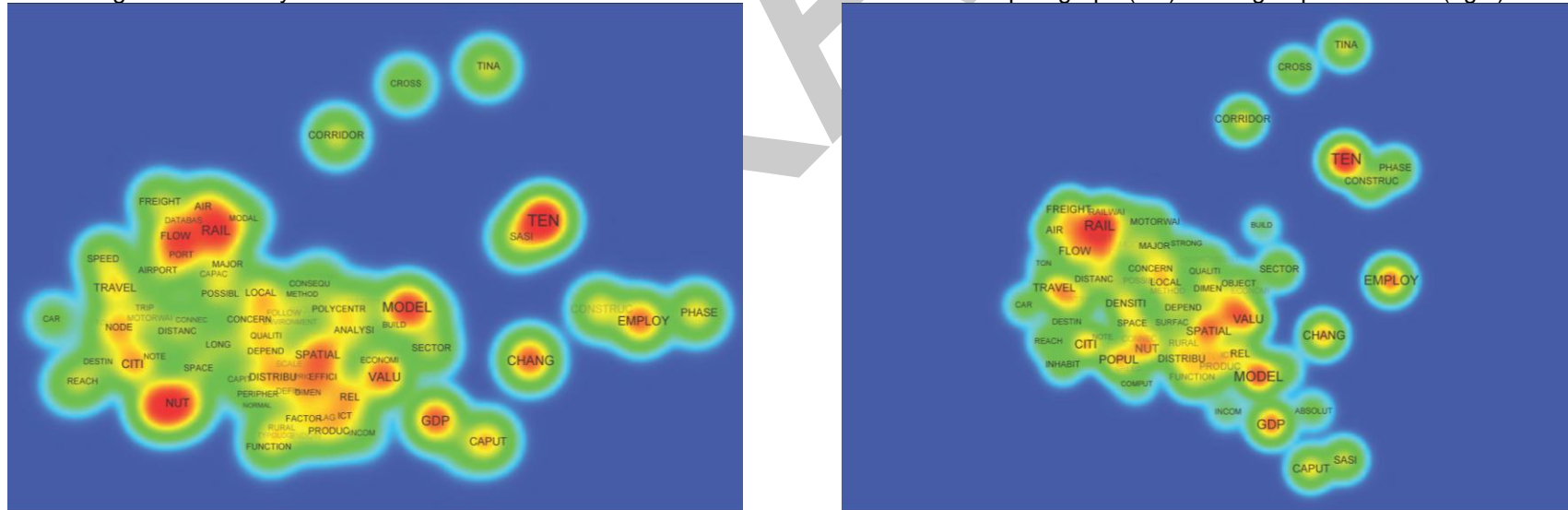


Figure 3d: Density view of data co-occurrence based on TF\*IDF distribution within a paragraph (left) and a group of 5 words (right)



Note: The word 'scenario', present in (3c), was added to a exclusion list in (3d)

Figure 4a: Density view of data co-occurrence based on TF distribution within a paragraph (left) and a group of 5 words (right)

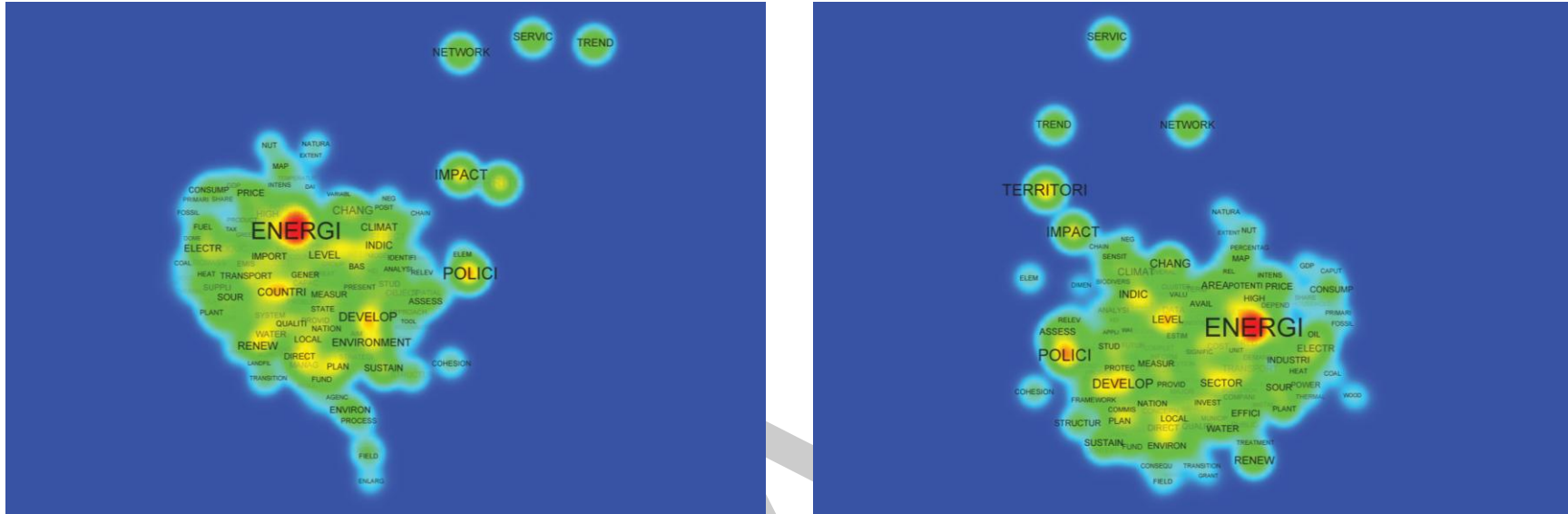
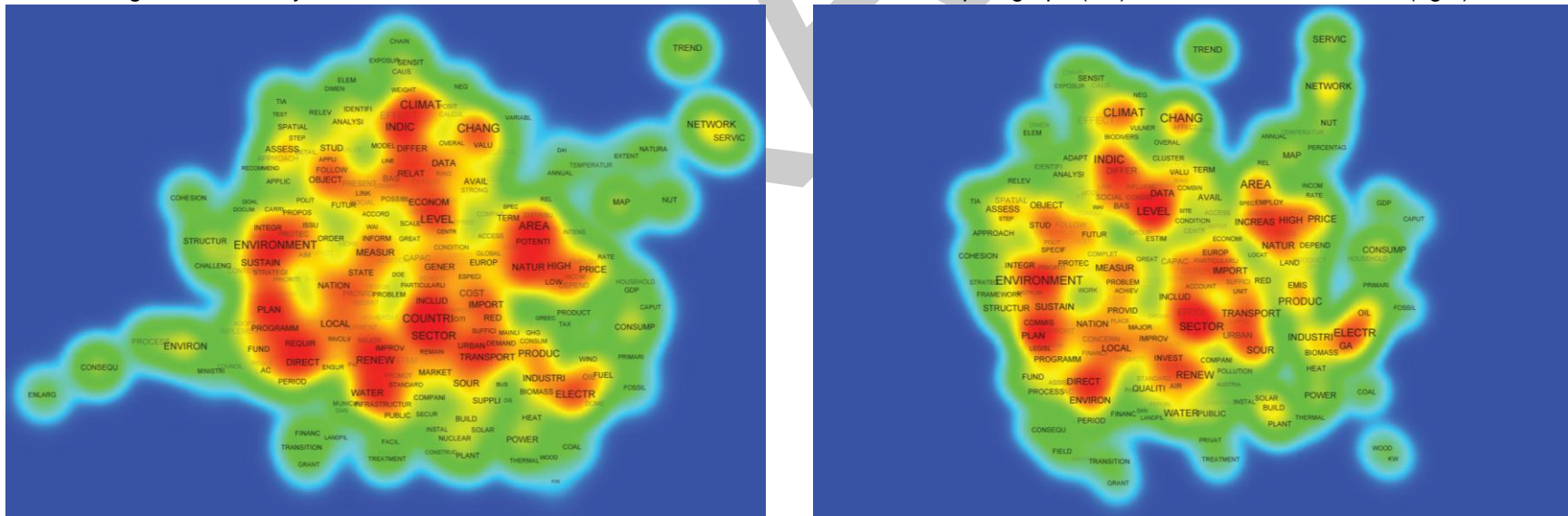


Figure 4b: Density view of data co-occurrence based on TF distribution within a paragraph (left) and a window of 5 words (right)



Note: The words 'energi', 'polici', 'territo', 'impact', and 'develop', present in (4a), were added to an exclusion list in (4c). TF: Term Frequency.

Figure 4c: Density view of data co-occurrence based on TF\*IDF distribution within a paragraph (left) and a group of 5 words (right)

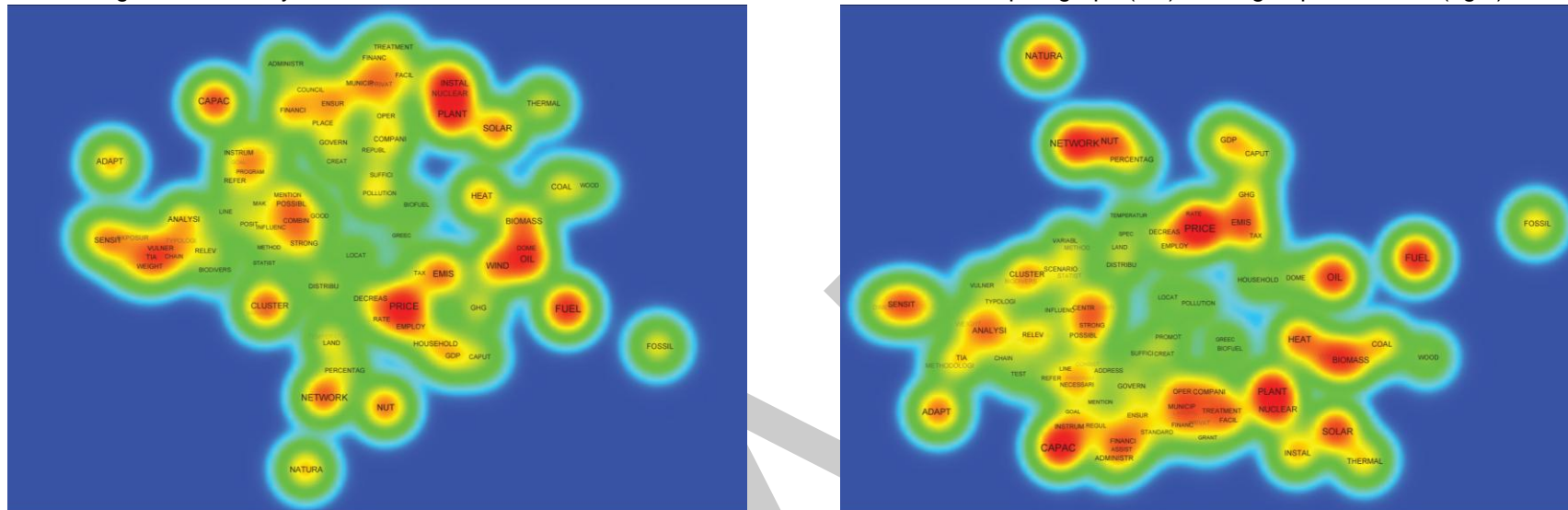
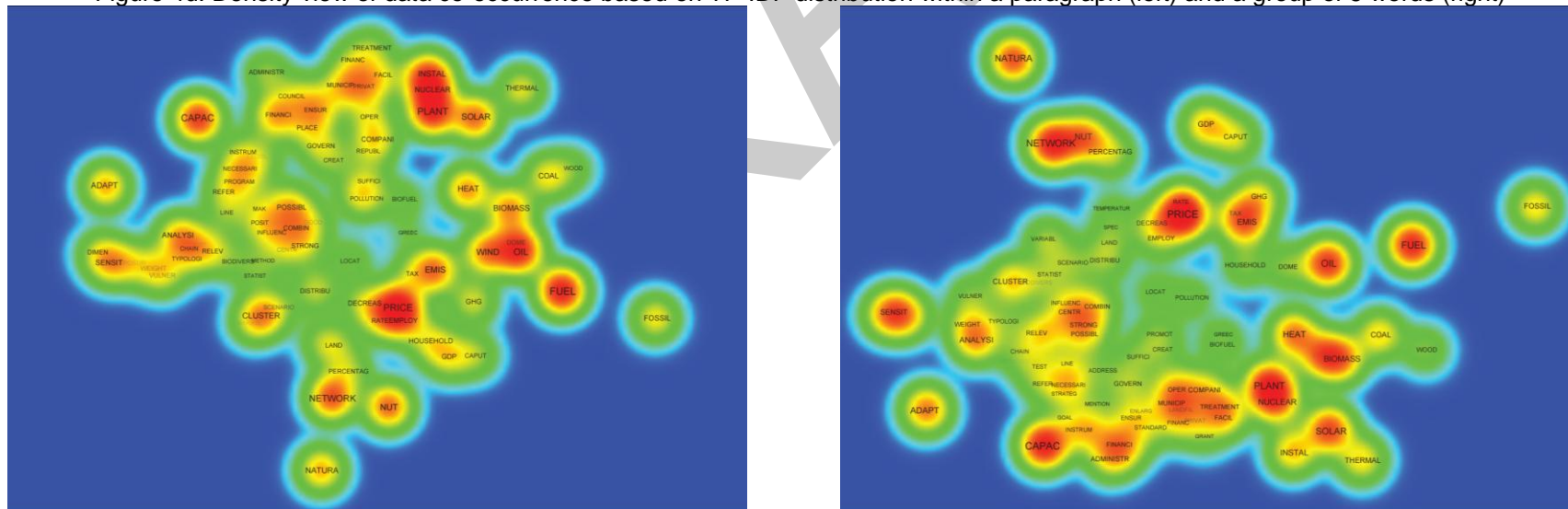


Figure 4d: Density view of data co-occurrence based on TF\*IDF distribution within a paragraph (left) and a group of 5 words (right)



Note: The word 'tia', present in (4c), was added to a exclusion list in (4d)



Figure 5a: Density view of data co-occurrence based on TF distribution within a paragraph (left) and a window of 5 words (right)

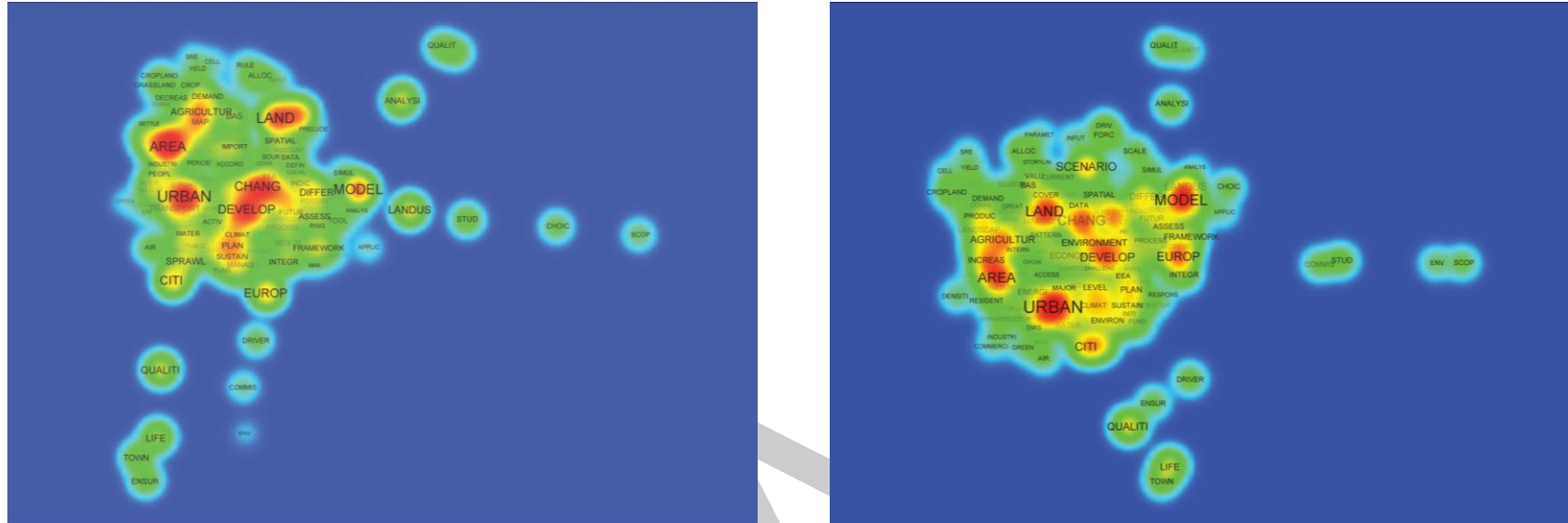
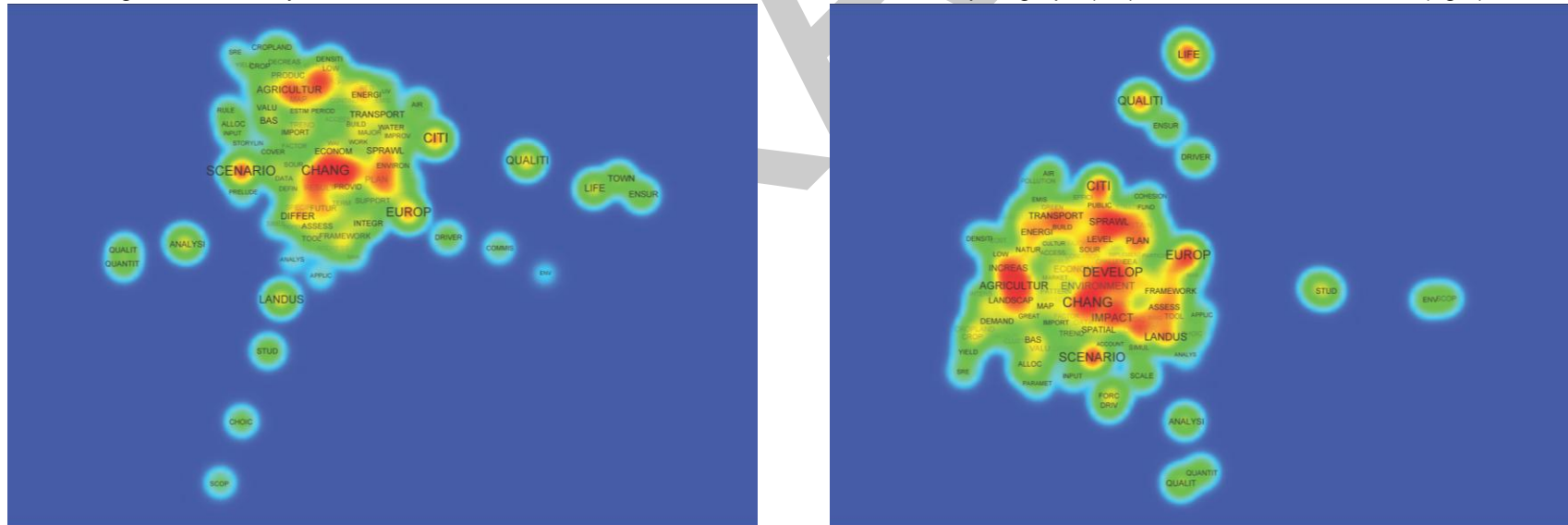


Figure 5b: Density view of data co-occurrence based on TF distribution within a paragraph (left) and a window of 5 words (right)



Note: The words 'urban', 'model', 'land', and 'area', present in (5a), were added to a exclusion list in (5b)

Figure 5c: Density view of data co-occurrence based on TF\*IDF distribution within a paragraph (left) and a group of 5 words (right)

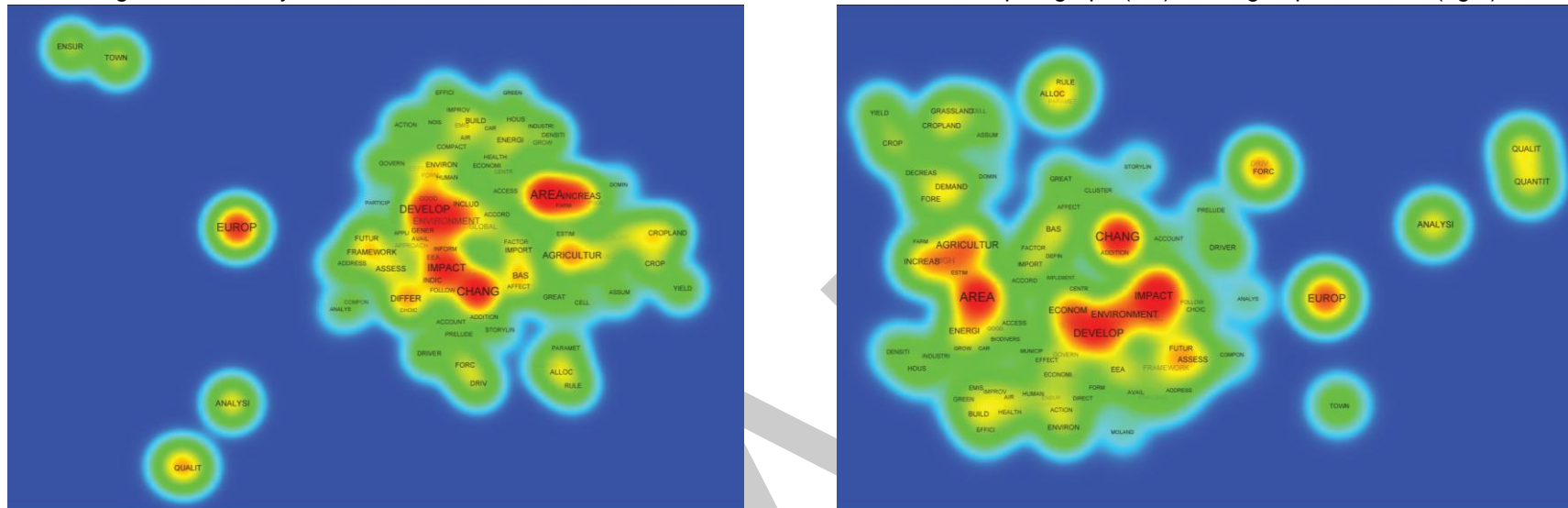
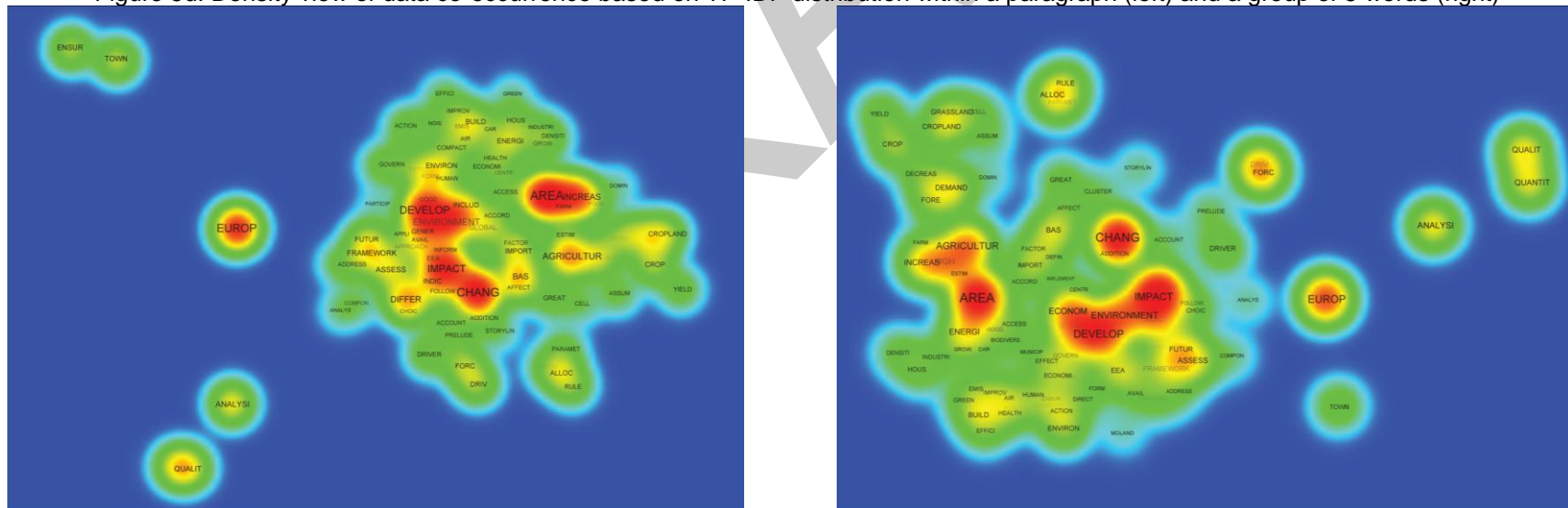


Figure 5d: Density view of data co-occurrence based on TF\*IDF distribution within a paragraph (left) and a group of 5 words (right)



Note: The word 'scenario', present in (5c), was added to a exclusion list in (5d)

Figure 6a: Density view of data co-occurrence based on TF distribution within a paragraph (left) and a group of 5 words (right)

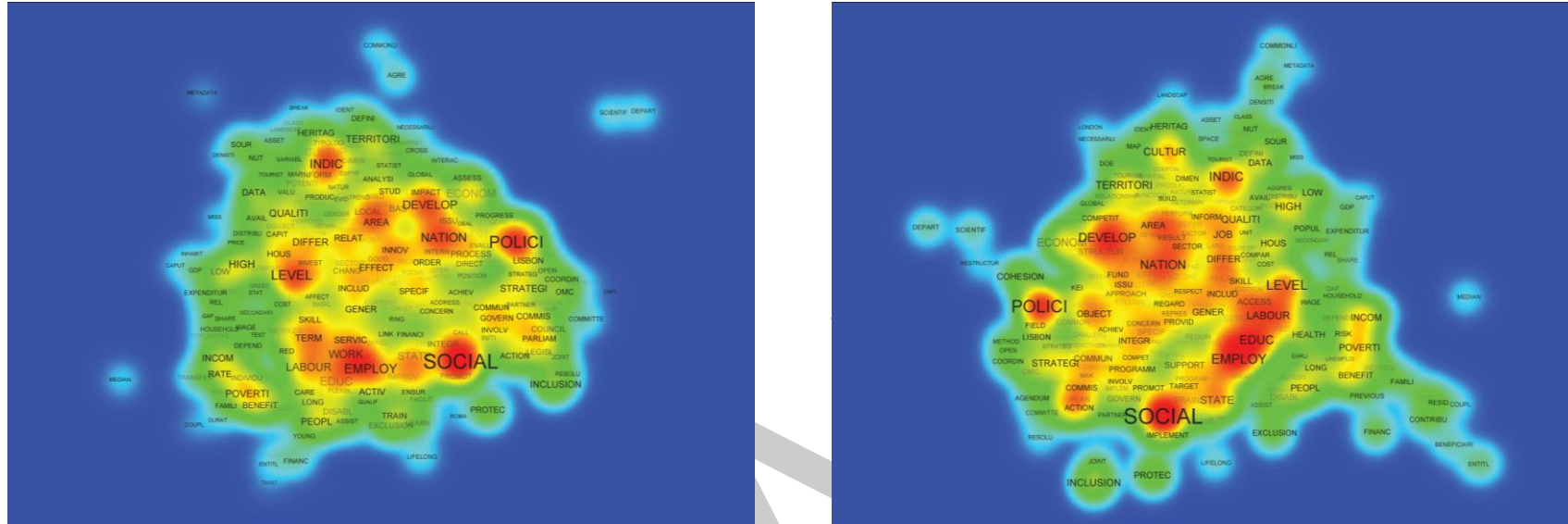
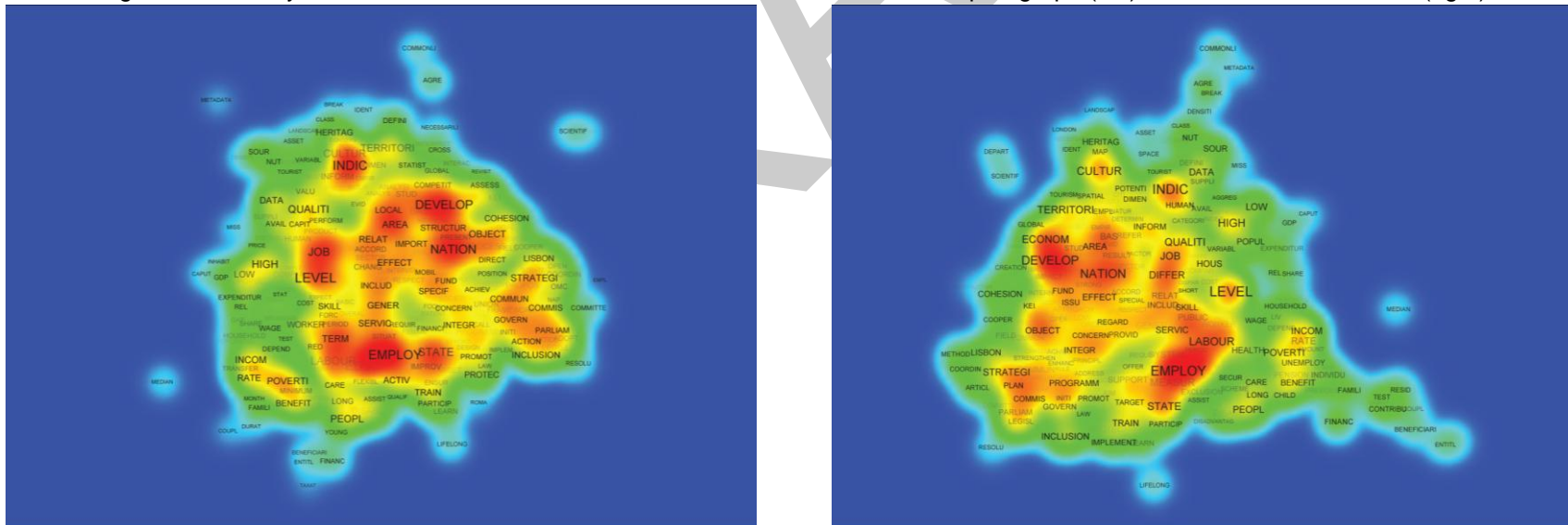


Figure 6b: Density view of data co-occurrence based on TF distribution within a paragraph (left) and a window of 5 words (right)



Note: The words 'socio', 'polici', present in (6a), were added to a exclusion list in (6b). TF: Term Frequency.

Figure 6c: Density view of data co-occurrence based on TF\*IDF distribution within a paragraph (left) and a group of 5 words (right)

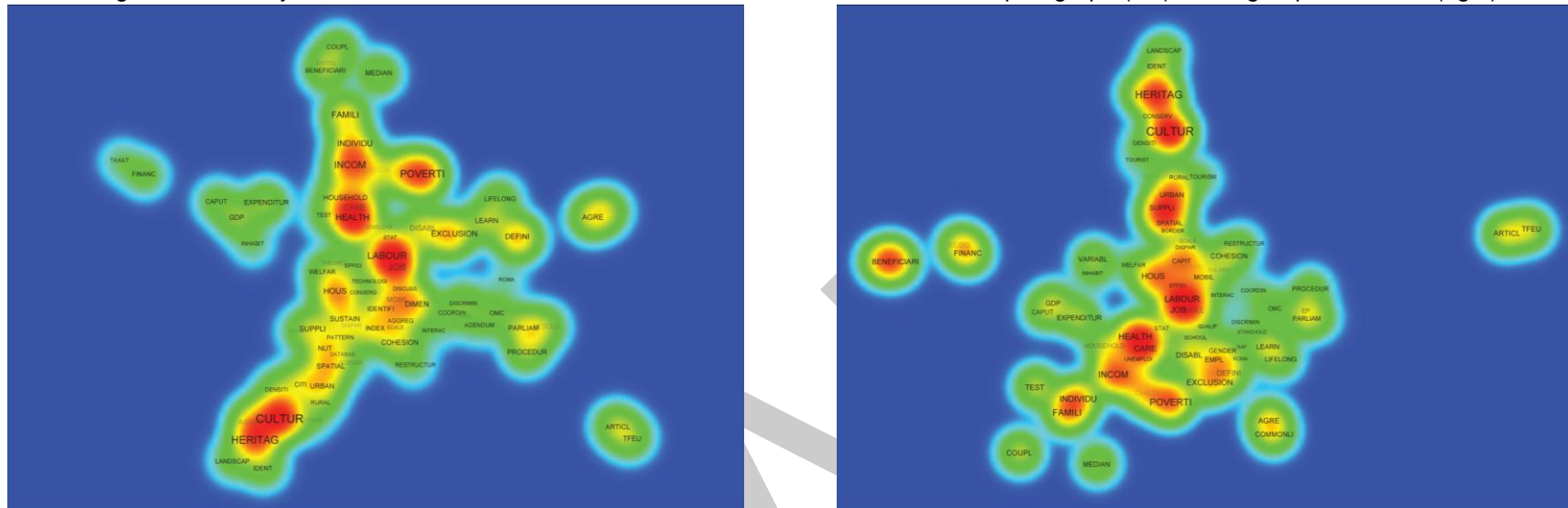
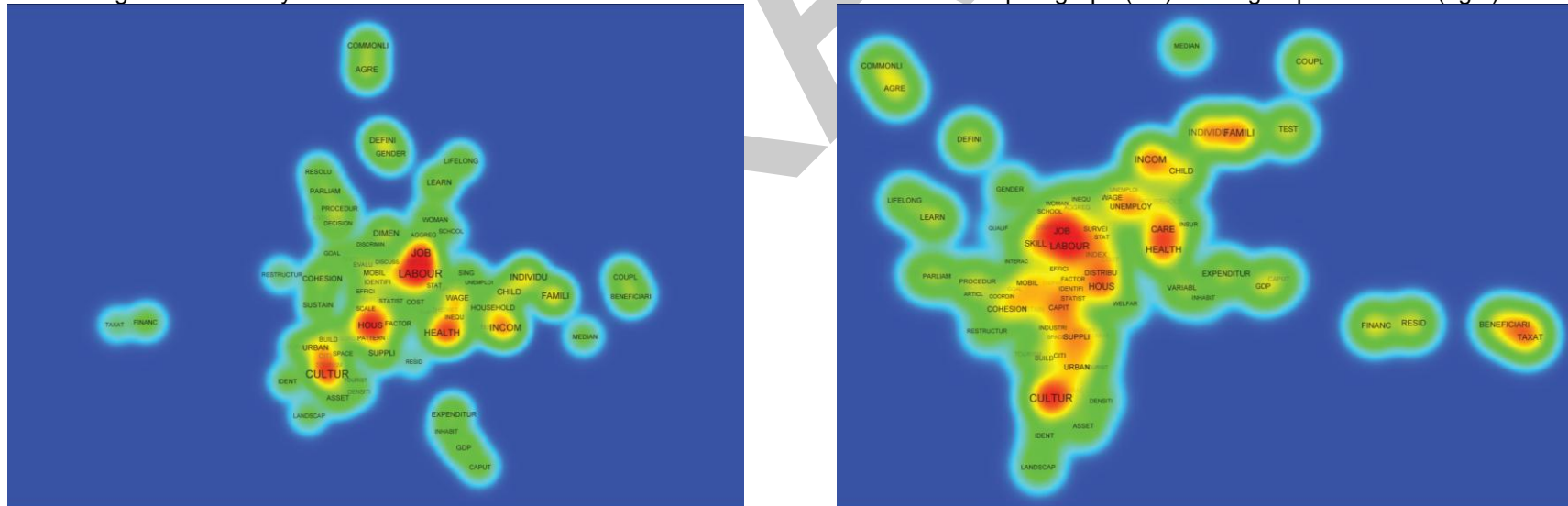


Figure 6d: Density view of data co-occurrence based on TF\*IDF distribution within a paragraph (left) and a group of 5 words (right)



Note: The words 'heritag', 'empl', 'spatial', 'pension', 'disabl', and 'poverti', present in (6c), was added to a exclusion list in (6d)

Figure 7a: Density view of data co-occurrence based on TF distribution within a paragraph (left) and a group of 5 words (right)

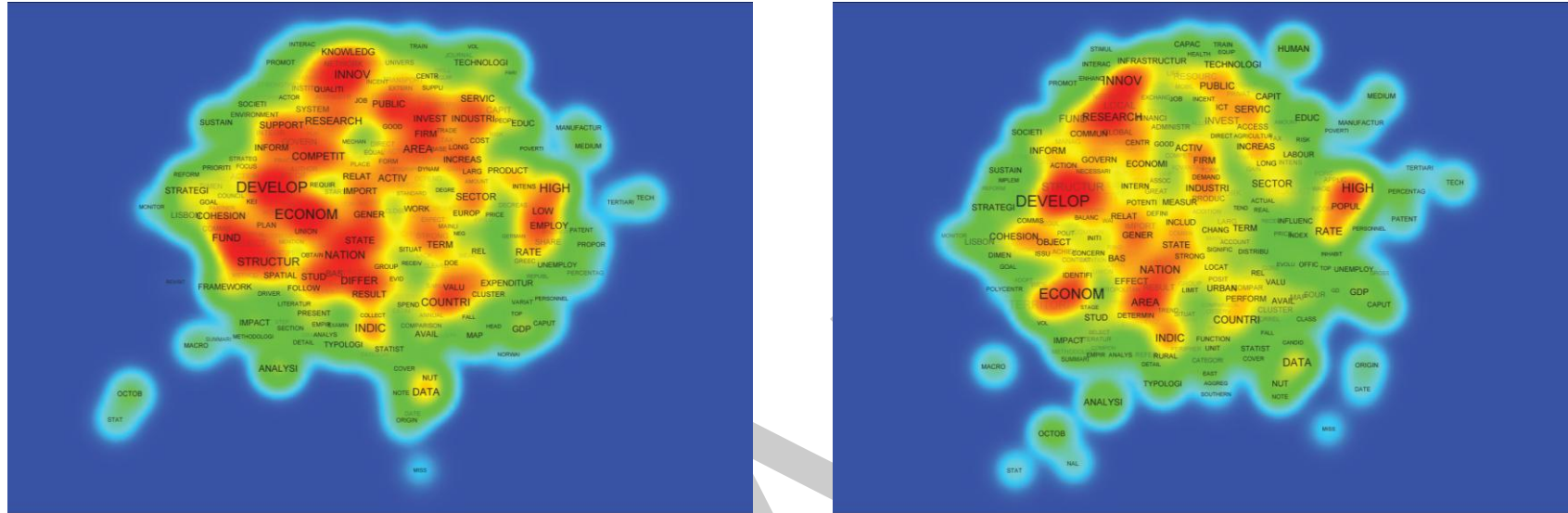
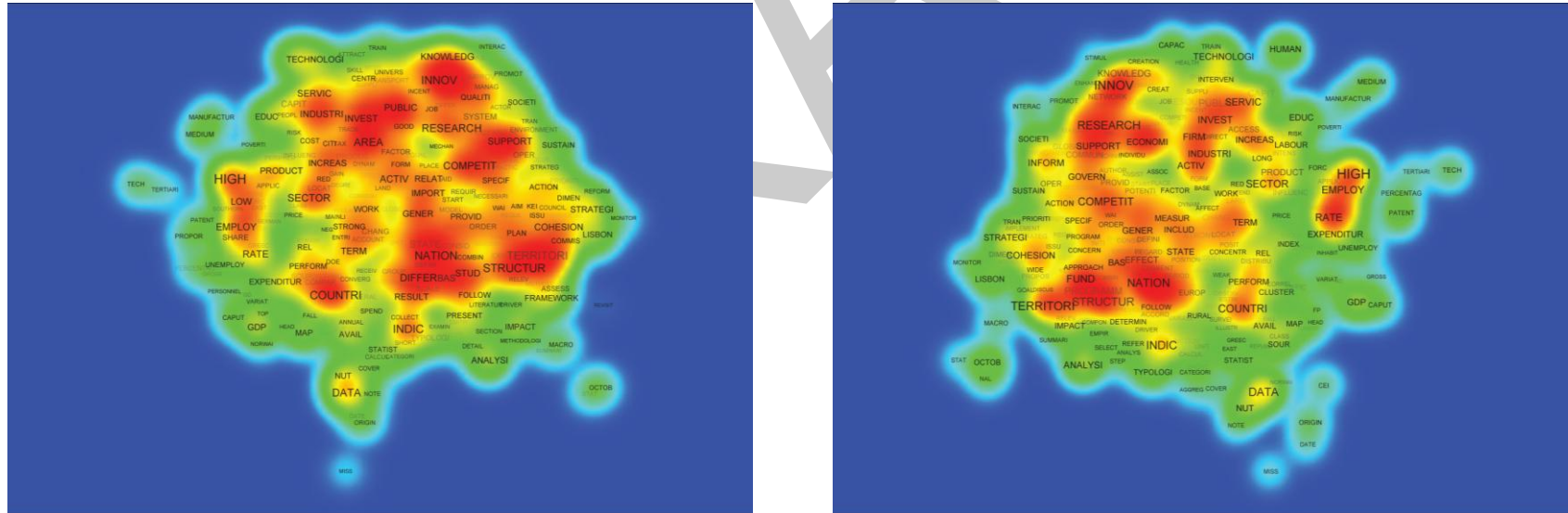


Figure 7b: Density view of data co-occurrence based on TF distribution within a paragraph (left) and a window of 5 words (right)



Note: The words 'develop', 'econom', present in (7a), were added to a exclusion list in (7b). TF: Term Frequency.

Figure 7c: Density view of data co-occurrence based on TF\*IDF distribution within a paragraph (left) and a group of 5 words (right)

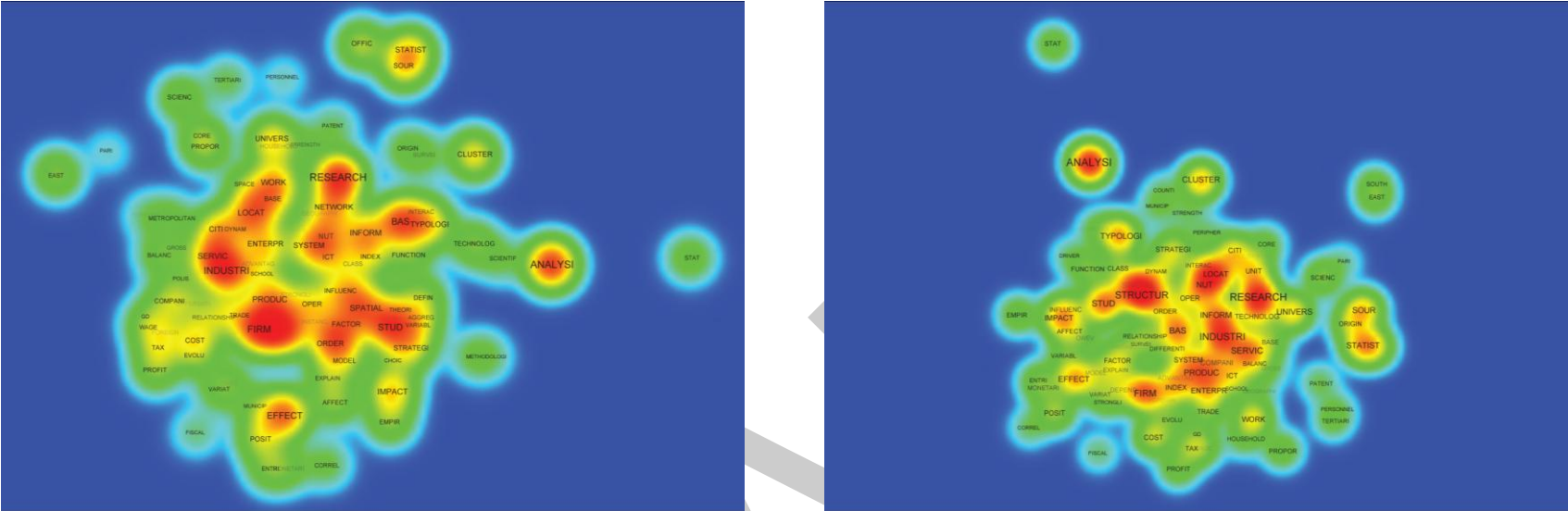
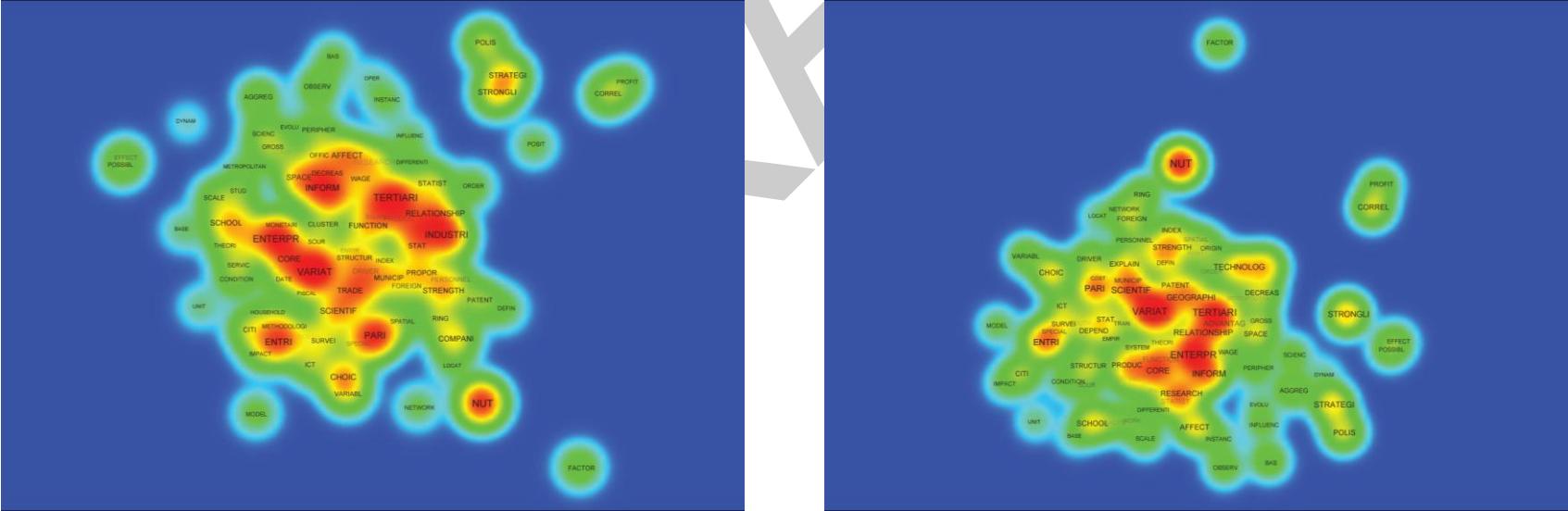
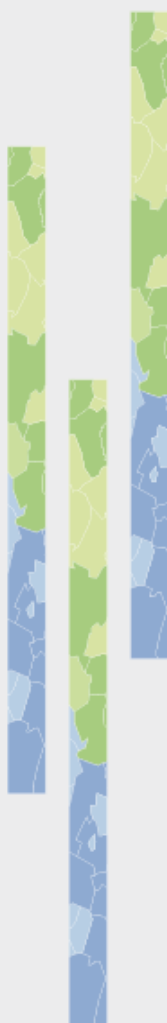
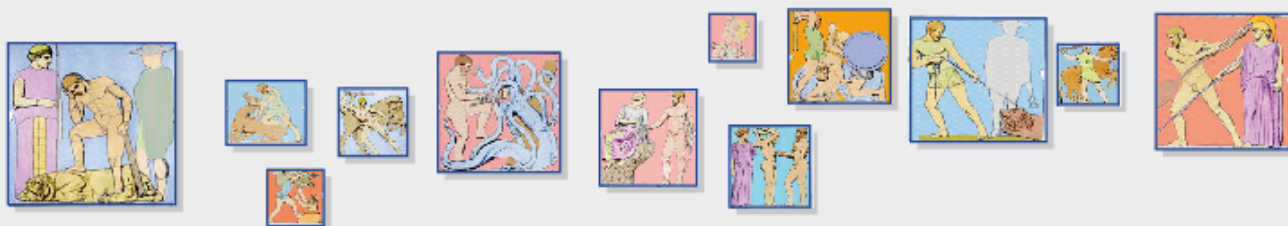


Figure 7d: Density view of data co-occurrence based on TF\*IDF distribution within a paragraph (left) and a window of 5 words (right)



Note: The words 'typologi', 'firm', 'tax', 'counti', and 'gdp', present in (7c), were added to a exclusion list in (7d)

DRAFT



# Coding scheme to label indicators: Guiding principles to TPGs involved in ongoing and future ESPON projects

## MAIN RESULTS

- The rationale defined by each TPG to label indicators and indices increases the degree of ambiguity
- Develop an harmonised scheme to code ESPON indicators and indices
- Illustrative examples demonstrate the usefulness of such approach

ESPON 2013 DATABASE

MARCH 2011





## **LIST OF AUTHORS**

Nuno Madeira, University of Luxembourg

Geoffrey Caruso, University of Luxembourg

### **Contact**

E-mail: [nuno.madeira@uni.lu](mailto:nuno.madeira@uni.lu); Tel. +352 46 66 44 9691

E-mail: [geoffrey.caruso@uni.lu](mailto:geoffrey.caruso@uni.lu); Tel. +352 46 66 44 6625

## Tables of contents

1. INTRODUCTION .....	4
2. CODING SCHEME TO LABEL INDICATORS .....	4
3. ILLUSTRATIVE EXAMPLES.....	6
4. CONCLUSIONS .....	7
REFERENCES .....	7

## 1. Introduction

This report provides background information and guidelines to those responsible for the delivery of ESPON data. One of the issues that emerged during the implementation of the ESPON 2013 DB concerns the rationale defined by Transnational Project Groups (TPGs) to label variables. Despite the usefulness of the different methods applied by TPGs, the fact is that these methods increase the degree of ambiguity within the ESPON 2013 DB. The implications are relevant, because if no harmonisation is applied it will invariably lead to incoherence.

Therefore, the aim of the present document is to provide guidelines on how to create codes for datasets that will be integrated in the database. Taking into consideration that different TPGs may have to deliver the same variable, it is then essential to apply a harmonised and identical code for those that share the same designation. In a way, this would bring some coherence to the database.

## 2. Coding scheme to label indicators

### *Prior considerations*

Naming indicators is an important component of indicator development. Therefore, TPGs should strive to be objective and consistent. Taking into consideration the ultimate list of ESPON indicators we noticed a wide variation of naming conventions that differ according to the criterion defined by each research team. Indeed, some applied research projects have chosen to use more simplistic names that capture the essence of what is being measured while others tend to be very descriptive. This same list discloses, however, some similar labels that essentially reveal different methodologies, data providers and, most likely, implications for the variable that is being classified as such.

In this section we introduce a coding scheme to label ESPON indicators and indices. Its purpose is to assemble relevant information about data using a minimum number of characters. We believe that it is impossible to fully harmonize a coding scheme given the variety of indicators delivered up to date. However, some harmonisation should be adopted.

With this regard, those responsible for the delivery of datasets within a TPG should first browse the list of acronyms and abbreviations proposed to identify variables (see Annex 1). In the event that a specific variable already integrates the above mentioned list, TPGs are encouraged to use the existing code. If not, TPGs should simply create a new code according to the scheme described in these guidelines.

Therefore, we invite TPGs to apply a coding scheme that comprises three fields or divisions. The information to be included in each division is based on the current state of the database and, essentially, tries to focus on the indicator itself, i.e. subject, nature of data, and calculation method(s). Other elements that might be used to label data will not be considered as they are already mentioned on metadata (i.e. time, space, data provider).

### *Procedure*

Each field or division that we consider for the coding scheme is accompanied by a list of abbreviations and/or acronyms that best identify those above-mentioned categories. The first two lists will certainly be adapted and enriched over time as the database integrates new data. For this purpose, we provide some abbreviations widely used to label indicators (see Annex 2). The third list should preferably remain fixed since it corresponds to measurement scales as recognised in the geographical and/or statistical literature (see

Annex 3). No maximum number of characters will be applied in whatever division. However, we recommend metadata developers to keep the code as short as possible.

The process for structuring the coding scheme is the following:

- 1) Start with three lower case letters best identifying the subject. If the main subject is described as an acronym use upper case letters. If possible, choose those three letters from the list of acronyms and abbreviation provided as an annex to this document. Try to use only three letters. However, some flexibility is allowed, especially for composed indicators.
- 2) Refine the subject using at least one character to identify derivations or restrictions when measuring the indicator. The goal here is to add items that could give more detailed information about the variable that is being measured;
- 3) Conclude the coding scheme by proving some information on the levels of measurement. For this purpose, we invite metadata developers of each TPG to use the list of abbreviations that relate to the nature of data, preferably the list of abbreviations for measurement scales.

### *Guiding principles*

The procedure is not constrained to a limit of characters. This means that those who have the responsibility to deliver data will have flexibility enough to code indicators. However, it is important to stick to the above-mentioned structure and, more importantly, use the list provided as an annex to these guidelines. Obviously, this list is not exhaustive but rather based on the current state of the database. Moreover, we should be well aware that in some cases adaptations will be necessary, especially to obtain more degree of freedom when facing rather complex, but similar, indicators.

The list of acronyms and abbreviations we recommend in Annex 1 for potential subjects is rather straightforward, and usually tries to capture the first three consonants of a word, or even other letters representing at best the core subject of a variable.

The second part of the list refers to widely used abbreviations that impose restrictions and/or use derivations. A typical example in demographic data is the distinction of gender by "f" for female and "m" for male. Other abbreviations refer to commonly used terms to describe a variable such as *index*, *rate*, or *change*. Sometimes it directly relates to the nature of data (particularly terms such as *volume*, *absolute*, *relative*), though it is only loosely related to the measure itself.

To what regards the third part of the coding scheme, we believe that it should contain an unambiguous description of the level of measurement (expression coined by Stevens, 1946) since data needs to be described as accurately as possible. Levels of measurement are particularly important because it allows the user to draw a direct relationship between data classification and the cartographic representation of data, as well as to capture its use within mathematical operations. Together with metadata, it is an important feature that could lead in the future to an automatic, clever data management system. In literature, four levels of measurement are commonly distinguished following the proposals made by Stevens (1946) on the theory of scales and measurements. In ascending order of precision these are: nominal, ordinal, interval, and ratio data. Statistical analysis and most of the spatial analysis handbooks refer to those four scales (Chrisman, 2002; Haining, 2003; Slocum et al., 2005, Haining, 2010).

At **nominal** level of measurement, the numbers are used to classify data. On the contrary, the **ordinal** scale illustrates some ordered or ranked relationship between categories. Despite the fact that both levels correspond to categorical data the major difference between them lies on the hierarchical, non-sequential relationship. For instance, land use types are classified as nominal data while income categories are explicitly ordinal.

The **interval** and **ratio** scales are quantitative data (i.e. numerical measures). The interval level has equal units of measurement, thus making it possible to interpret not only the order of scale but also the distance between them. Nevertheless, the zero point of an interval scale is arbitrary and is not a true zero. The ratio scale of measurement has fixed origin or zero point. This is the most advanced scale. Most of common statistical methods of analysis however require only interval level of measurement.

Geographers, particularly those involved in GIS and cartography argue that scales should be refined when dealing with geographic data (Forrest, 1999; Chrisman, 2002; Slocum et al., 2005). We choose to follow the naming and definition of measurement scales proposed by Forrest (1999). In addition, Chrisman (2002, 31-33) and Slocum et al. (2005, 60-61) also discuss the rationale for these subdivisions.

Against this background, the above-mentioned authors state that nominal data are divided up into four categories: *unique* (i.e. no duplicated value), *dichotomous* (i.e. binary data), *categorical* and *categories with graded membership*; ordinal data are subdivided into *complete* and *classed ordinal* depending on whether all values are unique or not and, finally, ratio data are subdivided into six subtypes. The first two are often referred to as *volumes* of *absolute numbers* in cartographic literature and mapped with proportional symbols: *extensive ratio* (where additive properties apply) and *count* (number of something). Then follows those ratios that are mapped using choropleth maps and often referred to as *relative data*: *derived ratio* (resulting from the division of any quantity by another), *density ratio* (the denominator is a geographical surface), and *constrained ratio* (values bounded between 0 and 1, representing proportions or probabilities). The last subtype, less in use within the EU territorial development research field, is the *cyclic* ratio (e.g. angles).

The corresponding list acronyms and abbreviations are displayed in both Annex 1 and 2 and a short description with examples for each level of measurement is provided in Annex 3. Whenever possible, TPGs should use the most descriptive abbreviation.

### 3. Illustrative examples

We now apply the coding scheme on several indicators to demonstrate the usefulness of this approach. For this purpose, we have randomly selected a set of indicators that show distinct features when are subject of conversion by the coding scheme. Table 1.1 illustrates the result of this method that wishes to capture, as much as possible, the content of each indicator. First we add the subject(s) attributed to each indicator that is being measured. Secondly, we determine the most characteristic derivations or restrictions induced by those indicators and, lastly, we determine the level of measurement as accurately as possible. In order to differentiate each category we suggest metadata developers to use underscore or low line symbol. If needed, TPGs should use dots to separate items that belong to the same category or division<sup>1</sup>. For age intervals, we recommend dashes.

(a)

Subject(s)							Derivations / Restrictions					Level of measurement									
m	i	g	.	p	o	p	-	c	h	.	t				-	r	t	c			

(b)

<sup>1</sup> The option for using dots as a separator is to reduce potential misunderstandings when importing variables to files with the CSV extension.

Subject(s)							Derivations / Restrictions	Level of measurement													
a	c	c	.	a	i	r	-	a	b	s					-	r	t	e			

(c)

Subject(s)							Derivations / Restrictions	Level of measurement												
e	d	u	.	s	c	d	-	t						-	r	t	c			

(d)

Subject(s)							Derivations / Restrictions	Level of measurement													
p	o	p					-	2	0	-	3	9	.	t	-	r	t	c			

(e)

Subject(s)							Derivations / Restrictions	Level of measurement												
C	O	2	.	r	o	d	-	v	o	l				-	r	t	e			

Table 1.1 Some illustrative examples of ESPON indicators that have applied the coding scheme, where (a) reflects "Migratory population change", (b) "Potential accessibility by air [absolute level]", (c) "Number of persons with secondary education degree", (d) "Population aged 20-39 years", and (e) "CO2 emissions by road traffic".

## 4. Conclusions

The result is rather helpful and easy to comprehend. However, these guidelines should be seen as an attempt to harmonise coding schemes. Most likely, additional improvements will be needed to further increase the quality of this proposal. At this point, is not possible to foresee or describe many of the indicators that will come out from the current and future applied research projects. This will require the involvement of the ESPON research community through a continuous, dynamic process.

## References

- Chrisman, N. (2002) Exploring Geographic Information Systems. New York: Wiley.
- Forrest, D. (1999) Geographic information. Its nature, classification, and cartographic representation. *Cartographica*, 36(2): 31-53.
- Haining, R. (2003) Spatial Data Analysis. Theory and Practice. Cambridge: Cambridge University Press.
- Haining, R. (2010) The Nature of Georeferenced Data. In Fisher M. M. & A. Getis (eds.) *Handbook of Applied Spatial Analysis. Software Tools, Methods and Applications*. Springer-Verlag: Berlin, pp. 197-217.
- Slocum et al. (2005) Thematic Cartography and Geographic Visualisation. New Jersey: Prentice Hall Series in Geographic Information Science.
- Stevens, s. (1946) On the theory of scales of measurement. *Science*, 103(2684): 677-680.

## Annex 1: Non-exhaustive list of acronyms and abbreviations

The list of acronyms and abbreviations compiles some of the subject-matter terms that have been identified in ESPON indicators. As stated elsewhere, the goal is to have some consistency on metadata structures and therefore facilitate information-sharing. For this purpose, we recommend TPGs to apply this list whenever feasible on the coding scheme in order to facilitate communication and understanding. The level of coverage is limited to data delivered up to June 2010. However, regular updates will inform TPGs about codes that have been used to harmonise terminology and coding schemes.

### A

Access(ibility)	acc
Administration	adm
Airports	air
Arable (land)	arb

### B

Birth	bth
Border(s)	brd
Broadband access (to Internet)	bro

### C

Carbon dioxide	CO2
Cinema(s)	cnm
City	cty
Clean	cln
Coastal	cst
Community	com
Component(s)	cmp
Construction	con
Core city	cc
Cultural	clt

### D

Death	dth
Density	den
Dependency	dep
Desiderability	dsb
Diversity	div
Doctor(s)	drs

### E

Easy-to-find-a-job	job
Easy-to-find-good-housing	hou
Economic old age dependency	eod
Economic	eco
Education	edu
Efficiency	eff
Electricity (Gas and Water Supply)	gas
Emigration	emi
Employment	emp
Energy	ene
Environment(al)	env
European Union	EU
Evaporation	eva

### F

Facilities	fcl
Farm(ers)	frm
Fertility	fer
Firm(s)	fir
Force	frc
Foreigner(s)	fgn
Frost	frt
Freight	frg

**G**

Greenspace	grs
Gross Domestic Product	GDP
Growth	gwh

**H**

Heritage	hrt
High and medium tech manufacturing activities	htc
Holdings	hld
Hospitals	hst
Higher Education Institutions	hei
Hotels and Restaurants	hot
Household(s)	hsd

**I**

Identity	idn
Immigration	imi
Industry and or/industrial	ind
Internet	itn
Island	isl
Infrastructure	inf

**L**

Labour	lab
Landscape	lds
Large Urban Zone	LUZ
Life Expectancy	lif
Life-long learning	lll

**M**

Manufacturing	man
Market	mrk
Metropolitan	met
Migration	mig
Mining (and Quarrying)	min
Mortality	mor
Mortality	mrt
Mountain(ous)	mtn

**N**

Natural grassland, heathland	grl
Natural	nat
Neighbourhood	nbh
Net extra-Europe migration	nee
Net inter-country migration	nie
Net internal Migration	nim
Noise	nse
Maritime	mar
Nomenclature des unités territoriales statistiques	NUTS

**O**

Old Age Dependency	oad
--------------------	-----

**P**

Pastures and mosaic farmland	ptr
Pollution	pol
Population	pop
Potential (impact)	pim
Precipitation	prc
Primary	prm
Public	pub
Purchasing power parity	ppp
Private	pvt



Purchasing power standards pps

## **Q**

Quality qua

## **R**

Rail ral

Rainfall rfl

Real State (renting and business activities) res

Replacement rpl

Residential (areas) rsd

Resources res

Retail trade (and wholesale) ret

Road rod

Rural rur

## **S**

Safe(ty) sfe

Schools sch

Service(s) ser

Secondary sec

(Post) Secondary pos

(Upper) Secondary ups

Student(s) stu

Sensitivity sns

Sparsely spr

Sport spt

Summer smm

## **T**

Temperature tmp

Territorial (Impact) tim

Tertiary trt

Tourism tou

Transport (and Storage) tra

## **U**

Unemployment ump

Urban urb

## **V**

Vegetation veg

Very old age dependency vod

Vulnerability vln

## **W**

Website web

Working age population wap

## **Y**

Young age dependency yng

## Annex 2: Non-exhaustive list of derivations and/or restrictions

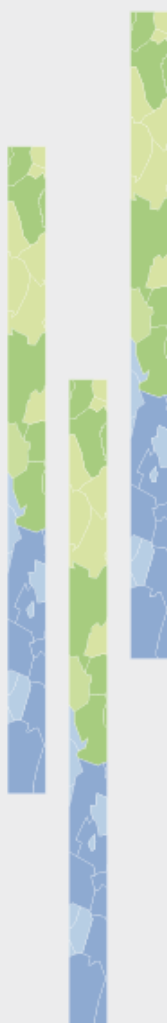
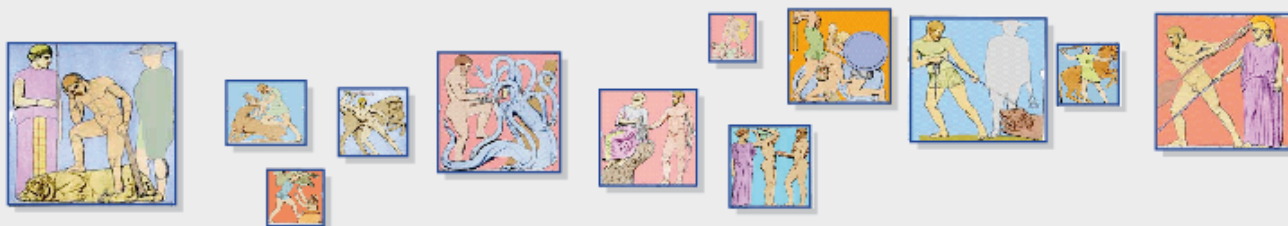
<i>Term</i>	<i>Recommended code</i>	<i>Comments</i>
Absolute	abs	The absolute value of an indicator implies a real number. This means that its value is closely related to the notions of magnitude, distance or norm and therefore is always greater than or equal to 0. An illustrative example for absolute values is the number of births.
Average	av	The value represents a central tendency. An easy way to obtain this value is to calculate the sum and then divide by the number of occurrences. There are many different descriptive statistics that can be chosen as measurement of the central tendency, including arithmetic mean, median distribution, and mode.
Change	ch	This value explains a temporal change over a period of time. In a way, it can be interpreted as a pattern that allows us to understand a particular spatial phenomenon. For example, population growth is the change of population over time, and can be quantified as the change in the number of individuals.
Index	ix	One of the most important features in the construction of an index number are its coverage, base period, weighting system, and method of averaging statistical results.
Share	sh	This value corresponds to a distribution, a portion that is calculated as a percentage of the derived total quantity. For instance, the prevalence of unemployment is usually measured using the 'unemployment rate' which is defined as the percentage of those in the labour force who are unemployed.
Weight	wgt	This concept is relevant for index numbers and transformations, such as 'GDP as constant prices'.
Female	f	The code for this term is applied if data needs to be categorised by gender.
Male	m	The code for this term is applied if data needs to be categorised by gender.
Total	t	This code corresponds to the total membership of a defined class of people, events or objects.
20-39	20-39	Age class between 20 and 39 years old of a certain phenomenon. For instance, 'sex ratio at age 20-39'. The same applies to other classes.
65+	65+	In this case, the membership is just focused on the age class over 65 years old.

Non-exhaustive list of derivations and/or restrictions

### Annex 3: Levels and scales of measurement

	<i>Stevens' scales (1946)</i>	<i>Forrest's extended levels (1999)</i>	<i>Required information and nature of data</i>	<i>Illustrative examples</i>	<i>Recommended abbreviation</i>
Qualitative	<b>Nominal</b>				<b>nom</b>
		Unique	All different (no duplication)	Country names, NUTS identifiers	nou
		Dichotomous	Membership (presence or absence)	Coastal areas, regions benefitting from EU funds	nod
		Categorical	Categories	Land use type, main religion	noc
	Graded membership	Categories plus degree of membership	Soil type with percentage conformance	nog	
Ranking (no quantity)	<b>Ordinal</b>				<b>ord</b>
		Complete ordinal	Unique ordering	Any ranking of regions or cities (without ex-aequo)	oru
	Classed ordinal	Categories plus ordering	Densely/intermediate/weakly populated regions	orc	
Quantitative	<b>Interval</b>		Measure plus arbitrary zero point	Temperatures in degrees	<b>int</b>
	<b>Ratio</b>				<b>rto</b>
		Extensive ratio	Measure (additive rules apply)	GDP, CO2 emissions, transported tons	rte
		Count	Measure (with unit =1)	Population, number of households, firms, births	rtc
		Derived ratio	Measure (quantity divided by)	GDP per inhabitant, labour productivity, cars per household	rtd
		Density ratio	Measure (quantity divided by area)	Population density	rde
		Cyclic ratio	Measure plus length of cycle	Angles, slope orientation	rty
	Constrained ratio	Probability or proportion, range [0,1]	Unemployment rate, share of young people	rtp	

Source: Levels and scales of measurement proposed by Stevens (1946) and Forrest (1999).



## Towards an ESPON thesaurus? Some preliminary considerations for the thematic structuring of the ESPON database

### MAIN RESULTS

- Construction of corporate thesaurus for structuring knowledge and facilitate information retrieval needs to comply with international guidelines
- Literature review shows a wide range of examples developed by international agencies and organisations on structured and controlled vocabulary
- Qualitative and quantitative text analysis improves the ability in defining standards and therefore to further advance on the harmonisation and coherence of concepts developed by ESPON
- For this purpose, it is important to determine links between different terms, namely by word co-occurrence analysis
- Lastly, we argue that such exercises are important to support the thematic structuring of the ESPON database by linking terms with data for information retrieval by potential users

**ESPON 2013 DATABASE**

**MARCH 2011**



## **LIST OF AUTHORS**

Nuno Madeira, University of Luxembourg

Geoffrey Caruso, University of Luxembourg

### **Contact**

E-mail: [nuno.madeira@uni.lu](mailto:nuno.madeira@uni.lu); Tel. +352 46 66 44 9691

E-mail: [geoffrey.caruso@uni.lu](mailto:geoffrey.caruso@uni.lu); Tel. +352 46 66 44 6625

## TABLE OF CONTENTS

Introduction .....	4
1. Main features for thesaurus construction .....	4
<i>The Equivalence Relationship</i> .....	5
<i>The Hierarchical Relationship</i> .....	5
<i>The Associative Relationship</i> .....	6
2. Thesauri developed by international organisations .....	7
<i>International Labour Organisation</i> .....	7
<i>United Nations Educational, Scientific and Cultural Organisation</i> .....	7
<i>United Nations Environment Programme</i> .....	8
<i>European Environmental Agency</i> .....	8
<i>Organisation for Economic and Co-operation Development</i> .....	9
<i>Food and Agricultural Organisation of the United Nations</i> .....	9
<i>Statistical Office of the European Communities</i> .....	10
3. Towards an harmonised vocabulary at the EU level .....	12
4. Text analysis applications for thematic structuring .....	13
5. Final considerations .....	14
References .....	15
Annexes .....	17

## **Introduction**

The thematic structuring of a geographical database aims at synthesising information through typologies and indicators that can address European political priorities and support applied research. The anonymous part of the application form specifies the main goals to be achieved by this research project. Within this context, the thematic dimension of data plays a crucial role, particularly due to the necessity of connecting geographical information to political objectives. Our goal is therefore to explore ways of progress on this issue and further advance on the structuring of the database into different themes and sub-themes that may support ESPON needs and its scientific objectives.

Over the last few years a great number of statistical agencies and international organisations have developed corporate glossary databases to integrate data and metadata in order to reinforce the communication and dissemination of standard concepts and benefit from the potentialities of terminology management tools for information retrieval by the end-users.

The seminal book *"Organizing Knowledge: An Introduction to Managing Access to Information"* says, in its most recent edition, that "...organization of knowledge is the other face of information retrieval. The better organized that knowledge is, the easier it is to retrieve specific items of knowledge..." (Rowley & Hartley, 2008: 4). The point of departure is therefore the relationship between data, information and knowledge. More concretely, how these concepts apply together for the formulation of a common set of terminology that supports the management of controlled vocabulary for storage and information retrieval?

According to these authors, depending on the environment and purpose of our needs different tools for organizing knowledge and information retrieval may apply. These tools have characteristics in common that improve the ability in defining standards and further advance in the harmonisation and coherence of concepts. One of the most interesting examples developed over the last decades is the thesaurus and its potentialities for alphabetical display of words and expressions that are linked by standardised semantic relationships. In the following chapter of this report some of these features are briefly described.

### **1. Main features for thesaurus construction**

From the moment information and data started to be easily accessible online many international agencies and organisations demonstrated their interest in having a corporate thesaurus to structure knowledge facilitate information retrieval. In order to ensure a minimum structure several international guidelines have been designed such as those formulated by the ISO. From a strictly information retrieval sense, a thesaurus is defined in the current international standard, ISO 11179, as the vocabulary of a controlled indexing language, formally organised so that relationships between concepts are made explicit. Its aim is to match the vocabulary used by the indexer with the language of the searcher, and thus improve the retrieval of relevant information.

The following section introduces some basic concepts about thesaurus. According to Noguera-Iso et al (2005) a thesaurus is an organised collection of terms enriched with relations that are linked to one another by cross-referencing to the database. A term is a word (or expression) that represents a conceptual category and for that reason a thesaurus can be seen as a structured and controlled vocabulary of words either represented in monolingual or multilingual format.

The adoption of such semantic tools has the benefit of ensuring that a subject will be described using the same preferred term each time it is indexed and this will make it easier to find all information about a specific topic during the search process. This means that using a thesaurus improves search results due to the inclusion of information about the relationships of words or expressions that represent standardised relationship indicators. The thesaurus specifies which word or expression will be used as the preferred

(or authorised) term for a particular concept and then connects the non-preferred terms or synonyms to it. The purpose is to provide a wide range of paths to reach those terms. Ultimately, it works as a classification tool to ease the communication process for indexers and information seekers that need to speak a common language.

Against this background, it is appropriate to take advantage from certain initiatives that stimulate the implementation of international standards and guidelines for wide dissemination of data and metadata exchange. Both INSPIRE and SDMX initiatives have the benefit of improving harmonisation and coherence of data. Since the ESPON DB 2013 project combines two types of data and metadata: (1) territorial statistical metadata (i.e. NUTS, WUTS, LAU) and (2) spatial metadata (i.e. environmental data), it is crucial to integrate both categories.

Due to the particularities of the ESPON programme access to data and metadata will be made available only in English language. With this regard, it should be mention that the management of monolingual thesauri is less complex but at the same time reduces the effectiveness for non-speaking English users of using ESPON results and evidence.

The purpose of building semantic tools (i.e. thesauri) is to maintain not only relationships among surface terms but also be able to deal with ambiguities. To this end, the thematic structuring of the database must be based on ESPON themes, sub-themes and keywords. The purpose is two-fold: in one hand, it would structure the database into different themes that address ESPON lexical concepts developed over the previous programme and, on the other hand, it would allow a more comprehensive use of the database by potential end-users (i.e. researchers, academics, practitioners, policy-makers).

However, some difficulties emerge within this discourse due to the complexity of political priorities. At this stage it is not obvious how the thematic structuring of the ESPON database should be organised. For this same reason it is important to have a minimum knowledge of what type of thematic fields should be added to an hypothetic ESPON thesaurus when storing data (and metadata) provided by the other applied research projects. With this regard, text and lexical analysis software applications show clear advantages and the ESTI framework, as described in the "*ESPON Handbook for Data Collection, Harmonisation and Quality Control*", can provide reasonable solutions for the automatic exploration of most of the data under investigation.

This brief overview is also important to understand that the structure of a thesaurus is generally defined a priori and terms are organised by cross-references that have generally three different types: (1) equivalence relationship; (2) associative relationship; and (3) hierarchical relationship. The following section provides a general description of these types of standardised relations (Nogueras et al, 2005; Severino, 2007).

### *The Equivalence Relationship*

This relationship concerns relations among terms that are considered equivalent inside the thesaurus. In other words, it covers synonyms. When two or more terms express the same concept, one of these is selected as preferred term. The equivalence relationship is expressed by the following: U or USE which leads from a non-preferred term to synonyms or quasi-synonyms which act as preferred terms, and UF or USED FOR the reciprocal, which records entry terms leading to the synonyms (e.g. spatial planning is a synonym for spatial development).

### *The Hierarchical Relationship*

The hierarchical relationship is the most important feature that distinguishes a thesaurus from an unstructured list of terms, such as glossary. It is based on position ranking, from a superior to a subordinate position. Terms that are superior in a category are at a super-ordinate level; terms that fall under or below a category are at a subordinate level. The broader term (BT) represents a class or a whole, while the narrower term (NT) refers



to components or part of the broader concept (e.g. Mediterranean countries is the broader term and Malta the narrower term).

In this sense, the hierarchical relationships links terms to other terms expressing more general and more specific concepts, hierarchically related terms are grouped under general subdivisions, which in turn are grouped into areas of knowledge. Normally, hierarchical relationships are indicated by the prefixes BT (broader term), i.e. label for the super-ordinate descriptor, between a specific descriptor and a more generic descriptor; and NT (narrower term), i.e. label for the subordinate descriptor, between a more generic descriptor and a more specific descriptor.

### *The Associative Relationship*

The associative relationship is used when two terms overlap in meaning. This relationship may be symmetrical or asymmetrical. In both cases, it indicates that a term has similarities with other concepts. However, some features must be distinguished.

A related term relationship alerts users to the fact that other information of interest may be classified under a different, but related, set of terms. Related terms relationships are only established between preferred terms. This type of relationship is symmetrical and therefore the relationship between terms is non-hierarchical (e.g. GDP is related to economic development and economic development is related to GDP). It means that similar terms (or related terms) are conceptually associated, but not equivalent or hierarchically related. The associative relationship is normally expressed by the abbreviation RT (related term). To what regards asymmetrical terms, the relationship is similar but not reciprocal. In other words, there is no related-term reference in the opposite direction (e.g. ageing population is related with demographic trends, but there searching for demographic trends is unlikely to be interested in ageing population).

## 2. Thesauri developed by international organisations

The following section describes some of the examples of online thesauri developed by international organisations, namely ILO (International Labour Organisation) and UNESCO (United Nations Educational, Scientific and Cultural Organisation), EnVoc (United Nations Environment Programme), GEMET (General Multilingual Environmental Thesaurus of the European Environmental Agency), Theseus (Statistical Office of the European Communities), and OECD Macrothesaurus (Organisation for Economic Co-operation and Development).

### *International Labour Organisation*

The ILO thesaurus may be browsed in two ways: alphabetically and hierarchically. It derives from a list of more than 4000 terms in the field of economic and social development (e.g. labour and employment policy, vocational training, working conditions) used to index and retrieve information. Potential new terms are monitored to see how their use develops in the literature and discussions are held with indexers and subject specialists to decide on their inclusion in the ILO thesaurus. Every term is presented in English, French, and Spanish. Many are followed by definitions or explanatory notes. Besides, ILO Thesaurus is available electronically and it can be easily downloaded in the three above mentioned languages.

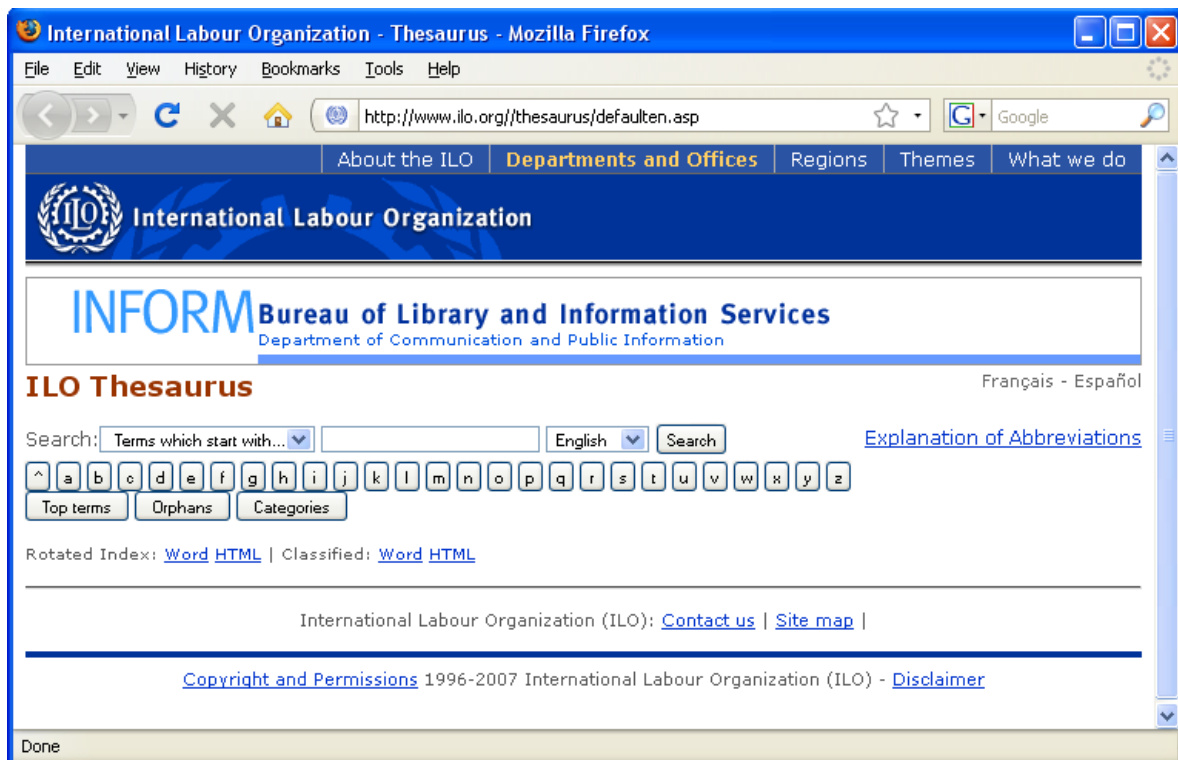


Figure 1. Print screen of the ILO Thesaurus online interface.  
Source: [www.ilo.org/thesaurus](http://www.ilo.org/thesaurus)

### *United Nations Educational, Scientific and Cultural Organisation*

The UNESCO thesaurus is a controlled vocabulary developed by the United Nations Educational, Scientific and Cultural Organisation which includes subject terms for several areas of knowledge (education, culture, social and human sciences), names of countries and groupings of countries (political, economic, geographic, ethnic and religious, and linguistic groupings). The terms are linked together by the three types of relationships described before, i.e. hierarchical, associative and equivalence relationships.

The UNESCO Thesaurus also includes scope notes which explain the meaning and application of terms. The same happens to French and Spanish equivalents of English

preferred terms. A more complete description of the structure of this thesaurus can be obtained in the report "*UNESCO Thesaurus: A Structures List of Descriptors for Indexing and Retrieving Literature in the Fields of Education, Science, Social and Human Science, Culture, Communication and Information*". A general description of the structure for each thesaurus by themes, sub-themes and related terms is illustrated as an annex to this report and addresses a particular focus on themes potentially related to the ESPON database.

#### *United Nations Environment Programme*

EnVoc is a multilingual thesaurus developed by UNEP (United Nations Environment Programme) with a controlled and structured vocabulary for the use in indexing, storing and retrieval of environmental information. It contains categorized and alphabetical lists of subjects and keywords in context list and is available in the six official United Nations languages (i.e. Arabic, Chinese, English, French, Russian, and Spanish). Additionally a number of other governments have undertaken the translation into their national languages which testifies the importance given to environment terminology and the value of the thesaurus as a reference tool (UNEP, 1997).

This initiative has its origin on the Infoterra thesaurus of environmental terms also developed by UNEP and constitutes revised version that reflects emerging environmental concerns and new technologies, especially in the field of environmental information. The thesaurus has evolved from an unstructured list of keywords with very limited application to a structured environmental thesaurus with a broad application base comprising several different types of expertise (e.g. librarians, database developers, GIS specialists).

As a subject, environment is multidisciplinary by nature. Its study, therefore, covers a wide range of linkages to disciplines and consequently it would be very difficult, though not impossible, to conduct a thesaurus containing every environmental term. The terms listed in EnVoc are drawn from a relatively high level in the environmental terminology hierarchy. This strategy eliminates unnecessary duplication with more specialised thesauri and the end product is closely integrated, compact and more practical to use.

#### *European Environmental Agency*

The processes of exchanging environmental information have significantly increased around the world during the last years and that reality motivated the development of a standard multilingual vocabulary capable of promoting and facilitating information exchange among countries on key environmental issues. The project on global environmental thesaurus was designed by a joint cooperation of different institutions with profound experience in this field.

The example provided by EEA (European Environmental Agency) is very comprehensive because it created important synergies among different organisations that had previously worked on the definition of common terminologies.

Indeed, shortly after setting up EEA it was launched an initiative called GEMET (General Environmental Multilingual Thesaurus). The idea was to merge in a unique thesaurus the best terminologies of six multilingual thesauri, including EnVoc, in order to propose a reference thesaurus to different organisations and, at the same time, provide an agreed common language basis for the exchange of environmental information. The initiative assumed great relevance in satisfying three basic functions, namely: define a system of controlled term, build a multilingual dictionary, and develop a glossary. In the late 1990s the final version of GEMET was published by EEA. The current version is available in 27 languages, and contains over 6,000 descriptors and access is open to all users in several formats. It can be browsed and searched on-line, accessed through Web services or downloaded as HTML or SKOS files (cf. [www.eionet.europa.eu/gemet](http://www.eionet.europa.eu/gemet)).

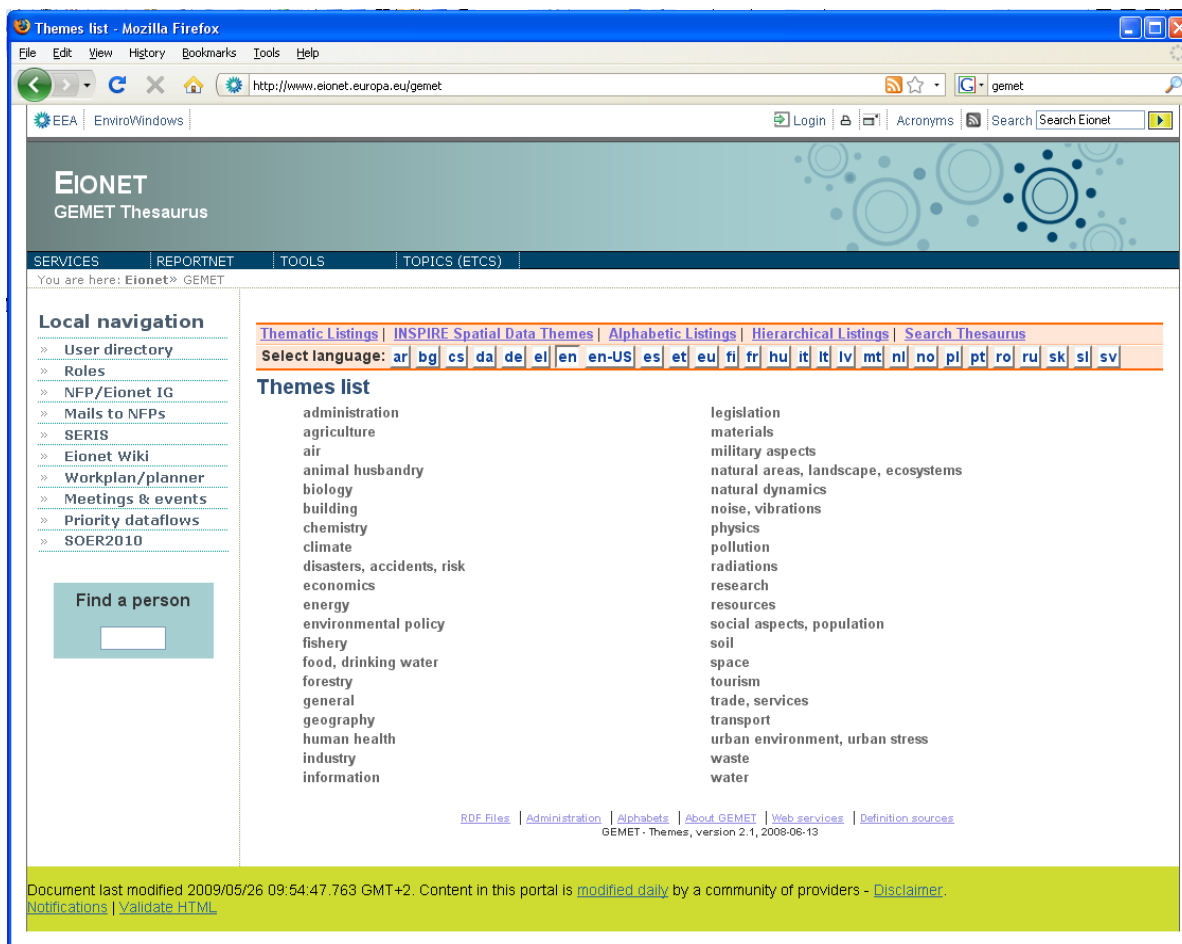


Figure 2. Print screen of the GEMET Thesaurus online interface.  
Source: www.eionet.europa.eu/gemet

### *Organisation for Economic and Co-operation Development*

This thesaurus has not been entirely developed by OECD. It represents a joint cooperation initiative in association with United Nations and the International Organisation of Labour to create some unity and consistency on economic and social development terms that are used in the international context. To this end, the constitution of the OECD Macrothesaurus used information from several sources and compiled the terms into a single scheme that would be useful to all constituents. Table 1 in Annex gives an overview of the structured word tree by themes and sub-themes.

### *Food and Agricultural Organisation of the United Nations*

The AGROCOV Thesaurus (cf. www.fao.org/agrovoc) was developed by FAO (Food and Agricultural Organisation of the United Nations) and the Commission of the European Communities, in the early 1980s. The initiative was designed to cover the terminology of all subject fields in agriculture, forestry, fisheries, food and related domains (e.g. environment). Due to the potentialities for indexing and retrieving data at world-wide level it is available in the five official languages at FAO (i.e. English, French, Spanish, Chinese and Arabic).

The AGROVOC Thesaurus is currently structured as a traditional thesaurus and stored in a relational database. It consists of words and expressions (terms), organized in semantic relationships and used to identify or search resources. The main role is to standardize the indexing process in order to make searching simpler and more efficient, and to provide the user with the most relevant resources.

As a multilingual, structured and controlled vocabulary, AGROVOC is made up of terms, which consist of one or more words representing always one and the same concept. For each term, a word block is displayed, showing the semantic relationship to other words or expressions. Scope notes are also used to clarify the meaning and the context of terms (words and expressions).

The rationale behind the process of structuring this thesaurus is similar to the examples described in the previous sections of this technical report, particularly to what regards the semantic relationships between concepts. These relationships (i.e. equivalence, hierarchical, associative) provide the scope and structure for the thesaurus. However, such initiatives can greatly benefit from the add-value of using extended sets of relationships that go beyond the standardised relationships to perform more granular and more consistent indexing, and to enable more effective searching and browsing for users.

Ultimately, it is worth mentioning that AGROVOC is available free of charge and users are encouraged to propose new terms for inclusion in the database. Figure 3 illustrates a print screen of the online interface.

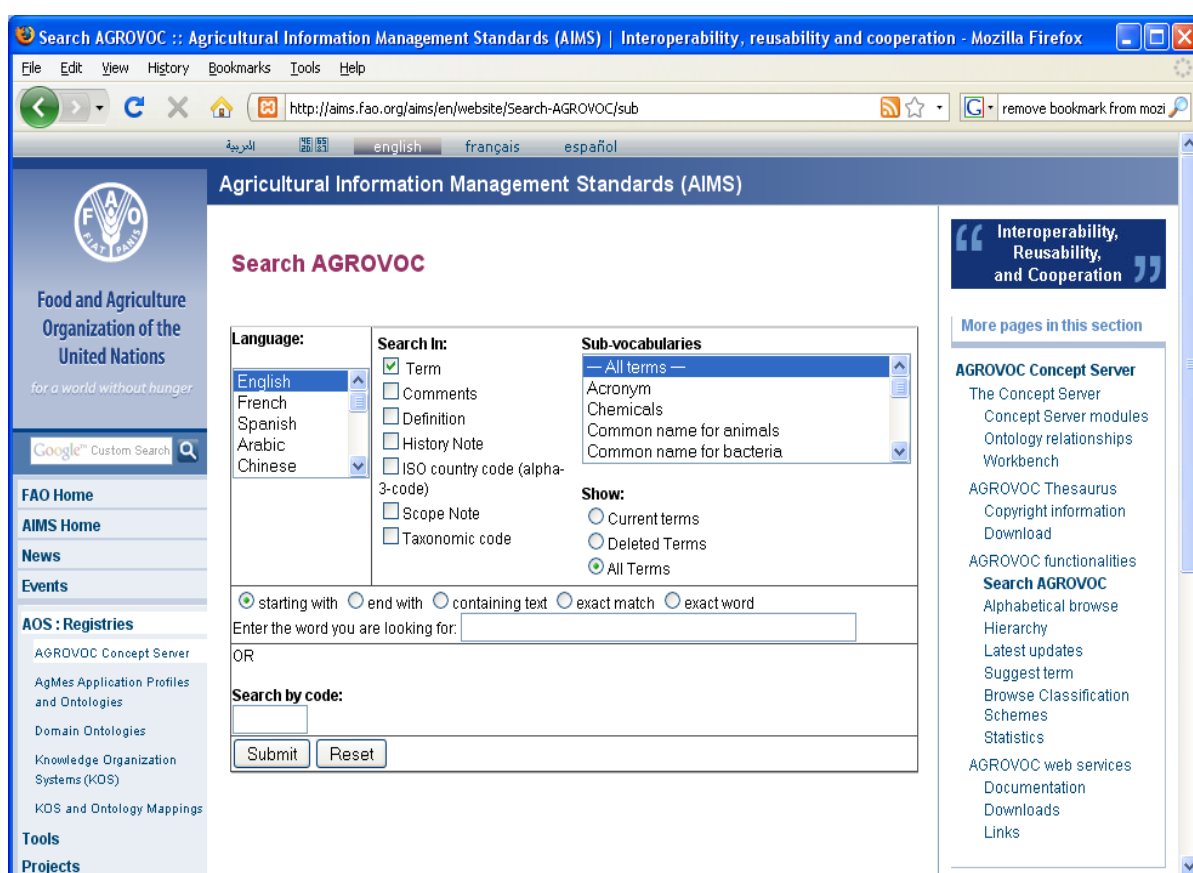


Figure 3. Print screen of the AGROVOC Thesaurus online interface.  
Source: aims.fao.org/aims

### Statistical Office of the European Communities

Eurostat, like many other international organisations, also developed a thesaurus-based system for consultation, indexation and retrieval of statistical data in order to help end-users to find information in its major reference database, *NewCronos*<sup>1</sup>. However, shortly after the beginning of this initiative the project was abandoned due to a negative priority given by Eurostat (De Norre; Groenez & Pellegrino, 2004). As an indicative reference, Figure 1 shows the linkages defined by Eurostat for statistical information systems.

<sup>1</sup> *NewCronos* is a unique database of macroeconomic and social statistics at EU level that includes data from candidate countries, EFTA countries, and major economic partners of the European Union (e.g. Japan, United States).

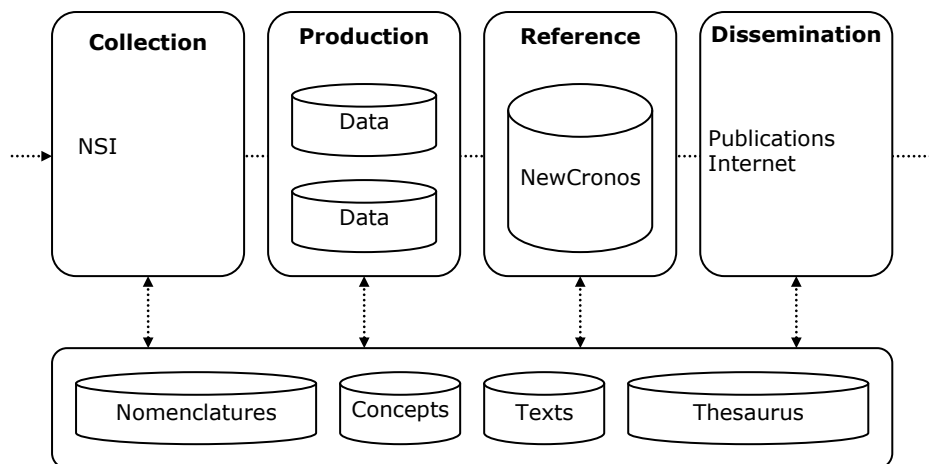


Figure 1. Eurostat's architecture for statistical information systems.

The initiative dates back the beginning of this decade and the logic behind the process has been described by De Norre & Groenez (2002) as a structured list of expressions used for indexing the content of data (i.e. statistical tables) and therefore ease the searching process by potential end-users. *Theseus*, as it was called, has been initially designed to be a multilingual thesaurus that would link major metadata of *NewCronos* to the contents of the thesaurus.

The agreement to further advance *Theseus* involved the development of an information management system and a web-interface for consultation. The content also required expertise from a group of specialists having the appropriate linguistic and content knowledge to regularly follow the evolution of the *NewCronos* database. In addition, it required expertise in adopting a more natural lexical instead of a rather technical language. For this reason, *Theseus* intended to suppress the ambiguities of natural lexical, in particular synonymy and homonymy.

The grounds for adopting a thesaurus were similar to the ones that lead other organisations to demonstrate the same approach, i.e. improving the information retrieval process by offering a multilingual tool to support potential users in searching data and information. To this end, *Theseus* structured its vocabulary in standardised semantic relationships and domains of knowledge. Therefore, the technical aspects of *Theseus* have been structured, at the first level, in semantic classes that would assemble all terms that have a specific meaning for a given sub-topic.

However, the purpose of this initiative goes far beyond the features offered by other online thesauri previously described in this report. In fact, this initiative also involved the capacity of linking metadata developed to define and describe statistical data and their structure to reference database *NewCronos*. This process was the basis for implementing the search facilities and three different approaches were considered in the indexation procedure.

The automatic indexation created too many problems and was abandoned without leaving any visible results. According to De Norre & Groenez (2002), automatic indexation was not compatible with high quality indexation and often required manual validation. The outputs obtained by this process were quite arguable.

The second approach followed a two-step procedure. Its rationale was analysed and developed by using a semi-automatic indexation that, firstly, compared metadata texts with *Theseus* keywords and, secondly, made use of information science expertise to determine appropriate links between metadata and keywords. The difficulties encountered during the test phase revealed that this option was not the best possibility. In general, the indexation process was too imprecise and did not take into account implicit concepts.

Finally, the third approach proved to be much more efficient. The manual indexation procedure provided some useful information on the quality of the metadata mostly due to the introduction of dictionaries defining both the *NewCronos* structures (i.e. themes, domains, collections, groups, subjects) and tables (i.e. titles).

Despite the efforts in finding a proper indexation procedure for the Eurostat thesaurus, the truth is that other major problems needed a solution. One of those main issues raised during the indexation procedure was how to keep the index coherent in respect with the daily evolution of *NewCronos* and *Theseus*. In other words, it was necessary to find a proper logging and reporting procedure of the changes to support the work of information science specialists. Consequently, depending of the type of changes specialised staff could opt on the best suitable indexation procedure (i.e. manual, semi-automatic) to make the necessary updates which, in this case, were applied in a much more limited scope.

Ideally, *Theseus* proposed different search facilities based on the indexation. Some of the features included: search in via the index; selection of keywords or synonyms via thematic list of semantic categories, or permuted index (i.e. tracking the word in keyword expressions); combination of keywords (i.e. Boolean equations); and feedback to refine the search by offering context information derived from the thesaurus (i.e. semantic categories, hierarchical relations and scope notes).

The purpose of building a thesaurus seemed to be unquestionable due to the given importance attributed by Eurostat to its high-quality statistical information services, particularly when it comes to provide statistical information and promote harmonisation of statistical methods across Europe. Those ambitions anticipated as well some further technical developments regarding the integration with other tools developed by Eurostat, such as RAMON (Metadata Server), or CODED (Concepts and Definitions Database).

### **3. Towards an harmonised vocabulary at the EU level**

The examples described previously demonstrate an increasing interest by international organisations in creating thesaurus for structuring their knowledge and support information retrieval.

The Eurovoc thesaurus is a multilingual and thematic thesaurus focusing in all fields which are of importance for the activities of the European institutions. It has been developed as a result of cooperation between the European Parliament, the European Commission and Publications Office with support from DG Information Society and Media. The thematic scope of Eurovoc is defined by 21 fields, divided into 127 microthesauri, and the last version contains 6645 descriptors, 6669 hierarchical and 3636 associative relationships.

Some fields are more developed than others due to the interests expressed by the European institutions. However, this situation does not mean that the thematic structuring has been organised using a logic method. On the contrary, the grouping of descriptors into fields is to a certain extent arbitrary. This means that certain descriptors can relate to two or more subject fields, but in order to make the thesaurus easier to manage and to limit its size it is generally accepted that limits have to be established according an hierarchy. Descriptors which could fit into two or more subject fields are assigned only to the field in which seems to be most natural for users.

In terms of languages and standards adopted by the Eurovoc thesaurus, the method is rather similar with other online thesauri developed by international organisations. It is published in all official languages of the European Community and complies with standards defined by the International Standards Organization, namely ISO 2788 for monolingual thesauri and ISO 5964 for multilingual thesauri. The structure of thesauri has been defined with regard to semantic relationships. Within the framework of the Eurovoc thesaurus, four different relationships may apply, namely: (1) micro-thesaurus, (2) equivalence, (3) hierarchical, (4) associative relationship and, finally, (5) scope note.

All descriptors are accompanied by a reference to a microthesaurus in order to show, under the abbreviation of MT, which microthesaurus or microthesauri they belong. For instance, within the field *04 Politics* it has been included several microthesauri, such as *0406 Political framework* or *0411 Political party* (cf. Table 2 in Annex).

The equivalence relationship between descriptors and non-descriptors is shown by the abbreviations: UF (used for), between the descriptor and the non-descriptor(s) it represents; USE between a non-descriptor and the descriptor which takes its place. The equivalence relationship covers relationships of several types: (a) genuine synonymity, or identical meanings; (b) near-synonymity, or similar meanings; (c) antonymy, or opposite meanings; (d) inclusion, when a descriptor embraces one or more specific concepts which are given the status of non-descriptors; because are not often used.

The hierarchical relationships between descriptors are shown by the abbreviations:

BT (broader term) between a specific descriptor and a more generic descriptor, together with a number showing the number of hierarchical steps between the specific descriptor and each broader term. Occasionally, certain descriptors in fields 72 Geography and 76 International Organizations are regarded as multi-thematic due to the presence of several broader terms at the next higher level, and descriptors with no broader terms are top terms. NT (narrower term) between a generic descriptor and a more specific descriptor, together with a number showing the number of hierarchical steps between the generic term and each narrower term.

The associative relationship between descriptors is shown by the abbreviation RT (related term) between two associated descriptors. The most significant features of the associative relationship are connected with its symmetrical approach, its incompatibility with the hierarchical relationship. Indeed, the descriptors under the same top term cannot be linked by an associative relationship, i.e. if two descriptors are linked by a hierarchical relationship there cannot be an associative relationship between them and inversely. Moreover, this semantic relationship can be interpreted in various forms (e.g. cause and effect, sequence in time or space, among others).

Ultimately, some descriptors are accompanied by scope notes, introduced by the abbreviation SN, and containing either a definition, if this clarifies the meaning of the descriptor or guidance on how to use the descriptor when indexing and formulating queries.

The Luxembourg-based Office for Official Publications of the European Communities (OPOCE) acts, on behalf of the European Commission, on management issues related to the Eurovoc thesaurus, especially regarding the necessary adaptations to the latest developments in which the Community institutions are active.

This special feature of Eurovoc is aimed at meeting the needs expressed by its users while keeping in mind the objective of covering in homogeneous manner, the fields of activity of the European Community. To this end, Eurovoc is subject to maintenance actions based on both internal and external needs and all users are invited to contribute to the development of Eurovoc by suggesting the addition, deletion or modification of descriptors, non-descriptors and semantic relationships.

Finally, it is worth mentioning that OPOCE plays a central role in the dissemination policy of the whole European Commission by providing tools, such as the web-based EU-Bookshop application, for searching and retrieving publications. The Eurovoc Steering Committee has officially adopted the version Eurovoc 4.3 on 28 November 2008 and shortly after that has been made available online (cf. <http://europa.eu/eurovoc>).

#### **4. Text analysis applications for thematic structuring**

In order to ensure a comprehensive thematic structuring of the ESPON database and therefore identify, retrieve and navigate statistical and geographical data it is essential to integrate spatial concepts developed by the previous ESPON programme. Within this





retrieval. To this end, the examples provided illustrate different approaches that vary according to the needs. However, the methods used for thesaurus construction are generally the same and, to a certain extent, represent joint cooperation initiatives among international organisations that cover similar fields or categories. The purpose of such initiatives is to create some consistency and unity. For instance, the constitution of the OECD Macrothesaurus corresponds to a joint initiative among United Nations, the International Labour Organisation, and the Organisation for Economic Co-operation and Development. In the same way, the development of GEMET is a result of the cooperation between the United Nations Environment Programme and the European Environmental Agency.

The purpose of having a specific thesaurus for the ESPON database would certainly contribute to generate consistency and enhance searches more effectively by end-users. In this sense, quantitative text analysis tools have the ability to support the definition of appropriate vocabulary for indexing terms. It has been argued that some of the lexical developed by the previous ESPON programme influenced the terminology adopted by other EU institutions and therefore similar exercises could further support the constitution of an ESPON thesaurus. Against this background, the ability to define standards and improve harmonisation and coherence of spatial concepts would preserve the overall meaning of words and expressions for information retrieval by potential end-users.

Next steps towards the thematic structuring of the ESPON database will involve qualitative and quantitative text analysis on various scientific reports (e.g. ESPON, INTERREG, INTERACT, DGs, CoR, EESC, European Parliament), determine links between different terms, namely by word co-occurrence analysis and types of relationships and, ultimately, link those terms with data for information retrieval (indicators, typologies) in order to structure the ESPON database.

## References

- De Norre, B. & D. Groenez (2002). *A Multilingual Thesaurus for Accessing Eurostat's Reference Databases*. Joint UNECE/EUROSTAT Work Session on Statistical Metadata. Luxembourg, 6-8 March 2002.
- De Norre, B.; Groenez, D. & M. Pellegrino (2004). *Integrating Statistical Terminology Tools within Eurostat's Dissemination Policy*. Joint UNECE/EUROSTAT/OECD Work Session on Statistical Metadata. Geneva, 9-11 February 2004.
- Diaz Munoz, P. (2008). The role of Statistical Data and Metadata Exchange in global statistical infrastructure. *Statistical Journal of the IAOS*, 25(2008): 47-54.
- Felluga, B.; Plini, P.; Cunningham, G. & S. Lucke (2000). *The Global Environmental Thesaurus Project*. Global Conference on Access to Environmental Information. Dublin, 11-15 September 2000.
- Nogueras, J. ; Zarazaga, F. & P. Muro (2005). *Geographic Information Metadata for Spatial Data Infrastructures. Resources, Interoperability and Information Retrieval*. Springer: Heidelberg.
- OECD (2007). *Data and Metadata Reporting and Presentation Handbook*. Organisation for Economic Co-operation and Development: Paris.
- Roe, S. & A. Thomas (2004). *The Thesaurus. Review, Renaissance, and Revision*. Routledge: London.
- Rowley, J. & R. Hartley (2008). *Organizing Knowledge: An Introduction to Managing Access to Information*. Ashgate: London.
- Severino, F. (2007). The term development in the thesauri of international organisations. *The European Journal of Development Research*, 19(2). 327-351.

UNEP (1997). *EnVoc Multilingual Thesaurus of Environmental Terms*. United Nations Environment Programme: Nairobi.

Williamson, N. & C. Beghtol (2004). *Knowledge Organization and Classification in International Information Retrieval*. Routledge: London.

## Annexes

ILO Thesaurus	UNESCO Thesaurus
01 International relations 02 Social policy, social protection and social security 03 Economic development 04 Law, human rights, government and politics 05 Social sciences, culture, humanities and art 06 Education and training 07 Rural development, agriculture, forestry 08 Economic activities 09 Trade 10 Transport NT1 Transport NT1 Goods Transport NT1 transport Infrastructure NT1 Vehicles NT1 Air, sea, inland waterway, road and railway NT2 Mode of transportation NT2 Air transport NT2 Sea transport NT3 Port NT3 Waterway NT4 inland waterway NT 5 canal BT Waterway RT <i>Inland water transport</i> NT2 Inland water transport NT2 Railway transport NT2 Road transport NT1 Loading and packing NT1 International, inland and urban transport 11 Financing 12 Management 13 Labour and employment 14 Population, race relations and migration 15 Health and safety 16 Environmental sciences NT1 Ecology NT1 Natural resources NT2 Natural resources NT2 Resources development BT Natural resources NT1 Environment 17 Earth sciences 18 Research and science 19 Library and information science	Countries and Country Groupings Africa NT1 Africa NT2 Central Africa NT2 East Africa NT2 North Africa NT3 Algeria BT <i>Arab states</i> BT <i>French speaking Africa</i> BT <i>French speaking countries</i> BT <i>Islamic countries</i> BT <i>Maghreb</i> BT <i>Mediterranean countries</i> BT <i>North Africa</i> NT3 Egypt BT <i>Arab states</i> BT <i>French Speaking countries</i> BT <i>Mediterranean countries</i> BT <i>North Africa</i> NT3 Libya NT3 Morocco NT3 Sahara NT3 Tunisia NT2 Southern Africa NT2 West Africa NT1 Americas and the Caribbean NT1 Asia and the Pacific NT1 Economic Groupings NT1 Ethnic and Religious Groupings NT1 Europe NT1 Geographic Groupings NT1 Linguistic Groupings NT1 Political Groupings Culture Education Information and Communication Politics, Law and Economics Science Social and Human Sciences Urban Planning

Table 1. Brief comparison between ILO and UNESCO thesauri.

Source:

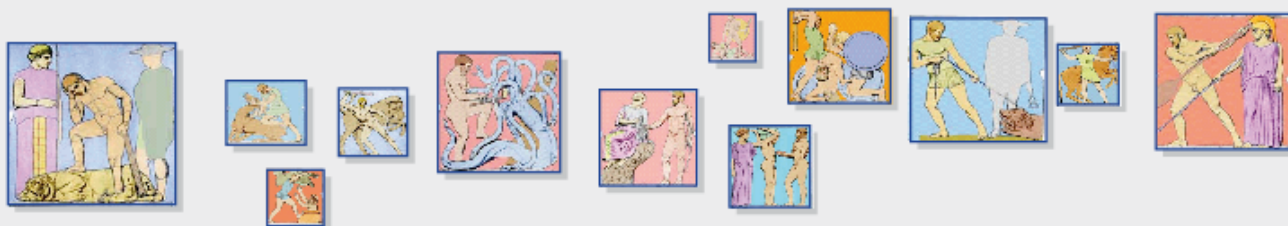
Note: Both examples illustrate a detailed overview of selected categories and its semantic relationships.

<p>1. INTERNATIONAL COOPERATION; INTERNATIONAL RELATIONS</p> <ol style="list-style-type: none"> <li>1. International Cooperation</li> <li>2. International Relations</li> <li>3. International Organizations</li> <li>4. Countries and Regions</li> </ol> <p>2. ECONOMIC POLICY; SOCIAL POLICY; PLANNING</p> <ol style="list-style-type: none"> <li>1. Economic Policy; Planning</li> <li>2. Social Policy</li> <li>3. Social Problems</li> <li>4. Social Services</li> </ol> <p>3. ECONOMIC CONDITIONS; ECONOMIC RESEARCH; ECONOMIC SYSTEMS</p> <ol style="list-style-type: none"> <li>1. Economic research; Economics</li> <li>2. Economic Conditions</li> <li>3. Economic Systems</li> </ol> <p>4. INSTITUTIONAL FRAMEWORK</p> <ol style="list-style-type: none"> <li>1. Law; Legislation</li> <li>2. Human Rights</li> <li>3. Politics</li> </ol> <p>5. CULTURE; SOCIETY</p> <ol style="list-style-type: none"> <li>1. Social Sciences</li> <li>2. Culture</li> <li>3. Society</li> <li>4. Ethics; Religion</li> <li>5. Art</li> <li>6. Languages</li> <li>7. Communication</li> </ol> <p>6. EDUCATION; TRAINING</p> <ol style="list-style-type: none"> <li>1. Educational Sciences</li> <li>2. Educational Development; Educational Policy</li> <li>3. EDUCATIONAL SYSTEMS</li> <li>4. EDUCATIONAL INSTITUTIONS</li> <li>5. CURRICULUM; TEACHING; LEARNING</li> <li>6. STUDENTS; TEACHING PERSONNEL</li> </ol> <p>7. AGRICULTURE</p> <ol style="list-style-type: none"> <li>1. AGRICULTURAL ECONOMICS</li> <li>2. LAND ECONOMICS</li> <li>3. AGRICULTURAL ENTERPRISES</li> <li>4. AGRICULTURAL EQUIPMENT</li> <li>5. AGRICULTURAL PRODUCTION</li> <li>6. AGRICULTURAL RESEARCH</li> <li>7. PLANT PRODUCTION</li> <li>8. FORESTS</li> <li>9. ANIMAL PRODUCTION</li> <li>10. FISHERY</li> </ol> <p>8. INDUSTRY</p> <ol style="list-style-type: none"> <li>1. INDUSTRIAL ECONOMICS</li> <li>2. INDUSTRIAL ENTERPRISES</li> <li>3. INDUSTRIAL ENGINEERING; INDUSTRIAL EQUIPMENT</li> <li>4. INDUSTRIAL PRODUCTION; INDUSTRIAL PRODUCTS</li> <li>5. INDUSTRIAL RESEARCH</li> <li>6. FOOD INDUSTRY</li> <li>7. WOODWORKING INDUSTRY; PULP AND PAPER INDUSTRY</li> <li>8. TEXTILE INDUSTRY; LEATHER INDUSTRY</li> <li>9. RUBBER INDUSTRY</li> <li>10. CONSTRUCTION INDUSTRY; CERAMICS INDUSTRY; GLASS INDUSTRY</li> <li>11. ENERGY</li> <li>12. CHEMICAL INDUSTRY</li> <li>13. MINING</li> <li>14. METALWORKING INDUSTRY</li> <li>15. ELECTRONICS; ELECTRICAL EQUIPMENT</li> <li>16. COMMUNICATION INDUSTRY</li> </ol> <p>9. TRADE</p> <ol style="list-style-type: none"> <li>1. DEMAND; MARKET; CONSUMPTION</li> <li>2. PRICES</li> <li>3. MARKETING</li> <li>4. DOMESTIC TRADE</li> <li>5. INTERNATIONAL TRADE</li> </ol> <p>10. TRANSPORT</p> <ol style="list-style-type: none"> <li>1. TRANSPORT ECONOMICS</li> <li>2. GOODS; PASSENGERS</li> <li>3. TRANSPORT INFRASTRUCTURE</li> </ol>	<ol style="list-style-type: none"> <li>4. MEANS OF TRANSPORT</li> <li>5. MODES OF TRANSPORT</li> <li>6. LOADING; PACKAGING</li> <li>7. INTERNATIONAL TRANSPORT; URBAN TRANSPORT</li> <li>8. TRAFFIC</li> <li>9. FREIGHT</li> </ol> <p>11. PUBLIC FINANCE; BANKING; INTERNATIONAL MONETARY RELATIONS</p> <ol style="list-style-type: none"> <li>1. PUBLIC FINANCE; TAXATION</li> <li>2. CURRENCIES; FINANCING</li> <li>3. INTERNATIONAL MONETARY SYSTEM</li> </ol> <p>12. MANAGEMENT; PRODUCTIVITY</p> <ol style="list-style-type: none"> <li>1. ENTERPRISES</li> <li>2. ECONOMIC CONCENTRATION</li> <li>3. ENTREPRENEURS</li> <li>4. MANAGEMENT</li> <li>5. EQUIPMENT</li> <li>6. TECHNOLOGY</li> <li>7. PRODUCTION; PRODUCTIVITY</li> <li>8. PRODUCTS; PRODUCT DEVELOPMENT</li> <li>9. COST ACCOUNTING; PROFIT</li> </ol> <p>13. LABOUR</p> <ol style="list-style-type: none"> <li>1. HUMAN RESOURCES</li> <li>2. EMPLOYMENT SERVICES; OCCUPATIONAL QUALIFICATIONS; PERSONNEL MANAGEMENT</li> <li>3. OCCUPATIONAL SAFETY</li> <li>4. DISMISSAL; LABOUR MOBILITY</li> <li>5. LABOUR RELATIONS</li> <li>6. WAGES; WAGE INCENTIVES</li> <li>7. LEISURE</li> <li>8. OCCUPATIONS</li> </ol> <p>14. DEMOGRAPHY; POPULATION</p> <ol style="list-style-type: none"> <li>1. POPULATION DYNAMICS</li> <li>2. AGE GROUPS</li> <li>3. ETHNIC GROUPS</li> <li>4. HABITAT; RURAL AREAS; URBAN AREAS</li> <li>5. FERTILITY; FAMILY PLANNING</li> <li>6. MORTALITY</li> <li>7. MIGRATION</li> </ol> <p>15. BIOLOGY; FOOD; HEALTH</p> <ol style="list-style-type: none"> <li>1. BIOLOGY; PARASITOLOGY; BIOCHEMISTRY</li> <li>2. ANATOMY; GENETICS; PHYSIOLOGY</li> <li>3. FOOD; NUTRITION</li> <li>4. MEDICINE; DISEASES</li> <li>5. PHARMACOLOGY; TOXICOLOGY</li> </ol> <p>16. ENVIRONMENT; NATURAL RESOURCES</p> <ol style="list-style-type: none"> <li>1. ECOLOGY</li> <li>2. NATURAL RESOURCES</li> <li>3. DISASTERS; POLLUTION</li> <li>4. POLLUTION CONTROL; ENVIRONMENTAL ENGINEERING</li> <li>5. RESOURCES CONSERVATION</li> </ol> <p>17. EARTH SCIENCES; SPACE SCIENCES</p> <ol style="list-style-type: none"> <li>1. METEOROLOGY</li> <li>2. CLIMATE</li> <li>3. GEOGRAPHY</li> <li>4. GEOPHYSICS; GEOLOGY; SOIL SCIENCES</li> <li>5. HYDROLOGY; WATER</li> <li>6. OCEANOGRAPHY</li> <li>7. SPACE SCIENCES</li> </ol> <p>18. SCIENCE; RESEARCH; METHODOLOGY</p> <ol style="list-style-type: none"> <li>1. RESEARCH; SCIENCE</li> <li>2. ORGANIZATION OF RESEARCH</li> <li>3. RESEARCH METHODS; THEORY</li> <li>4. DATA COLLECTING</li> <li>5. EXPERIMENTS</li> <li>6. MEASUREMENT</li> <li>7. MAPPING</li> <li>8. MATHEMATICS; STATISTICAL ANALYSIS</li> <li>9. COMPARISON; EVALUATION</li> <li>10. FORECASTS; TIME FACTOR</li> </ol> <p>19. INFORMATION; DOCUMENTATION</p> <ol style="list-style-type: none"> <li>1. INFORMATION</li> <li>2. DOCUMENTS</li> <li>3. TERMINOLOGY</li> <li>4. CONFERENCES</li> </ol>
--	--

Table 2. List of themes and sub-themes of the OECD Macrothesaurus.

<p>04 POLITICS  0406 political framework  0411 political party  0416 electoral procedure and voting  0421 parliament  0426 parliamentary proceedings  0431 politics and public safety  0436 executive power and public service  08 INTERNATIONAL RELATIONS  0806 international affairs  0811 cooperation policy  0816 international balance  0821 defence  10 EUROPEAN COMMUNITIES  1006 Community institutions and European civil service  1011 European Union law  1016 European construction  1021 Community finance  12 LAW  1206 sources and branches of the law  1211 civil law  1216 criminal law  1221 justice  1226 organisation of the legal system  1231 international law  1236 rights and freedoms  16 ECONOMICS  1606 economic policy  1611 economic growth  1616 regions and regional policy  1621 economic structure  1626 national accounts  1631 economic analysis  20 TRADE  2006 trade policy  2011 tariff policy  2016 trade  2021 international trade  2026 consumption  2031 marketing  2036 distributive trades  24 FINANCE  2406 monetary relations  2411 monetary economics  2416 financial institutions and credit  2421 free movement of capital  2426 financing and investment  2431 insurance  2436 public finance and budget policy  2441 budget  2446 taxation  2451 prices  28 SOCIAL QUESTIONS  2806 family  2811 migration  2816 demography and population  2821 social framework  2826 social affairs  2831 culture and religion  2836 social protection  2841 health  2846 construction and town planning  32 EDUCATION AND COMMUNICATIONS  3206 education  3211 teaching  3216 organisation of teaching  3221 documentation  3226 communications  3231 information and information processing  3236 information technology and data processing  36 SCIENCE  3606 natural and applied sciences  3611 humanities  40 BUSINESS AND COMPETITION  4006 business organisation</p>	<p>4011 business classification  4016 legal form of organisations  4021 management  4026 accounting  4031 competition  44 EMPLOYMENT AND WORKING CONDITIONS  4406 employment  4411 labour market  4416 organisation of work and working conditions  4421 personnel management and staff remuneration  4426 labour law and labour relations  48 TRANSPORT  4806 transport policy  4811 organisation of transport  4816 land transport  4821 maritime and inland waterway transport  4826 air and space transport  52 ENVIRONMENT  5206 environmental policy  5211 natural environment  5216 deterioration of the environment  56 AGRICULTURE, FORESTRY AND FISHERIES  5606 agricultural policy  5611 agricultural structures and production  5616 farming systems  5621 cultivation of agricultural land  5626 means of agricultural production  5631 agricultural activity  5636 forestry  5641 fisheries  60 AGRI-FOODSTUFFS  6006 plant product  6011 animal product  6016 processed agricultural produce  6021 beverages and sugar  6026 foodstuff  6031 agri-foodstuffs  6036 food technology  64 PRODUCTION, TECHNOLOGY AND RESEARCH  6406 production  6411 technology and technical regulations  6416 research and intellectual property  66 ENERGY  6606 energy policy  6611 coal and mining industries  6616 oil industry  6621 electrical and nuclear industries  6626 soft energy  68 INDUSTRY  6806 industrial structures and policy  6811 chemistry  6816 iron, steel and other metal industries  6821 mechanical engineering  6826 electronics and electrical engineering  6831 building and public works  6836 wood industry  6841 leather and textile industries  6846 miscellaneous industries  72 GEOGRAPHY  7206 Europe  7211 regions of EU Member States  7216 America  7221 Africa  7226 Asia and Oceania  7231 economic geography  7236 political geography  7241 overseas countries and territories  76 INTERNATIONAL ORGANISATIONS  7606 United Nations  7611 European organisations  7616 extra-European organisations  7621 world organisations  7626 non-governmental organisations</p>
---	---

Table 3. List of themes and sub-themes of the Eurovoc Thesaurus.  
Source: <http://europa.eu/eurovoc>.



## UPDATE OF ESPON DATABASE 2006 INTO THE 2013 VERSION

### MAIN RESULTS

- Two ESPON Databases can be considered within the previous programme. An internal, used by ESPON projects and an external available on ESPON Website which synthesizes and presents as a clear way ESPON 2001-2006 results.
- The update of the ESPON Database available on Web appears more logical.
- Thanks to the new data and metadata model developed in the ESPON DB Project, the level of description of each indicator is improved.
- Expected integration in the ESPON 2013 DB: December 2010

**ESPON 2013 DATABASE**

**MARCH 2011**



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

27 pages

# LIST OF AUTHORS

Ronan Ysebaert, UMS RIATE, Paris, France

Bogdan Moisuc, LIG, Grenoble, France

## Contact

(1) [ronan.ysebaert@ums-riate.fr](mailto:ronan.ysebaert@ums-riate.fr)

(2) [Bogdan.Moisuc@imag.fr](mailto:Bogdan.Moisuc@imag.fr)

(1) *Tel. + 33 1 57 27 65 32*

(2) *Tel. + 33 4 76 82 72 25*



# TABLE OF CONTENT

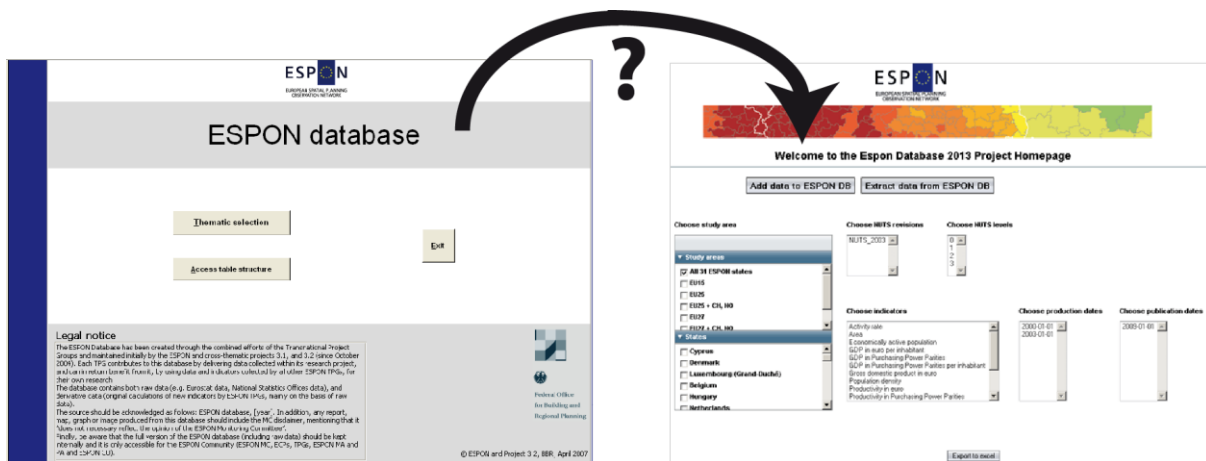
Introduction.....	3
1 Overview of the ESPON 2006 Database .....	4
1.1 The "internal" ESPON 2006 Database .....	4
1.2 The "external" ESPON 2006 Database.....	7
2 What strategy for updating the ESPON 2006 Database?.....	10
2.1 Trying to integrate the external ESPON 2006 Database (available on the ESPON Website) .....	10
2.2 Integrate both Eurostat indicators and ESPON indicators.....	10
2.3 Description of datasets who will integrate the ESPON DB 2013 .....	12
3. Integration of the ESPON 2006 Database in the 2013 DB architecture - a real added value.....	13
3.1. Elements added by the ESPON 2013 Database Project to the metadata of ESPON 2006 indicators .....	13
3.2. Improvements of the metadata and data information .....	15
3.3. New metadata structure for existing information.....	16
APPENDIX 1 - Updated indicators from the ESPON 2006 Database – Data from TPG's (ESPON Project Indicators) .....	17
APPENDIX 2 - Updated indicators from the ESPON 2006 Database – Data from Eurostat (ESPON Basic Indicators).....	25

# Introduction

One of the main objective of challenge 1 – collection of basic regional data – is to “*be able in a short delay to connect the new information elaborated by ESPON 2013 Program with former datasets elaborated by ESPON 2006 Program*” (First Interim Report of ESPON 2013 Database Project). It is a clear fact that ensuring the continuity between the different ESPON databases is a crucial bullet point for the development of the ESPON Program. In other words, it raises the question of the integration and implementation of indicators produced by ESPON 2006 Program into the new version of the database.

However this work is not obvious at all. Some questions have to be answered before making any update: What information is contented in the former ESPON Database? Is all information interesting? If not, which indicators have to be integrated? How making possible the concordance between the two different metadata template? ... and so on.

The aim of this paper is to try to establish some milestones in this way. In a first section, we will make a global overview of the previous ESPON Database. Considering this expertise, we will propose a strategy for the update of indicators. Finally, we will discuss about the degree of precision of the update considering the possibilities which exist with the new metadata template.



# 1 Overview of the the ESPON 2006 Database

In concrete terms, ESPON 2006 Database is developed in two parts: an “internal” database, dedicated to ESPON Projects during the ESPON Program and an “external” one where people can download datasets with a limited number of constraints.

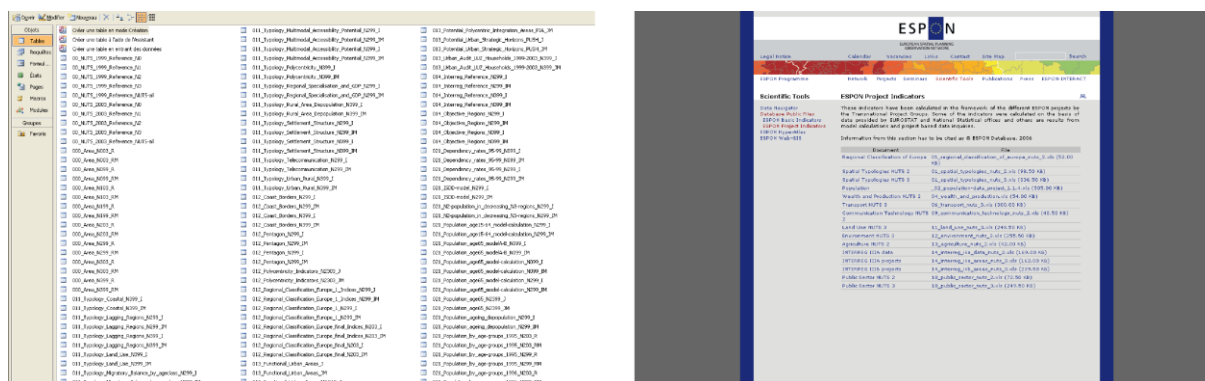


Figure 1 – The “Internal” (on left) and “external” (on right) ESPON 2006 Databases

## 1.1 The “internal” ESPON 2006 Database

UMS RIATE analysed the content of the internal ESPON 2006 Database to have a global overview of the information available within ESPON in June 2006<sup>1</sup>. The geographical coverage of the ESPON 2006 Database contains the ESPON area during the period of the program, that is to say the current ESPON area minus Iceland and Liechtenstein.

In June 2006, ESPON internal Database contains more than 4350 indicators<sup>2</sup> grouped into 361 tables structured **by themes** (figure 2). Regarding the structuring of the database, population and employment/labour market themes are over represented with 1036 and 1458 indicators. Conversely, some themes contain few indicators, such as tourism or communication (less than 20 indicators)

<sup>1</sup> It depicts the situation in June 2006. Some indicators has been added since this period (last version = April 2007). However, we can consider that these modifications do not change considerably the structure of this internal database.

<sup>2</sup> We consider that an indicator is a column of data in a table. For instance, if GDP in euros is described for NUTS0, NUTS1, NUTS2 and NUTS3 we consider that there is 4 indicators.

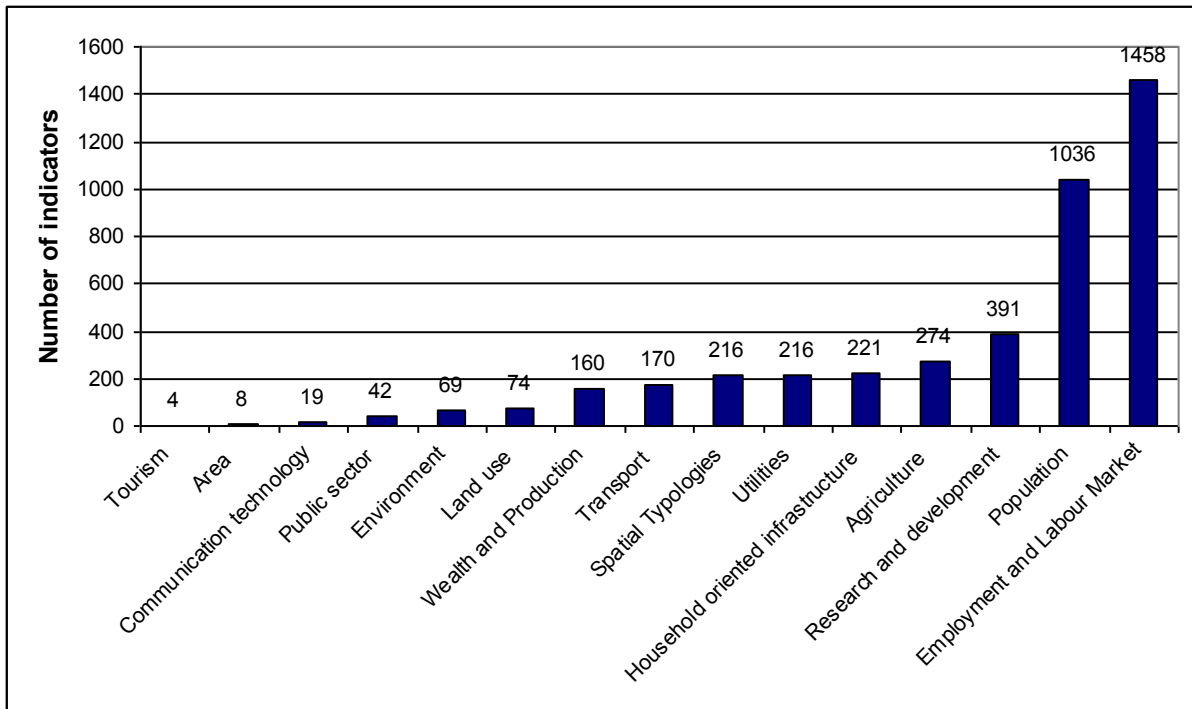


Figure 2 – Number of indicators by themes into the internal ESPON 2006 Database

The analysis of the quantity of information by **type of NUTS** (figure 3) reveals that most of the indicators have been collected in NUTS 2 (more than 3000 indicators, 70 % of the total) and in the NUTS 1999 version (3200 indicators, 74 % of the total).

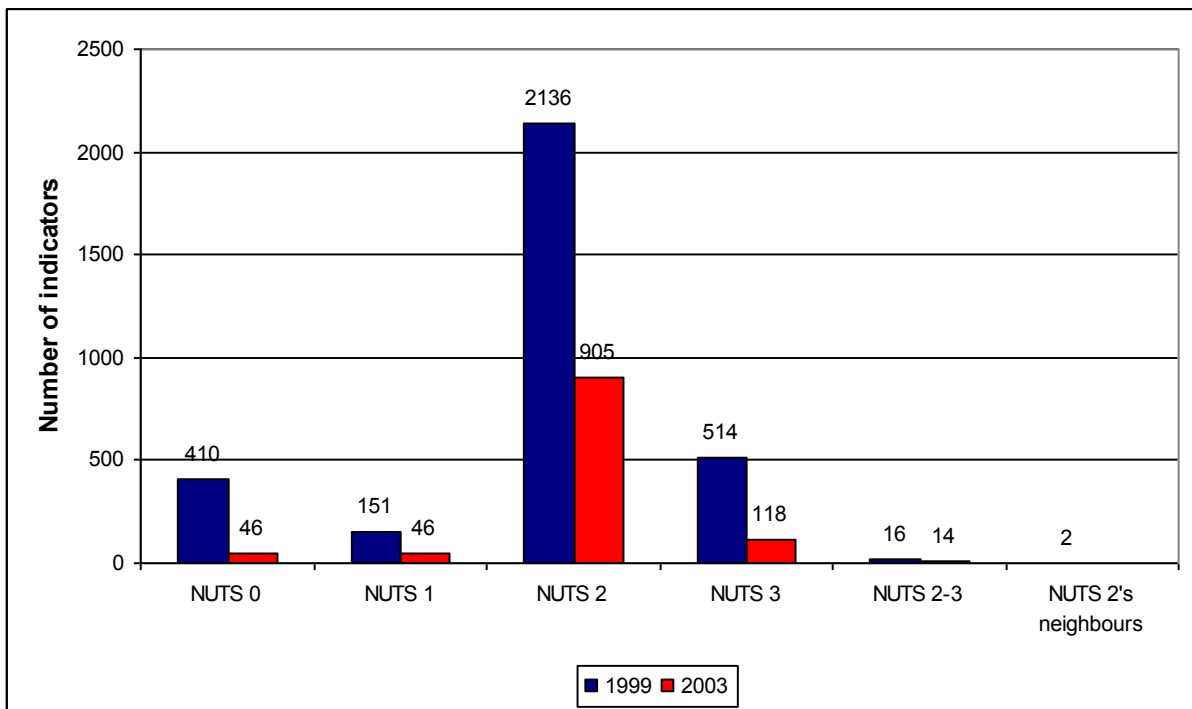


Figure 3 – Number of indicators by type of NUTS into the internal ESPON 2006 Database

Most of the indicators contained in the ESPON 2006 Database have a **period of reference** comprised from 1999 to 2003, 60 % of the total (figure 4). The Database proposed also some indicators described by an evolution. It is possible to have data from 1974 but the degree of completeness is small.

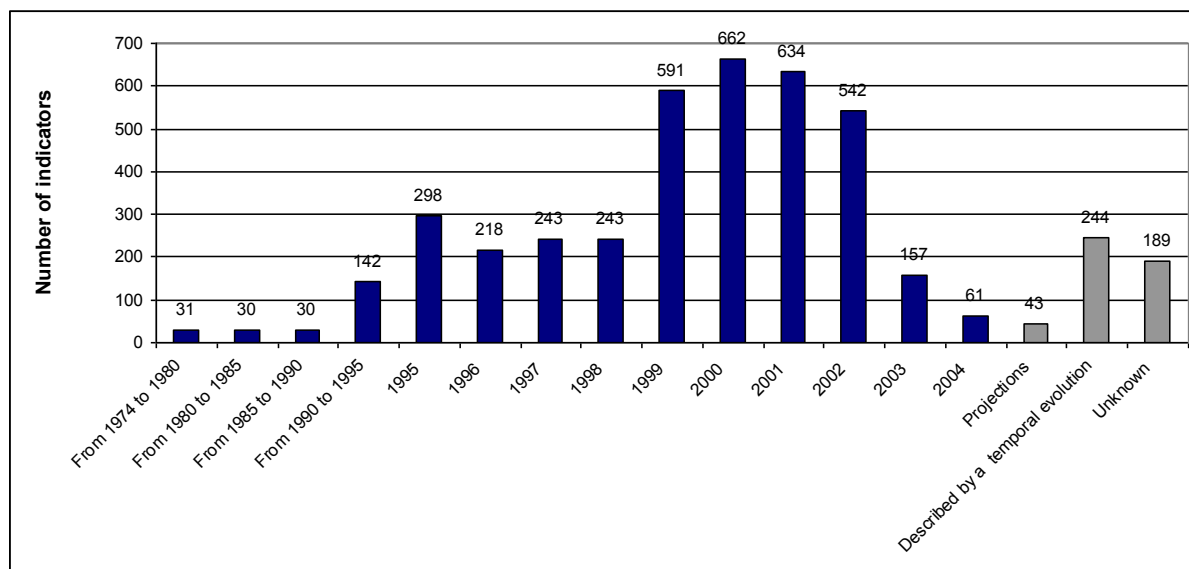


Figure 4 – Period of reference of indicators into the internal ESPON 2006 Database

Considering the **origin of the data**, 78 % of the indicators have a data source coming from Eurostat (Eurostat Regio mainly). The other indicators come from ESPON projects directly (typologies, calculation of innovative indicators) and other data centers, like the European Environmental Agency.

**The degree of completeness** of values in the different indicators of the database is very heterogeneous (figure 5). The main problem raised by researchers during the last ESPON program considering the database is that it was impossible to have an overview of the quality of information before downloading the dataset. Indeed, most of the indicators are described by a very poor completeness. Around 50 % of the indicators available can be characterised by too important missing values – no data, very poor or poor quality – to be used in the all ESPON area.

However, some indicators present a good degree of completeness. But it represents a minor part of the database (around 40 %) It is particularly the case for indicators described in NUTS0. In NUTS2 and NUTS3 the thematic field where this quality is the most important is wealth and production and population. In NUTS 2 or NUTS 3, most of the datasets are described by missing values in French overseas territories, Canaries, Madeira or Ceuta and Mellila.

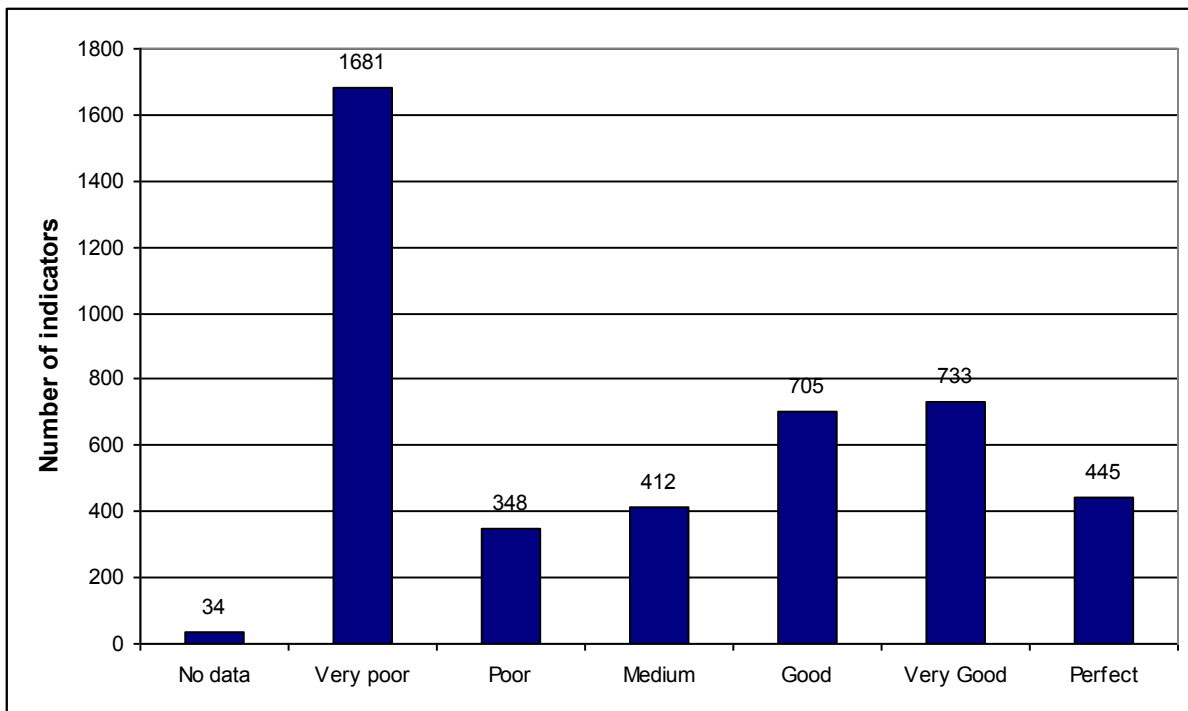


Figure 5 – Degree of completeness<sup>3</sup> of indicators into the internal ESPON 2006 Database

## 1.2 The “external” ESPON 2006 Database

The so called “external” ESPON 2006 Database is available on ESPON website<sup>4</sup> and is divided in two parts: “Basic indicators” and “project indicators”. Basically the main difference between these two fields is that “Basic indicators” coming mainly from Eurostat (excepted for Switzerland and Norway where data come from National Statistical Offices) and has being made in form by ESPON community; on the contrary “project indicators” can be considered as an ESPON Projects production (new indicators, synthetic typologies and so on).

When considering the content of the database by **thematic field** (figure 6), we can notice that population concentrate the most part of the information (80 indicators coming from ESPON, 21 coming from Eurostat). But ESPON makes available a lot of synthetic typologies (47) which summarizes the scientific support of ESPON program to European statistical information. It shows also the synthesis work that has been done during the previous ESPON program.

<sup>3</sup> This classification is a result of a typology: “Perfect” means that 100 % of the data are available; “very good” means that there is a missing values for 1 country; “good” for 2-3 countries; medium for 4-5 countries; “poor” for 6-7 countries and “very poor” for more than 8 countries. The aim of this typology is not to be objective at all (specific cases appear regularly) but to give a global overview of the quality of information available on ESPON 2006 Database.

<sup>4</sup> [http://www.espon.eu/mmp/online/website/content/tools/832/index\\_EN.html](http://www.espon.eu/mmp/online/website/content/tools/832/index_EN.html)

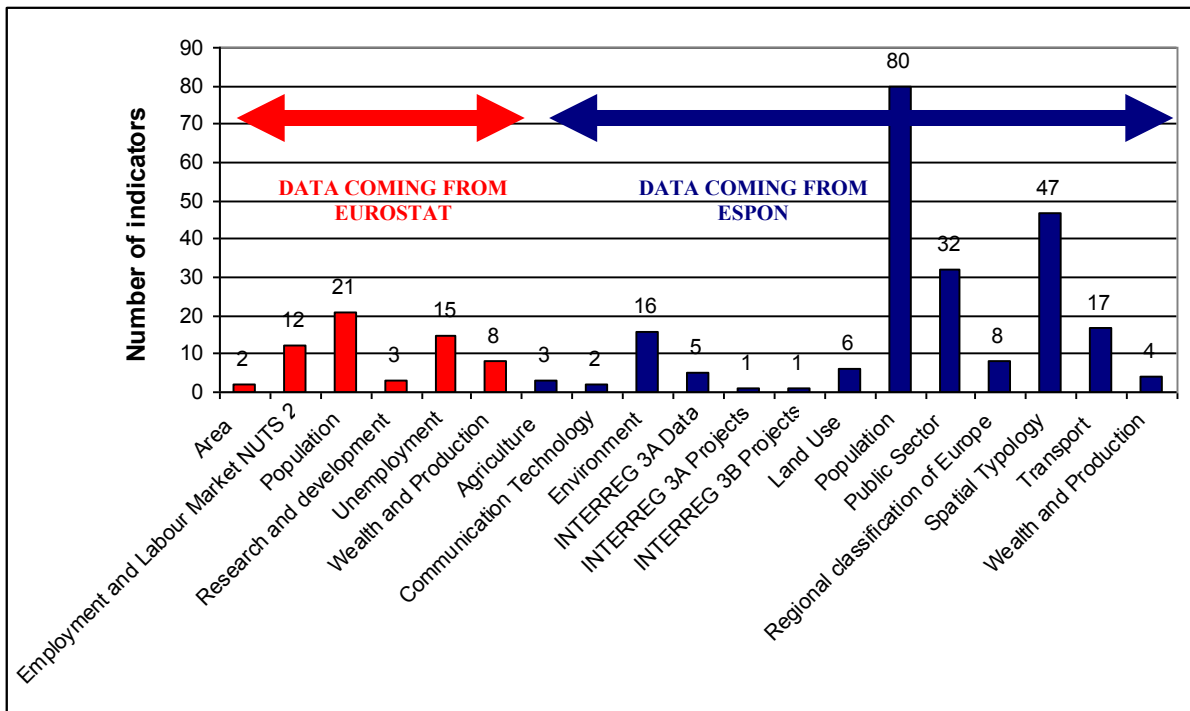


Figure 6 – Number of indicators by themes into the external ESPON 2006 Database

The geographical scope of ESPON 2006 Database available on line is the **regional one**. It contains data for NUTS 2 or NUTS 3 level (or a mix of NUTS 1-2-3). Most indicators are described in NUTS 1999 version and the main geographical level of analysis is NUTS 2.

	NUTS VERSION	
	NUTS 1999	NUTS 2003
NUTS2	126	33
NUTS3	89	23
Mix	12	0

Table 1 – Number of indicators by type of NUTS into the external ESPON 2006 Database

**The degree of completeness** of this database is great since most of the indicators are above 90 % (figure 7).

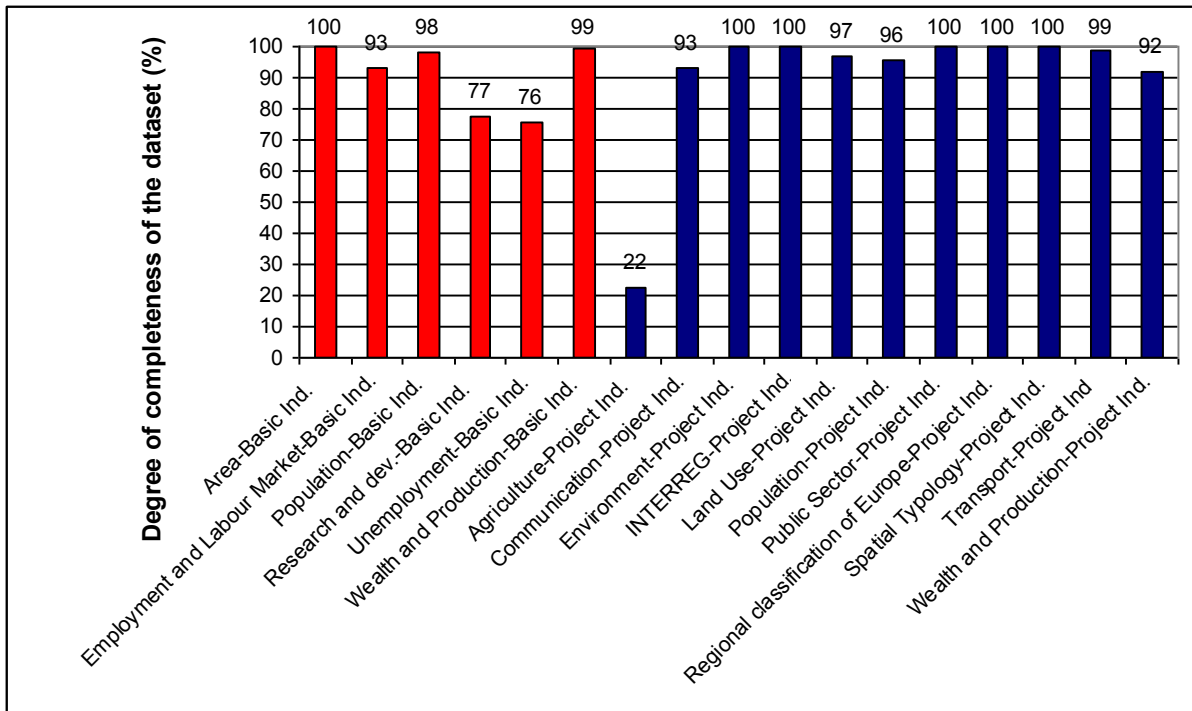


Figure 7 – Degree of completeness of the datasets of the external ESPON 2006 database by themes



## 2 What strategy for updating the ESPON 2006 Database?

### 2.1 Trying to integrate the external ESPON 2006 Database (available on the ESPON Website)

ESPON 2013 Database project proposes to focus the work of integration of indicators from the previous program into the external ESPON 2006 Database. Regarding to the part 1 of the paper, different reasons explain this choice:

- The **degree of completeness** of the datasets is significantly higher in the external than in the internal one. It does not make sense to update indicators which are characterised by 80 % of missing values.
- The fact that indicators from ESPON projects have been made available to the external world shows that the reflection needed for choosing the **most pertinent** indicators produced by the program has already been done.
- It is impossible to integrate in a short term and in a good way more than **4 000 indicators** (internal ESPON Database) and continue to integrate data in a new database structure. The risk is high to make confusions.

### 2.2 Integrate both Eurostat indicators and ESPON indicators

As shown in the previous section, the origin of the data of ESPON 2006 Database is heterogeneous. However, data come mainly from Eurostat and ESPON projects. ESPON 2013 Database project proposes to integrate both indicators developed by ESPON Projects (called "ESPON Project indicators" in ESPON website) and collected from Eurostat website (called "basic indicators" in ESPON website).

The integration of selected indicators from ESPON 2006 Projects (ESPON Project indicators) does not raise any questions for us: they are the result of innovative methodology using knowledge in statistics and spatial analysis (spatial typologies, regional classification of Europe, estimations of missing values). This data has consequently have to be integrated in the ESPON Database in that form.

However, the question is more open for data coming from Eurostat. We have decided to store this information taking into account that it is not possible to obtain historical data on the Eurostat website. Eurostat proposes only data in the current NUTS version (e.g. NUTS 2006 version). It is possible to download this historical data by using the NewCronos Database (the historical database from Eurostat), which is not obvious for basic end-users (it requires using sql queries...). But it is important to keep in mind that indicators proposed by Eurostat are regularly revised. As an example we have compared the difference of values between two sources for a same indicator (total population in 2003): ESPON 2006 Database (origin of data: Eurostat, 2005) and Eurostat in 2009 (figure 8). If the data are the same for most of the NUTS2, 25 % of the values of total population in 2003 have been re-estimated (France, Slovakia, United Kingdom) since 2005. Indeed, considering that Eurostat revised regularly its basic indicators (population, employment, production), it is better to update these indicators by using directly official sources (national statistical offices, Eurostat). It implies that some values stored in the ESPON 2013 Database and coming from Eurostat are not necessary the same as the ones that are downloadable at present.

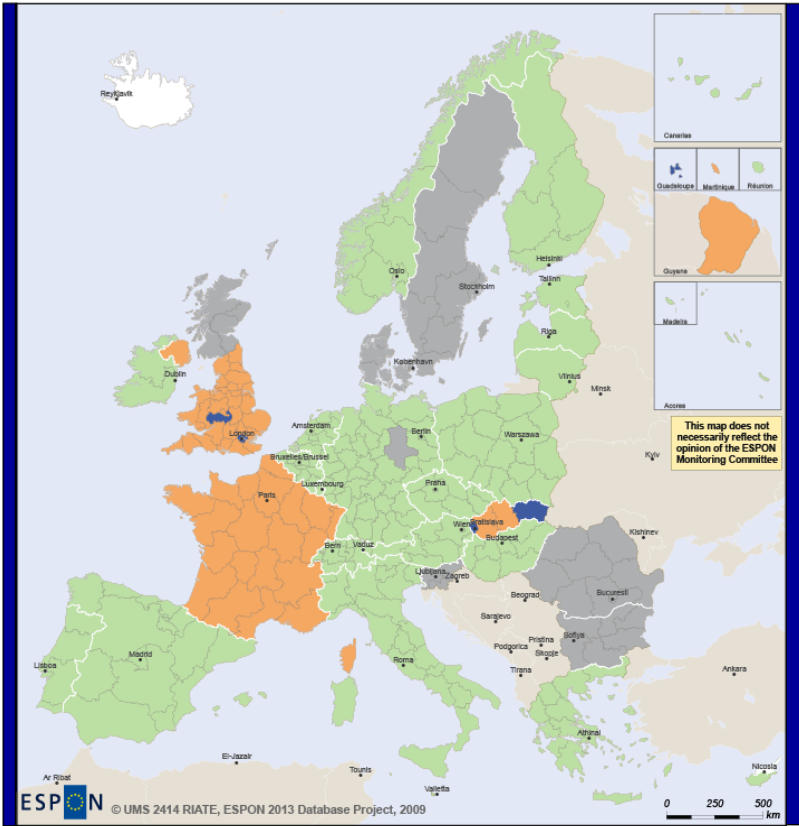


Figure 8 – NUTS 2 where population in 2003 has been re-estimated according to Eurostat

Source: ESPON 2006 Database, Eurostat, 2009  
 Source for administrative boundaries: UMS RIATE  
 Origin of data: © European Communities, 2009  
 NUTS 2 - 2006

Population 2003 (Eurostat 2009) / Population 2003 (ESPON 2006 derived from Eurostat)

- Re-estimation (population according Eurostat 2009 > population according ESPON 2006)
- Same values
- Re-estimation (population according Eurostat 2009 < population according ESPON 2006)
- No information (change of NUTS version)

## 2.3 Description of datasets who will integrate the ESPON DB 2013

More concretely, the ESPON 2013 Database Project proposes to integrated indicators coming from the ESPON 2006 External Database. As a whole, this concerns 15 datasets (including 198 indicators) coming from ESPON Project indicators and 10 datasets (including 54 indicators) coming from ESPON Basic indicators (figure 9<sup>5</sup>)

### ESPON Project Indicators

These indicators have been calculated in the framework of the different ESPON projects by the Transnational Project Groups. Some of the indicators were calculated on the basis of data provided by EUROSTAT and National Statistical offices and others are results from model calculations and project based data inquiries.

Information from this section has to be cited as © ESPON Database, 2006

Name	Type	Size
Regional Classification of Europe	XLS	52,00 KB
Spatial Typologies NUTS 2	XLS	98,50 KB
Spatial Typologies NUTS 3	XLS	336,50 KB
Population	XLS	505,00 KB
Wealth and Production NUTS 2	XLS	54,00 KB
Transport NUTS 3	XLS	300,00 KB
Communication Technology NUTS 2	XLS	40,50 KB
Land Use NUTS 3	XLS	248,50 KB
Environment NUTS 3	XLS	255,50 KB
Agriculture NUTS 2	XLS	42,00 KB
INTERREG IIIA data	XLS	169,00 KB
INTERREG IIIA projects	XLS	162,00 KB
INTERREG IIIB projects	XLS	239,50 KB
Public Sector NUTS 2	XLS	72,50 KB
Public Sector NUTS 3	XLS	249,50 KB

### ESPON Basic indicators

The tables presented in this section include basic information on population, employment, unemployment and economic output. This information is mainly based on regional statistics from EUROSTAT. It covers the regions of the EU25 and Candidate Countries and in some cases it has been complemented with additional data from National Statistics Offices (NSO) in order to cover the entire ESPON territory (EU 25 + Romania and Bulgaria + Switzerland and Norway).

Information from this section has to be cited as © ESPON Database, 2006

Name	Type	Size
Area NUTS 2	XLS	37,50 KB
Area NUTS 3	XLS	129,50 KB
Population NUTS 2	XLS	112,50 KB
Population NUTS 3	XLS	219,50 KB
Employment and Labour Market NUTS 2	XLS	88,00 KB
Unemployment NUTS 2	XLS	52,00 KB
Unemployment NUTS 3	XLS	189,00 KB
Wealth and Production NUTS 2	XLS	51,50 KB
Wealth and Production NUTS 3	XLS	188,00 KB
Research and Development NUTS 2	XLS	44,50 KB

Figure 9 – Targeted datasets to update in the ESPON 2013 Database

[http://www.espon.eu/main/Menu\\_ScientificTools/ESPON2006Tools/DatabasePublicFiles/projectindicators.html](http://www.espon.eu/main/Menu_ScientificTools/ESPON2006Tools/DatabasePublicFiles/projectindicators.html)

<sup>5</sup> Listing and description of all indicators contented in these datasets in annexes.

### 3. Integration of the ESPON 2006 Database in the 2013 DB architecture - a real added value.

The integration of ESPON 2006 indicators to the ESPON 2013 Database infrastructure implies to adapt the data and metadata model to the new requirements. In some cases it implies (3.1) to add information to the metadata of the ESPON 2006 Database, which is not fully compliant with the ISO-19115 norm. In other cases, thanks to information available in the different ESPON Reports and metadata mentioned in the datasets downloadable on ESPON Website, it is possible to improve the quality of metadata (3.2). Finally, the metadata and data models proposed by the ESPON 2013 Database Project makes possible a better structuring of the information contained in the metadata of ESPON 2006 indicators (3.3)

#### 3.1. Elements added by the ESPON 2013 Database Project to the metadata of ESPON 2006 indicators

In relation with the data and metadata models proposed by the ESPON 2013 Project, some elements were not described in the datasets available on ESPON Website. This has been corrected for the following elements:

- **Metadata point of contact (figure 10).** The metadata template requires a contact person. Considering that it is complicated to find valid contact 4 years after the end of these projects, we have decided to put ESPON Coordination Unit as a metadata point of contact.

Metadata point of contact	
name	ESPON
email	<a href="mailto:info@espon.eu">info@espon.eu</a>
organization	ESPON Coordination Unit
function	Project launcher
role	Data collector

Figure 10 – Adding information concerning point of contact and upload date

- **Date of the scope (figure 11):** Defining the precise date of upload allows defining when the data have been published. Thanks to this it is possible to see immediately that it corresponds to the first ESPON Programme. Considering that this information is not available in the dataset we have chosen to indicate “May 2006” as an upload date.

It corresponds more or less to the end of scientific activities in the first ESPON Program. This date can be changed if necessary.

- **Data provider (figure 11):** for each datasets uploaded, we have considered “ESPON 2006 Database” as a first level data provider and the real data provider (ESPON Projects, Eurostat etc.) as a second level data provider, in the “estimation methodology” field. As an example the data provider of a given scope could be “ESPON 2006 Database”, and in the estimation methodology field it could be specified that this ESPON 2006 DB data came primarily from Eurostat (regional database) or another data source (National Statistical Offices of Switzerland for instance).

1		
<b>label lineage</b>		
<b>provider</b>	ESPON 2006 Database	
<b>date</b>	05/2006	
<b>URL</b>		
<b>methodology</b>	Data comes from Eurostat (Regional database) but are described in an outdated NUTS version. Consequently they are not downloadable in Eurostat website.	
<b>methodology URI</b>		
<b>reliability</b>		
<b>estimation quality</b>	high	FAUX
<b>constraints</b>		
<b>public data</b>		VRAI
<b>public methodology</b>		VRAI
<b>copyrights</b>	© ESPON Database, 2006	
2		
<b>label lineage</b>		
<b>provider</b>	ESPON 2006 Database	
<b>date</b>	05/2006	
<b>URL</b>		
<b>methodology</b>	Data comes from the National Statistical Office of Switzerland (Office fédéral de la Statistique)	
<b>methodology URI</b>		
<b>reliability</b>		
<b>estimation quality</b>	high	FAUX
<b>constraints</b>		
<b>public data</b>		VRAI
<b>public methodology</b>		VRAI
<b>copyrights</b>	© ESPON Database, 2006	

Figure 11 – Definition of date and provider of ESPON 2006 datasets scope

### 3.2. Improvements of the metadata and data information

Considering the different fields that have to be filed in the metadata and data templates, it allows to envisage new possibilities for using the ESPON 2006 indicators

- Precise the methodology (figure 12).** For some indicators, as typologies, the methodology description has been improved. Thanks to the information contented in the final reports of concerned projects, we have added substantial elements in the methodology field.

Identification	
code	RCECLI
name	RCE - classified lisbon performance
units	
abstract	Regional Classification of Europe - classified lisbon performance
methodology classification	<b>INDICATORS</b> Degree of Lisbon performance as an aggregate of 5 indicators: - Productivity (GDP per person employed 2000) (+) - Employment rate (Employed population / population aged 15-64 2003) (+) - Expenditure on R&D (Expenditure on R&D / Total GDP 2001) (+) - R&D Business Enterprise Sector (Personnel / 1000 active person 2001) (+) - High educated population (Highly educated population / total educated population 2002) (+)
	<b>TRANSFORMATION</b> Aggregate of z-transformed indicator-values by thematic field + classified on the basis of mean value and standard deviation
	<b>CLASSES</b> 1=highly below average; 2=below average; 3=average; 4=above average; 5=highly above average.
	<b>theme</b> <b>keywords</b>
	Typology Lisbon

Figure 12 – More precise information in the methodology of the indicator

- Precise the data source (figure 13).** Thanks to the new data and metadata models, it is possible to integrate the URL of the ESPON Report linked to the data calculation, which is really a value added for valorise ESPON results.

scope	
label	1
lineage	
provider	ESPON 2006 Database
date	05/2006
URL	<a href="http://www.espon.eu/export/sites/default/Documents/Projects/ESPON2006Projects/CoordinatingCrossThematicProjects/Coordination/fr-3.1-full.pdf">http://www.espon.eu/export/sites/default/Documents/Projects/ESPON2006Projects/CoordinatingCrossThematicProjects/Coordination/fr-3.1-full.pdf</a>
methodology	Author: Das Bundesinstitut für Bau-, Stadt- und Raumforschung (BBR)
reliability	
estimation	FAUX
quality	high
constraints	
public dat	VRAI
public met	VRAI
copyright	© ESPON Database, 2006

Figure 13 – Identification of the source

### 3.3. New metadata structure for existing information

The metadata integrated in the ESPON 2006 Database template are done in “free text” fields. Thanks to the new metadata model, it is possible to better structure these fields (figure 14):

1. For each territorial unit contained in the dataset, the NUTS level and the NUTS version has been identified.

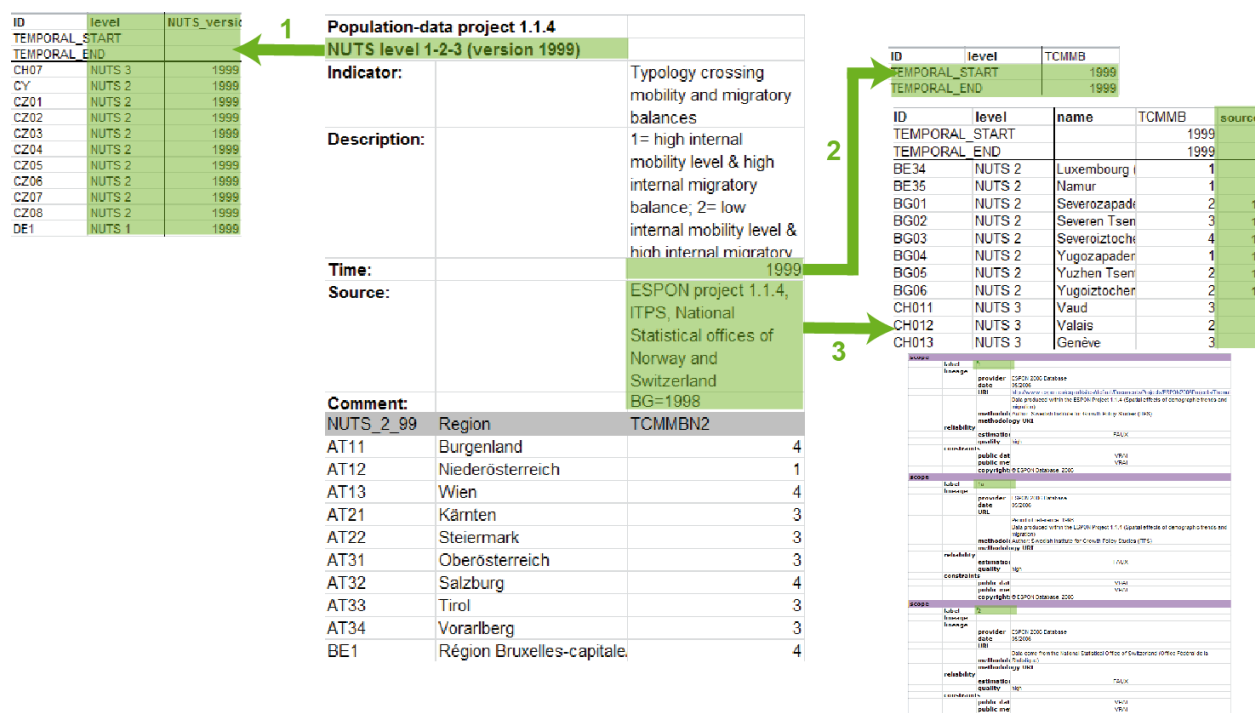


Figure 14 – Integration of metadata in the 2013 database structure (theoretical example)

2. The temporal validity of the indicator has been indicated in dedicated field
3. All sources and comment have been better integrated (column source which is linked to the metadata sheet where values are precisely described).

Consequently, it is now possible for the end-user to better query the ESPON 2006 database:

- What data are the available in the NUTS 1999 delineation?
- What are the available indicators with a temporal extent starting in 1999 and finishing in 2000?
- What is the information available coming from ESPON 1.1.4 Project?

These basic questions, typically adapted to the content of the ESPON 2006 Database will be possible in a near future.

Last but not least, this work makes possible to integrate the results from ESPON 2006 and ESPON 2013 Programs in the same database architecture.

## APPENDIX 1 - Updated indicators from the ESPON 2006 Database – Data from TPG’s (ESPON Project Indicators)

DATASET	NAME OF INDICATOR	Geog. Object	Temporal start/end	SOURCE (main)
01_regional_classification_of_europe_nuts_2_data	RCE - classified economy	NUTS2 (version 2003)	2002/2002	ESPON 2.4.2
01_regional_classification_of_europe_nuts_2_data	RCE - classified lisbon performance	NUTS2 (version 2003)	2002/2002	ESPON 2.4.2
01_regional_classification_of_europe_nuts_2_data	RCE - classified labour market	NUTS2 (version 2003)	2003/2003	ESPON 2.4.2
01_regional_classification_of_europe_nuts_2_data	RCE - classified demography	NUTS2 (version 2003)	2002/2002	ESPON 2.4.2
01_regional_classification_of_europe_nuts_2_data	RCE - classified naturalness	NUTS2 (version 2003)	2000/2000	ESPON 2.4.2
01_regional_classification_of_europe_nuts_2_data	RCE - classified natural hazards	NUTS2 (version 2003)	2002/2002	ESPON 2.4.2
01_regional_classification_of_europe_nuts_2_data	RCE - classified technological hazards	NUTS2 (version 2003)	2002/2002	ESPON 2.4.2
01_regional_classification_of_europe_nuts_2_data	RCE - classified accessibility	NUTS2 (version 2003)	2001/2001	ESPON 2.4.2
01_spatial_typologies_nuts_2	Settlement Structure Typology	NUTS2 (version 1999)	1999/1999	ESPON project 3.1
01_spatial_typologies_nuts_2	A typology of levels of household telecommunications uptake	NUTS2 (version 1999)	2002/2002	ESPON project 1.2.2
01_spatial_typologies_nuts_2	A typology of estimated levels of business telecommunications access and uptake	NUTS2 (version 1999)	2002/2002	ESPON project 1.2.2
01_spatial_typologies_nuts_2	A typology comparing levels of household and business telecommunications uptake	NUTS2 (version 1999)	2002/2002	ESPON project 1.2.2
01_spatial_typologies_nuts_2	An overall typology of combined household and business telecommunications development	NUTS2 (version 1999)	2002/2002	ESPON project 1.2.2
01_spatial_typologies_nuts_2	Typology Multimodal Accessibility Potential	NUTS2 (version 1999)	2001/2001	ESPON project 2.1.1
01_spatial_typologies_nuts_2	Typology of lagging regions	NUTS2 (version 1999)	2000/2000	ESPON project 2.1.1
01_spatial_typologies_nuts_2	Typologies of regional specialisation and GDP per capita 2001	NUTS2 (version 1999)	2001/2001	ESPON project 1.1.3
01_spatial_typologies_nuts_2	Typologies of regional specialisation and GDP per capita 1995-2001	NUTS2 (version 1999)	1995/2001	ESPON project 1.1.3
01_spatial_typologies_nuts_2	Coast	NUTS2 (version 1999)	2003/2003	ESPON project 2.1.1
01_spatial_typologies_nuts_2	Border	NUTS2 (version 1999)	2003/2003	ESPON project 2.1.1
01_spatial_typologies_nuts_2	Pentagon EU 15	NUTS2 (version 1999)	2003/2003	ESPON project 2.1.1
01_spatial_typologies_nuts_2	Pentagon EU 27 plus 2	NUTS2 (version 1999)	2003/2003	ESPON project 2.1.1
01_spatial_typologies_nuts_2	Part of Interreg	NUTS2	2000/2000	ESPON



	North-Sea Programme	(version 1999)		project 3.1
01_spatial_typologies_nuts_2	Part of Interreg CADSES Programme	NUTS2 (version 1999)	2000/2000	ESPON project 3.1
01_spatial_typologies_nuts_2	Part of Interreg Atlantic-Area Programme	NUTS2 (version 1999)	2000/2000	ESPON project 3.1
01_spatial_typologies_nuts_2	Part of Interreg Programme "Non continental and overseas cooperation areas"	NUTS2 (version 1999)	2000/2000	ESPON project 3.1
01_spatial_typologies_nuts_2	Part of Interreg Programme "Northern-Peripherie"	NUTS2 (version 1999)	2000/2000	ESPON project 3.1
01_spatial_typologies_nuts_2	Part of Interreg Alpine-Space Programme	NUTS2 (version 1999)	2000/2000	ESPON project 3.1
01_spatial_typologies_nuts_2	Part of Interreg Programme "Archimedes"	NUTS2 (version 1999)	2000/2000	ESPON project 3.1
01_spatial_typologies_nuts_2	Part of Interreg Programme "Baltic Sea"	NUTS2 (version 1999)	2000/2000	ESPON project 3.1
01_spatial_typologies_nuts_2	Part of Interreg Programme "Medoc-Area"	NUTS2 (version 1999)	2000/2000	ESPON project 3.1
01_spatial_typologies_nuts_2	Part of Interreg Programme "South-West-Europe"	NUTS2 (version 1999)	2000/2000	ESPON project 3.1
01_spatial_typologies_nuts_2	Part of Interreg Programme "North-West-Europe"	NUTS2 (version 1999)	2000/2000	ESPON project 3.1
01_spatial_typologies_nuts_2	"Objective 1" regions= regions situated within objective 1 regions	NUTS2 (version 1999)	2000/2000	ESPON project 3.1
01_spatial_typologies_nuts_2	Objective 2 regions includes regions containing at least one Objective 2 region (partly)	NUTS2 (version 1999)	2000/2000	ESPON project 3.1
01_spatial_typologies_nuts_3	Typology Settlement Structure	NUTS2 (version 1999)	1999/1999	ESPON project 3.1
01_spatial_typologies_nuts_3	Typology Multimodal Accessibility Potential	NUTS2 (version 1999)	2001/2001	ESPON project 2.1.1
01_spatial_typologies_nuts_3	Typologie of lagging regions	NUTS2 (version 1999)	2000/2000	ESPON project 2.1.1
01_spatial_typologies_nuts_3	Urban-rural typology	NUTS2 (version 1999)	1999/1999	ESPON project 1.1.2
01_spatial_typologies_nuts_3	Coast	NUTS2 (version 1999)	2003/2003	ESPON project 2.1.1
01_spatial_typologies_nuts_3	Border	NUTS2 (version 1999)	2003/2003	ESPON project 2.1.1
01_spatial_typologies_nuts_3	Pentagon EU 15	NUTS2 (version 1999)	2003/2003	ESPON project 2.1.1
01_spatial_typologies_nuts_3	Pentagon EU 27 plus 2	NUTS2 (version 1999)	2003/2003	ESPON project 2.1.1
01_spatial_typologies_nuts_3	Part of Interreg North-Sea Programme	NUTS2 (version 1999)	2000/2000	ESPON project 3.1
01_spatial_typologies_nuts_3	Part of Interreg CADSES Programme	NUTS2 (version 1999)	2000/2000	ESPON project 3.1
01_spatial_typologies_nuts_3	Part of Interreg Atlantic-Area Programme	NUTS2 (version 1999)	2000/2000	ESPON project 3.1
01_spatial_typologies_nuts_3	Part of Interreg Programme "Non continental and overseas cooperation	NUTS2 (version 1999)	2000/2000	ESPON project 3.1

	areas"			
01_spatial_typologies_nuts_3	Part of Interreg Programme "Northern-Peripherie"	NUTS2 (version 1999)	2000/2000	ESPON project 3.1
01_spatial_typologies_nuts_3	Part of Interreg Alpine-Space Programme	NUTS2 (version 1999)	2000/2000	ESPON project 3.1
01_spatial_typologies_nuts_3	Part of Interreg Programme "Archimedes"	NUTS2 (version 1999)	2000/2000	ESPON project 3.1
01_spatial_typologies_nuts_3	Part of Interreg Programme "Baltic Sea"	NUTS2 (version 1999)	2000/2000	ESPON project 3.1
01_spatial_typologies_nuts_3	Part of Interreg Programme "Medoc-Area"	NUTS2 (version 1999)	2000/2000	ESPON project 3.1
01_spatial_typologies_nuts_3	Part of Interreg Programme "South-West-Europe"	NUTS2 (version 1999)	2000/2000	ESPON project 3.1
01_spatial_typologies_nuts_3	Part of Interreg Programme "North-West-Europe"	NUTS2 (version 1999)	2000/2000	ESPON project 3.1
01_spatial_typologies_nuts_3	"Objective 1" regions= regions situated within objective 1 regions	NUTS2 (version 1999)	2000/2000	ESPON project 3.1
01_spatial_typologies_nuts_3	Objective 2 regions includes regions containing at least one Objective 2 region (partly)	NUTS2 (version 1999)	2000/2000	ESPON project 3.1
04_wealth_and_production	GDP per capita in PPS	NUTS2 (version 1999)	1999/1999	ESPON project 3.1
04_wealth_and_production	GDP per capita in PPS- Global deviation (EU15)	NUTS2 (version 1999)	1999/1999	ESPON project 3.1
04_wealth_and_production	GDP per capita in PPS- Medium deviation	NUTS2 (version 1999)	1999/1999	ESPON project 3.1
04_wealth_and_production	GDP per capita in PPS- Local deviation	NUTS2 (version 1999)	1999/1999	ESPON project 3.1
02_population-data_project_1.1.4_N2	Typology of migratory balances by age classes	NUTS2 (version 1999)	1995/2000	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Dependency rate	NUTS2 (version 1999)	1995/1995.1999/1999	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Share of NUTS 2 average population living in NUTS 3 regions with population decline	NUTS2 (version 1999)	1995/1999	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Share of NUTS 2 area comprising NUTS 3 regions with population decline	NUTS2 (version 1999)	1995/1999	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Population with 65 and more years (%) (Model A)	NUTS2 (version 1999)	2000/2000. 2025/2025, 2050/2050	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Population with 65 and more years (%) (Model B0)	NUTS2 (version 1999)	2000/2000. 2025/2025, 2050/2050	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Population with 65 and more years (%) (Model B1)	NUTS2 (version 1999)	2000/2000. 2025/2025, 2050/2050	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Population with 65 and more years (%) (Model B2)	NUTS2 (version 1999)	2000/2000.2025/2025, 2050/2050	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Population with 65 and more years (%) (Model B3)	NUTS2 (version 1999)	2000/2000. 2025/2025, 2050/2050	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Average score on indirect "ageing"/ "depopulating" indicators	NUTS2 (version 1999)	2000/2000	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Average score on indirect "ageing"/ "depopulating"	NUTS2 (version 1999)	2000/2000	ESPON project 1.1.4

	indicators, Grouped (quartiles)			
02_population-data_project_1.1.4_N2	National Total Fertility Rates	NUTS2 (version 1999)	1999/2000	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Ageing Population (4 groups) 65+/Tot.	NUTS2 (version 1999)	2000/2000	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Ageing "Labour Force" (4 groups) 55-64/20-64	NUTS2 (version 1999)	2000/2000	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	"Labour Force" Replacement (4 groups) 10-19/55-64	NUTS2 (version 1999)	2000/2000	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Post-Active Dependency (4 groups) 65+/20-64	NUTS2 (version 1999)	2000/2000	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Aged People vs. Youth (4 groups) 65+/15-24	NUTS2 (version 1999)	2000/2000	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Share of children (4 groups) 0-14/Tot.pop	NUTS2 (version 1999)	2000/2000	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Changes in Natural Growth Potential (4 groups) 20-29 years in 2020 (born 1991-2000)/20-29 years in 2000 (born 1971-1980)	NUTS2 (version 1999)	2000/2020	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Ageing Population (indexes) 65+/Tot.	NUTS2 (version 1999)	2000/2000	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Ageing "Labour Force" (indexes) 55-64/20-64	NUTS2 (version 1999)	2000/2000	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	"Labour Force" Replacement (indexes) 10-19/55-64	NUTS2 (version 1999)	2000/2000	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Post-Active Dependency (indexes) 65+/20-64	NUTS2 (version 1999)	2000/2000	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Aged People vs. Youth (indexes) 65+/15-24	NUTS2 (version 1999)	2000/2000	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Share of children (indexes) 0-14/Tot.pop	NUTS2 (version 1999)	2000/2000	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Changes in Natural Growth Potential (indexes) 20-29 years in 2020 (born 1991-2000)/20-29 years in 2000 (born 1971-1980)	NUTS2 (version 1999)	2000/2020	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Total fertility rate	NUTS2 (version 1999)	1990/1990; 1995/1995; 1999/1999	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	External immigration	NUTS2 (version 1999)	1996/1999	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Migratory balance by regions between	NUTS2 (version 1999)	1996/1999	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Absolute migratory balance	NUTS2 (version 1999)	1996/1999	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Migratory balance 17.5 to 27.5 years old	NUTS2 (version 1999)	1995/2000	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Migratory balance 32.5 to 42.5 years old	NUTS2 (version 1999)	1995/2000	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Migratory balance 52.5 to 67.5 years old	NUTS2 (version 1999)	1995/2000	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Synthetic cartography of migratory balances	NUTS2 (version 1999)	1995/2000	ESPON project 1.1.4

	for the main age classes			
02_population-data_project_1.1.4_N2	Variation of the population (%) (Model A)	NUTS2 (version 1999)	2000/2050	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Variation of the population (%) (Model B0)	NUTS2 (version 1999)	2000/2050	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Variation of the population (%) (Model B2)	NUTS2 (version 1999)	2000/2050	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Variation of the population (%) (Model B3)	NUTS2 (version 1999)	2000/2050	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Population between 15 and 64 years (%) (Model A)	NUTS2 (version 1999)	2000/2000; 2025/2025; 2050/2050	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Population between 15 and 64 years (%) (Model B1)	NUTS2 (version 1999)	2000/2000; 2025/2025; 2050/2050	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Population between 15 and 64 years (%) (Model B1)	NUTS2 (version 1999)	2000/2000; 2025/2025; 2050/2050	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Population between 15 and 64 years (%) (Model B2)	NUTS2 (version 1999)	2000/2000; 2025/2025; 2050/2050	ESPON project 1.1.4
02_population-data_project_1.1.4_N2	Population between 15 and 64 years (%) (Model B3)	NUTS2 (version 1999)	2000/2000; 2025/2025; 2050/2050	ESPON project 1.1.4
02_population-data_project_1.1.4_N3	Type of rural area	NUTS3 (version 1999)	2000/2000	ESPON project 1.1.4
02_population-data_project_1.1.4_N3	Relative depopulation, quartiles	NUTS3 (version 1999)	1990/2000	ESPON project 1.1.4
02_population-data_project_1.1.4_N3	Population change	NUTS3 (version 1999)	1990/2000; 1990/1995; 1995/2000	ESPON project 1.1.4
02_population-data_project_1.1.4_N23	Share (%) of population in the ages 65+	NUTS 2-3 (version 1999)	1990/1990; 1995/1995; 1999/1999	ESPON project 1.1.4
02_population-data_project_1.1.4_N123	Typology crossing mobility and migratory balances	NUTS 1-2-3 (version 1999)	1996/1999	ESPON project 1.1.4
02_population-data_project_1.1.4_N123	Internal migratory balance	NUTS 1-2-3 (version 1999)	1996/1999	ESPON project 1.1.4
02_population-data_project_1.1.4_N123	Total migratory balance	NUTS 1-2-3 (version 1999)	1996/1999	ESPON project 1.1.4
02_population-data_project_1.1.4_N123	External migratory balance	NUTS 1-2-3 (version 1999)	1996/1999	ESPON project 1.1.4
02_population-data_project_1.1.4_N123	Internal mobility by region	NUTS 1-2-3 (version 1999)	1996/1999	ESPON project 1.1.4
02_population-data_project_1.1.4_N123	Mobility by region relative to national mobility	NUTS 1-2-3 (version 1999)	1996/1999	ESPON project 1.1.4
02_population-data_project_1.1.4_N23a	Total population development	NUTS 2-3a (version 1999)	1996/1999	ESPON project 1.1.4
02_population-data_project_1.1.4_N23a	Natural population development	NUTS 2-3a (version 1999)	1996/1999	ESPON project 1.1.4
02_population-data_project_1.1.4_N23a	Net migration	NUTS 2-3a (version 1999)	1996/1999	ESPON project 1.1.4
ESPON_2006_06_transport_nuts_3	Number of commercial airports	NUTS3 (version 1999)	2001/2001	ESPON project 1.2.1
ESPON_2006_06_transport_nuts_3	Number of commercial seaports	NUTS3 (version 1999)	2001/2001	ESPON project 1.2.1
ESPON_2006_06_transport_nuts_3	Length of road network (km)	NUTS3 (version 1999)	2001/2001	ESPON project 1.2.1
ESPON_2006_06_transport_nuts_3	Length of railway network, km	NUTS3 (version 1999)	2001/2001	ESPON project 1.2.1

ESPON_2006_06_transport_nuts_3	Traffic in commercial airports	NUTS3 (version 1999)	2001/2001	ESPON project 1.2.1
ESPON_2006_06_transport_nuts_3	Connectivity to commercial airports by car of the capital or centroid representative of the NUTS3	NUTS3 (version 1999)	2001/2001	ESPON project 1.2.1
ESPON_2006_06_transport_nuts_3	Connectivity to commercial seaports by car of the capital or centroid representative of the NUTS3	NUTS3 (version 1999)	2001/2001	ESPON project 1.2.1
ESPON_2006_06_transport_nuts_3	Time to the nearest motorway access, by car of the capital or centroid representative of the NUTS3	NUTS3 (version 1999)	2001/2001	ESPON project 1.2.1
ESPON_2006_06_transport_nuts_3	Daily population accessible by car	NUTS3 (version 1999)	1999/1999	ESPON project 1.2.1
ESPON_2006_06_transport_nuts_3	Daily market accessible by car in terms of GDP	NUTS3 (version 1999)	2000/2000	ESPON project 1.2.1
ESPON_2006_06_transport_nuts_3	Potential accessibility air, ESPON space = 100	NUTS3 (version 1999)	2001/2001	ESPON project 1.2.1
ESPON_2006_06_transport_nuts_3	Potential accessibility rail, ESPON space = 100	NUTS3 (version 1999)	2001/2001	ESPON project 1.2.1
ESPON_2006_06_transport_nuts_3	Potential accessibility road, ESPON space = 100	NUTS3 (version 1999)	2001/2001	ESPON project 1.2.1
ESPON_2006_06_transport_nuts_3	Potential accessibility multimodal, ESPON space = 100	NUTS3 (version 1999)	2001/2001	ESPON project 1.2.1
ESPON_2006_06_transport_nuts_3	Accessibility time to market by road,	NUTS3 (version 1999)	1997/1997	ESPON Project 2.1.1
ESPON_2006_06_transport_nuts_3	Accessibility time to market by rail,	NUTS3 (version 1999)	1997/1997	ESPON Project 2.1.1
ESPON_2006_06_transport_nuts_3	Accessibility time to market by rail and road,	NUTS3 (version 1999)	1997/1997	ESPON Project 2.1.1
ESPON_2006_09_communication_technology_nuts_2	Share of Internet users	NUTS2 (version 1999)	2002/2002	ESPON Project 1.2.2
ESPON_2006_09_communication_technology_nuts_2	Share of firms with own website	NUTS2 (version 1999)	2002/2002	ESPON Project 1.2.2
ESPON_2006_11_land_use_nuts_3	Share of artificial surfaces	NUTS3 (version 1999)	1996/1996	Espon project 1.1.2
ESPON_2006_11_land_use_nuts_3	Share of artificial surfaces per 1000 inh.	NUTS3 (version 1999)	1996/1996	Espon project 1.1.2
ESPON_2006_11_land_use_nuts_3	Share of artificial surfaces per 100 million GDP PPS	NUTS3 (version 1999)	1996/1996	Espon project 1.1.2
ESPON_2006_11_land_use_nuts_3	Share of urban fabric	NUTS3 (version 1999)	1996/1996	ESPON Project 3.1
ESPON_2006_11_land_use_nuts_3	Share of arable land	NUTS3 (version 1999)	1996/1996	ESPON Project 3.1
ESPON_2006_11_land_use_nuts_3	Share of permanent crops	NUTS3 (version 1999)	1996/1996	ESPON Project 3.1
ESPON_2006_12_environment_nuts_3	Occurrence of snow avalanches	NUTS3 (version 1999)	2004/2004	ESPON Project 1.3.1
ESPON_2006_12_environment_nuts_3	Large scale droughts in Europe	NUTS3 (version 1999)	1904/1995	ESPON Project 1.3.1
ESPON_2006_12_environment_nuts_3	Regional earthquake hazard potential	NUTS3 (version 1999)	1998/1998	ESPON Project 1.3.1

		1999)		
ESPON_2006_12_environment_nuts_3	Extreme temperatures	NUTS3 (version 1999)	1961/1990	ESPON Project 1.3.1
ESPON_2006_12_environment_nuts_3	Regional flood hazard potential	NUTS3 (version 1999)	1987/2002	ESPON Project 1.3.1
ESPON_2006_12_environment_nuts_3	Forest fire hazard	NUTS3 (version 1999)	1997/2003	ESPON Project 1.3.1
ESPON_2006_12_environment_nuts_3	Occurrence of landslides	NUTS3 (version 1999)	2004/2004	ESPON Project 1.3.1
ESPON_2006_12_environment_nuts_3	Occurrence of storm surges	NUTS3 (version 1999)	2004/2004	ESPON Project 1.3.1
ESPON_2006_12_environment_nuts_3	Occurrence of tsunami runups and tsunami potential areas in Europe	NUTS3 (version 1999)	-1628/2003	ESPON Project 1.3.1
ESPON_2006_12_environment_nuts_3	Volcanic eruptions during the last 10 000 years	NUTS3 (version 1999)	-1628/2003	ESPON Project 1.3.1
ESPON_2006_12_environment_nuts_3	Approximate probability of having winter storms and for tropical storms probable maximum intensity	NUTS3 (version 1999)	2004/2004	ESPON Project 1.3.1
ESPON_2006_12_environment_nuts_3	Air traffics hazard potential	NUTS3 (version 1999)	1996/2003	ESPON Project 1.3.1
ESPON_2006_12_environment_nuts_3	Chemical plants hazard potential	NUTS3 (version 1999)	2001/2004	ESPON Project 1.3.1
ESPON_2006_12_environment_nuts_3	Potential risk of radioactive contamination on NUTS3 regions	NUTS3 (version 1999)	2003/2003	ESPON Project 1.3.1
ESPON_2006_12_environment_nuts_3	Classification of Oil-SUM values	NUTS3 (version 1999)	2002/2002	ESPON Project 1.3.1
ESPON_2006_12_environment_nuts_3	Sum of all weighted hazard values	NUTS3 (version 1999)	2004/2004	ESPON Project 1.3.1
ESPON_2006_13_agriculture_nuts_2	Share of UAA which is arable	NUTS2 (version 1999)	2001/2001	ESPON Project 2.1.3
ESPON_2006_13_agriculture_nuts_2	Share of UAA that is fallow	NUTS2 (version 1999)	2001/2001	ESPON Project 2.1.3
ESPON_2006_13_agriculture_nuts_2	Share of farm holders aged <65	NUTS2 (version 1999)	1997/1997	ESPON Project 2.1.3
ESPON_2006_14_interreg_iiia_areas_nuts_3	Table of belonging to the 64 INTERREG IIIA Programs	NUTS3 (version 2003)	2000/2000	ESPON
ESPON_2006_14_interreg_iiia_data_nuts_3	Typology of borders in NUTS3 regions participating in INTERREG IIIA Programmes	NUTS3 (version 2003)	2006/2006	ESPON INTERACT/K TH
ESPON_2006_14_interreg_iiia_data_nuts_3	Intensity of projects per INTERREG IIIA Programme	NUTS3 (version 2003)	2006/2006	ESPON INTERACT/K TH
ESPON_2006_14_interreg_iiia_data_nuts_3	Geographic type of land border of NUTS3 INTERREG IIIA programme areas	NUTS3 (version 2003)	2006/2006	ESPON INTERACT/K TH
ESPON_2006_14_interreg_iiia_data_nuts_3	Density of border crossings in INTERREG IIIA areas	NUTS3 (version 2003)	2006/2006	ESPON INTERACT/K TH
ESPON_2006_14_interreg_iiia_data_nuts_3	Economic disparities per programme	NUTS3 (version 2003)	2006/2006	ESPON INTERACT/K TH
ESPON_2006_14_interreg_iiib_areas_nuts_3	Table of belonging to the 14 INTERREG IIIB Programs	NUTS3 (version 1999)	2006/2006	ESPON Project 3.1
ESPON_2006_18_public_sector_nuts_2_3_data	Percentage of regional Pre-	NUTS2&NUTS3 (version 1999)	1998/2000	ESPON Project 2.2.2

	Accession-Aid (PHARE, PHARE CBC, ISPA) addressing capital-supply-potential	1999)		
ESPON_2006_18_public_sector_nuts_2_3_data	Percentage of regional Pre-Accession-Aid (PHARE, PHARE CBC, ISPA) addressing environmental quality	NUTS2&NUT S3 (version 1999)	1998/2000	ESPON Project 2.2.2
ESPON_2006_18_public_sector_nuts_2_3_data	Percentage of regional Pre-Accession-Aid (PHARE, PHARE CBC, ISPA) addressing geographical position	NUTS2&NUT S3 (version 1999)	1998/2000	ESPON Project 2.2.2
ESPON_2006_18_public_sector_nuts_2_3_data	Percentage of regional Pre-Accession-Aid (PHARE, PHARE CBC, ISPA) addressing potential of innovation	NUTS2&NUT S3 (version 1999)	1998/2000	ESPON Project 2.2.2
ESPON_2006_18_public_sector_nuts_2_3_data	Percentage of regional Pre-Accession-Aid (PHARE, PHARE CBC, ISPA) addressing institutional conditions	NUTS2&NUT S3 (version 1999)	1998/2000	ESPON Project 2.2.2
ESPON_2006_18_public_sector_nuts_2_3_data	Percentage of regional Pre-Accession-Aid (PHARE, PHARE CBC, ISPA) addressing labour market potential	NUTS2&NUT S3 (version 1999)	1998/2000	ESPON Project 2.2.2
ESPON_2006_18_public_sector_nuts_2_3_data	Percentage of regional Pre-Accession-Aid (PHARE, PHARE CBC, ISPA) addressing regional market potential	NUTS2&NUT S3 (version 1999)	1998/2000	ESPON Project 2.2.2
ESPON_2006_18_public_sector_nuts_2_3_data	Percentage of regional Pre-Accession-Aid (PHARE, PHARE CBC, ISPA) addressing urbanisation&localisation advantages	NUTS2&NUT S3 (version 1999)	1998/2000	ESPON Project 2.2.2
ESPON_2006_18_public_sector_nuts_2_3_data	Total Pre-Accession-Aid spending (PHARE, PHARE CBC, ISPA)	NUTS2&NUT S3 (version 1999)	1998/2000	ESPON Project 2.2.2
ESPON_2006_18_public_sector_nuts_2_3_data	Average Annual Pre-Accession-Aid spending (PHARE, PHARE CBC, ISPA)	NUTS2&NUT S3 (version 1999)	1998/2000	ESPON Project 2.2.2
ESPON_2006_18_public_sector_nuts_2_3_data	All Structural and Cohesion Fund expenditure	NUTS2&NUT S3 (version 1999)	1994/1999	ESPON Project 2.2.1
ESPON_2006_18_public_sector_nuts_2_3_data	Structural Fund expenditure (R)	NUTS2&NUT S3 (version 1999)	1994/1999	ESPON Project 2.2.1
ESPON_2006_18_public_sector_nuts_2_3_data	Structural Fund expenditure (S)	NUTS2&NUT S3 (version 1999)	1994/1999	ESPON Project 2.2.1
ESPON_2006_18_public_sector_nuts_2_3_data	Structural Fund expenditure (A)	NUTS2&NUT S3 (version 1999)	1994/1999	ESPON Project 2.2.1
ESPON_2006_18_public_sector_nuts_2_3_data	Cohesion Fund expenditure (T)	NUTS2&NUT S3 (version 1999)	1994/1999	ESPON Project 2.2.1
ESPON_2006_18_public_sector_nuts_2_3_data	Cohesion Fund expenditure (E)	NUTS2&NUT S3 (version 1999)	1994/1999	ESPON Project 2.2.1

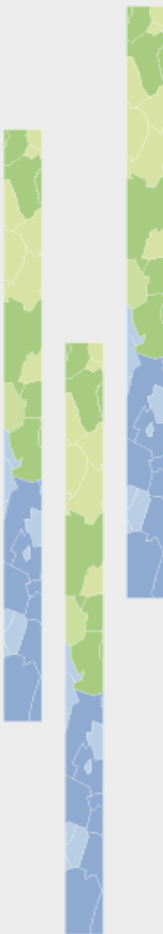
**APPENDIX 2 - Updated indicators from the ESPON 2006 Database – Data from Eurostat (ESPN Basic Indicators)**

<b>DATASET</b>	<b>NAME OF INDICATOR</b>	<b>Geog. Object</b>	<b>Temporal start/end</b>	<b>SOURCE (main)</b>
00_area_nuts_2-3_data	Total land area	NUTS2, NUTS3 (version 2003)	2003/2003	Eurostat
02_population_nuts_2	Population total	NUTS2 (version 2003)	2003/2003	Eurostat
02_population_nuts_2	Population density	NUTS2 (version 2003)	2002/2002	Eurostat
02_population_nuts_2	Share of female population	NUTS2 (version 2003)	2003/2003	Eurostat
02_population_nuts_2	Share of male population	NUTS2 (version 2003)	2003/2003	Eurostat
02_population_nuts_2	Share of population < 14 years	NUTS2 (version 2003)	2003/2003	Eurostat
02_population_nuts_2	Share of population > 65 years	NUTS2 (version 2003)	2003/2003	Eurostat
02_population_nuts_2	Share of high aged population (> 75 years)	NUTS2 (version 2003)	2003/2003	Eurostat
02_population_nuts_2	Share of female population < 14 years	NUTS2 (version 2003)	2003/2003	Eurostat
02_population_nuts_2	Share of female population > 65 years	NUTS2 (version 2003)	2003/2003	Eurostat
02_population_nuts_2	Share of female high aged population (> 75 years)	NUTS2 (version 2003)	2003/2003	Eurostat
02_population_nuts_2	Share of male population < 14 years	NUTS2 (version 2003)	2003/2003	Eurostat
02_population_nuts_2	Share of male population > 65 years	NUTS2 (version 2003)	2003/2003	Eurostat
02_population_nuts_2	Share of male high aged population (> 75 years)	NUTS2 (version 2003)	2003/2003	Eurostat
02_population_nuts_2	Development of total population in %	NUTS2 (version 2003)	1995/2003	Eurostat
02_population_nuts_2	Development of female population in %	NUTS2 (version 2003)	1995/2003	Eurostat
02_population_nuts_2	Development of male population in %	NUTS2 (version 2003)	1995/2003	Eurostat
02_population_nuts_3	Average Population	NUTS3 (version 2003)	2003/2003	Eurostat
02_population_nuts_3	Average male Population, share in %	NUTS3 (version 2003)	2003/2003	Eurostat
02_population_nuts_3	Average female Population, share in %	NUTS3 (version 2003)	2003/2003	Eurostat
02_population_nuts_3	Population density	NUTS3 (version 2003)	2002/2002	Eurostat
02_population_nuts_3	Development average population	NUTS3 (version 2003)	1995/2003	Eurostat
03_employment_and_labourmarket_nuts_2	Active population total 2001	NUTS2 (version 1999)	2001/2001	Eurostat
03_employment_and_labourmarket_nuts_2	Share of active population < 25 years	NUTS2 (version 1999)	2001/2001	Eurostat



		1999)		
03_employment_and_labourmarket_nuts_2	Persons employed	NUTS2 (version 1999)	2001/2001; 2002/2002	Eurostat
03_employment_and_labourmarket_nuts_2	Share of persons employed male	NUTS2 (version 1999)	2001/2001	Eurostat
03_employment_and_labourmarket_nuts_2	Share of persons employed female	NUTS2 (version 1999)	2001/2001	Eurostat
03_employment_and_labourmarket_nuts_2	Share of persons employed in Agriculture in % of total	NUTS2 (version 1999)	2001/2001	Eurostat
03_employment_and_labourmarket_nuts_2	Share of persons employed in Industry in % of total	NUTS2 (version 1999)	2001/2001	Eurostat
03_employment_and_labourmarket_nuts_2	Share of persons employed in Services in % of total	NUTS2 (version 1999)	2001/2001	Eurostat
03_employment_and_labourmarket_nuts_2	Share of employed persons, national, < 25 years, in % of total	NUTS2 (version 1999)	2002/2002	Eurostat
03_employment_and_labourmarket_nuts_2	Share of employed persons, national, < 25 years, in % of total, >65 years	NUTS2 (version 1999)	2003/2003	Eurostat
03_unemployment_nuts_2	Unemployment rate total	NUTS2 (version 2003)	2004/2004	Eurostat
03_unemployment_nuts_2	Unemployment rate female	NUTS2 (version 2003)	2004/2004	Eurostat
03_unemployment_nuts_2	Unemployment rate male	NUTS2 (version 2003)	2004/2004	Eurostat
03_unemployment_nuts_2	Unemployment rate of persons < 25 years	NUTS2 (version 2003)	2004/2004	Eurostat
03_unemployment_nuts_2	Development of unemployment rate	NUTS2 (version 2003)	1999/2004	Eurostat
03_unemployment_nuts_2	Development of unemployment rate, female	NUTS2 (version 2003)	1999/2004	Eurostat
03_unemployment_nuts_2	Development of unemployment rate, male	NUTS2 (version 2003)	1999/2004	Eurostat
03_unemployment_nuts_3	Unemployment rate total	NUTS3 (version 1999)	2001/2001	Eurostat
03_unemployment_nuts_3	Unemployment rate female	NUTS3 (version 1999)	2001/2001	Eurostat
03_unemployment_nuts_3	Unemployment rate male	NUTS3 (version 1999)	2001/2001	Eurostat
03_unemployment_nuts_3	Unemployment rate of persons < 25 years	NUTS3 (version 1999)	2001/2001	Eurostat
03_unemployment_nuts_3	Development of unemployment rate	NUTS3 (version 1999)	1998/2001	Eurostat
03_unemployment_nuts_3	Development of unemployment rate, female	NUTS3 (version 1999)	1998/2001	Eurostat
03_unemployment_nuts_3	Development of unemployment rate, male	NUTS3 (version 1999)	1998/2001	Eurostat
03_unemployment_nuts_3	Development of unemployment rate, young pop	NUTS3 (version 1999)	1998/2001	Eurostat
wealth_and_production_nuts_2-3_data	GDP in Purchasing Power Parities per inhabitant 2002	NUTS2, NUTS3 (version 2003)	2002/2002	ESPO project 3.2
wealth_and_production_nuts_2-3_data	GDP in Euro per inhabitant 2002	NUTS2, NUTS3 (version 2003)	2002/2002	ESPO project 3.2
wealth_and_production_nuts_2-3_data	Development of GDP in	NUTS2,	1998/2002	ESPO project

	Purchasing Power Parities per inhabitant 1998-2002	NUTS3 (version 2003)		3.2
wealth_and_production_nuts_2-3_data	Development of GDP in Euro per inhabitant 1998-2002	NUTS2, NUTS3 (version 2003)	1998/2002	ESPON project 3.2
07_research_and_development_nuts_2	Patent applications to the EPO per persons employed	NUTS2 (version 2003)	2002/2002	Eurostat
07_research_and_development_nuts_2	Total intramural R&D expenditure	NUTS2 (version 2003)	2002/2002	Eurostat
07_research_and_development_nuts_2	FuE Business Enterprise Sector, personnell	NUTS2 (version 2003)	2003/2003	Eurostat



# LISTING OF ESPON DATABASE INDICATORS

*SITUATION IN MARCH 2011*

## CONTENT

- This report proposes to list the indicators produced by ESPON Projects – under priority 1-2-3 – which are downloadable through the ESPON Database interface.
- The listing is organized by categories, following the ESPON Database structure: regional data, urban data, local data, neighborhood data, grid data, world data and historical data.

**ESPON 2013 DATABASE**

**MARCH 2011**



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

**74 pages**

# LIST OF AUTHORS

Ronan Ysebaert, UMS 2414 RIATE

Bogdan Moisuc, LIG Steamer

## Contact

ronan.ysebaert@ums-riate.fr

moisuc@imag.fr

tel. (+33) 1 57 27 65 32

tel. (+33) 4 76 82 72 11

# TABLE OF CONTENTS

**Forewords ..... 3**

**Regional data ..... 4**

**Local data..... 28**

**Neighborhood data..... 53**

**Urban data ..... 56**

**World data ..... 67**

**Grid data ..... 74**

# Forewords

This Technical report lists all the indicators available in the ESPON Database which have been delivered by ESPON Projects (version March 2011). The most part of this data is free of download.

However, this technical report does not take into accounts the indicators of the ESPON 2006 Database (Access format), the Europe In the World database (ESPON 3.4.1)

In light orange are shown the indicators produced by the ESPON Database Project.

Indicators are listed as below:

- **Provider:** Name of the project who has delivered the data.
- **Indicator code:** Code of the indicator in the database. In most of the cases, the ESPON coding scheme has still to be applied to the indicators.
- **Indicator name:** Name given to the indicator (in the field indicator name of the metadata)
- **Year(s):** temporal validity of the indicator (Year“a” means a single year; year“x”-year“y” means an evolution or an interval)
- **Theme-Sub theme:** Thematic classification of the indicator, following the thematic structure developed by ESPON Database Project.
- **Geo object:** Territorial or spatial unit of reference (NUTS, LAU2, FUA etc.)
- **Geo version:** Version of the geographical object (NUTS 2006, LAU2 2008 etc.)
- **Space:** Geographical coverage of the indicator. When the all ESPON Area (EU27+4) is covered “ESPON Area” is mentioned (but it does not necessarily mean that the dataset is complete).
- **Category:** Precise where it is possible to download the data in the ESPON DB Web interface.
- **Dataset title:** For regional data only, name of the dataset where is include the indicator.

## • Regional data (1)

Provider	Indicator - code	Indicator - name	Year(s)	Theme - Sub theme	Geo object	Geo version	Dataset title	Category
Accessibility (RAIL_Relativechang	Potential accessibility rail, relative change	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 1999, NUTS 2003	Accessibility	Regional data - 1	
Accessibility (RAIL_Absolutechan	Potential accessibility rail, absolute change	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 1999, NUTS 2003	Accessibility	Regional data - 2	
Accessibility (RAIL_Indexchange	Potential accessibility rail, index change	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 1999, NUTS 2003	Accessibility	Regional data - 3	
Accessibility (air_i_ch	Potential accessibility by air, change of standardised	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 2006	Accessibility	Regional data - 4	
Accessibility (mm_a	Multimodal potential accessibility, absolute change	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 2006	Accessibility	Regional data - 5	
Accessibility (air_a	Potential accessibility by air, absolute change	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 2006	Accessibility	Regional data - 6	
Accessibility (mm_r	Multimodal potential accessibility, relative change	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 2006	Accessibility	Regional data - 7	
Accessibility (air_r	Potential accessibility by air, relative change	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 2006	Accessibility	Regional data - 8	
Accessibility (ROAD_Index	Potential accessibility road, standardised	2006	0301 Transport - Accessibility	NUTS	NUTS 1999, NUTS 2003	Accessibility	Regional data - 9	
Accessibility (MM	Multimodal potential accessibility, absolute level	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 2006	Accessibility	Regional data - 10	
Accessibility (ROAD_Relativechar	Potential accessibility road, relative change	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 1999, NUTS 2003	Accessibility	Regional data - 11	
Accessibility (ROAD_Absolutecha	Potential accessibility road, absolute change	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 1999, NUTS 2003	Accessibility	Regional data - 12	
Accessibility (pop_t	Total population	Between 2001 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	Accessibility	Regional data - 13	
Accessibility (Air	Potential accessibility by air, absolute level	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 2006	Accessibility	Regional data - 14	
Accessibility (mm_i	Multimodal potential accessibility, standardised	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 2006	Accessibility	Regional data - 15	
Accessibility (mm_i_ch	Multimodal potential accessibility, change of standardised	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 2006	Accessibility	Regional data - 16	
Accessibility (air_i	Potential accessibility by air, standardised	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 2006	Accessibility	Regional data - 17	
Accessibility (ROAD_Indexchange	Potential accessibility road, index change	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 1999, NUTS 2003	Accessibility	Regional data - 18	
Accessibility (RAIL_Index	Potential accessibility rail, standardized	2006	0301 Transport - Accessibility	NUTS	NUTS 1999, NUTS 2003	Accessibility	Regional data - 19	
Accessibility (Rail	Potential accessibility, rail, absolute level	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 2006	accessibility_road_rail	Regional data - 20	
Accessibility (Pop	Total population	Between 2001 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	accessibility_road_rail	Regional data - 21	
Accessibility (Rail_i_EU27	Potential accessibility, rail, standardised	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 2006	accessibility_road_rail	Regional data - 22	
Accessibility (Rail_a_ch	Potential accessibility, rail, absolute change	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 2006	accessibility_road_rail	Regional data - 23	
Accessibility (Rail_r_ch	Potential accessibility, rail, relative change	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 2006	accessibility_road_rail	Regional data - 24	
Accessibility (Road	Potential accessibility, road, absolute level	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 2006	accessibility_road_rail	Regional data - 25	
Accessibility (Road_i_ESPON	Potential accessibility, road, standardised ESPON	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 2006	accessibility_road_rail	Regional data - 26	
Accessibility (Road_i_ch	Potential accessibility, road, change of standardised	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 2006	accessibility_road_rail	Regional data - 27	
Accessibility (Road_r_ch	Potential accessibility, road, relative change	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 2006	accessibility_road_rail	Regional data - 28	
Accessibility (Road_a_ch	Potential accessibility, road, absolute change	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 2006	accessibility_road_rail	Regional data - 29	
Accessibility (Road_i_EU27	Potential accessibility, road, standardised	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 2006	accessibility_road_rail	Regional data - 30	
Accessibility (Rail_i_ESPON	Potential accessibility, rail, standardised ESPON	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 2006	accessibility_road_rail	Regional data - 31	
Accessibility (Rail_i_ch	Potential accessibility, rail, change of standardised	Between 2001 and 2006	0301 Transport - Accessibility	NUTS	NUTS 2006	accessibility_road_rail	Regional data - 32	
DEMIFER - Der GEI_0105_%_ch_pe	GDP Growth - € per inhabitant	Between 2001 and 2005	0701 Economy - Aggregated account	NUTS	NUTS 2006	DEMIFER Labour Force Data	Regional data - 33	
DEMIFER - Der GPla_EU100	GDP in PPP per inhabitant in 2005	2005	0701 Economy - Aggregated account	NUTS	NUTS 2006	DEMIFER Labour Force Data	Regional data - 34	
DEMIFER - Der LF_part_1564	Labour Force Participation	2007	0602 Social Affairs - Labour market	NUTS	NUTS 2006	DEMIFER Labour Force Data	Regional data - 35	
DEMIFER - Der GEIa_EU100	GDP per inhabitant	2005	0701 Economy - Aggregated account	NUTS	NUTS 2006	DEMIFER Labour Force Data	Regional data - 36	
DEMIFER - Der GDP_PPS	Total GDP (millions PPS)	2005	0701 Economy - Aggregated account	NUTS	NUTS 2006	DEMIFER Labour Force Data	Regional data - 37	
DEMIFER - Der GDP_euro	Total GDP (Millions euros)	2005	0701 Economy - Aggregated account	NUTS	NUTS 2006	DEMIFER Labour Force Data	Regional data - 38	
DEMIFER - Der pop_t	Total population	2007	0201 Demography - Population struct	NUTS	NUTS 2006	DEMIFER Labour Force Data	Regional data - 39	
DEMIFER - Der Tertiary education(1	Tertiary Educated	2007	0601 Social Affairs - Education	NUTS	NUTS 2006	DEMIFER Labour Force Data	Regional data - 40	
DEMIFER - Der mig_Non-EU	Foreign Population from non-EU Countries	2007	0201 Demography - Population struct	NUTS	NUTS 2006	DEMIFER Labour Force Data	Regional data - 41	
DEMIFER - Der mig_EU27	Foreign Population from EU27 Countries	2007	0201 Demography - Population struct	NUTS	NUTS 2006	DEMIFER Labour Force Data	Regional data - 42	
DEMIFER - Der MIG_tot	Foreign Population	2007	0204 Demography - Migrations	NUTS	NUTS 2006	DEMIFER Labour Force Data	Regional data - 43	
DEMIFER - Der UNEMP_1564	Unemployment Rate	2007	0702 Economy - Employment	NUTS	NUTS 2006	DEMIFER Labour Force Data	Regional data - 44	
DEMIFER - Der unemp_long	Long-Term Unemployment	2007	0702 Economy - Employment	NUTS	NUTS 2006	DEMIFER Labour Force Data	Regional data - 45	
DEMIFER - Der GPI_0105_%_ch_pe	GDP Growth - PPP per inhabitant	Between 2001 and 2005	0701 Economy - Aggregated account	NUTS	NUTS 2006	DEMIFER Labour Force Data	Regional data - 46	
DEMIFER - Der real_DR_1574	*Real* Dependency Ratio	2007	0602 Social Affairs - Labour market	NUTS	NUTS 2006	DEMIFER Labour Force Data	Regional data - 47	
DEMIFER - Der 0-14_change_STQ	Change in child ages, STQ projection	Between 2005 and 2050	0201 Demography - Population struct	NUTS	NUTS 2006	DEMIFER policy scenarios	Regional data - 48	
DEMIFER - Der pop_change_CME	Change in population in policy scenario CME	Between 2005 and 2050	0201 Demography - Population struct	NUTS	NUTS 2006	DEMIFER policy scenarios	Regional data - 49	
DEMIFER - Der pop_change_LSE	Change in population in policy scenario LSE,	Between 2005 and 2050	0201 Demography - Population struct	NUTS	NUTS 2006	DEMIFER policy scenarios	Regional data - 50	
DEMIFER - Der pop_change_EME	Change in population in policy scenario EME	Between 2005 and 2050	0201 Demography - Population struct	NUTS	NUTS 2006	DEMIFER policy scenarios	Regional data - 51	
DEMIFER - Der pop_change_GSE	Change in population in policy scenario GSE	Between 2005 and 2050	0201 Demography - Population struct	NUTS	NUTS 2006	DEMIFER policy scenarios	Regional data - 52	
DEMIFER - Der 15-64 change EME	Change in working ages, EME scenario	Between 2005 and 2050	0201 Demography - Population struct	NUTS	NUTS 2006	DEMIFER policy scenarios	Regional data - 53	





## • Regional data (3)

Provider	Indicator - code	Indicator - name	Year(s)	Theme - Sub theme	Geo object	Geo version	Dataset title	Category
DEMIFER - Der SCE,GSE_LF,m,ch		Change in male labour force in policy scenario GSE	Between 2005 and 2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 108
DEMIFER - Der SCE,CME_LF,ch		Change in labour force in policy scenario CME	Between 2005 and 2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 109
DEMIFER - Der SCE,LSE_LF,ch		Change in labour force in policy scenario LSE'	Between 2005 and 2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 110
DEMIFER - Der SCE,EME_LF,ch		Change in labour force in policy scenario EME	Between 2005 and 2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 111
DEMIFER - Der SCE,GSE_LF,ch		Change in labour force in policy scenario GSE	Between 2005 and 2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 112
DEMIFER - Der LF,m,55-59_rt		Male Aged 55-59 Labour Force Participation	2005	0602 Social Affairs - Labour market	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 113
DEMIFER - Der SCE,CME_LF,m,55-5		Male Aged 55-59 Labour Force Participation in policy scenar	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 114
DEMIFER - Der SCE,LSE_LF,m,55-5		Male Aged 55-59 Labour Force Participation in policy scenar	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 115
DEMIFER - Der SCE,EME_LF,m,55-5		Male Aged 55-59 Labour Force Participation in policy scenar	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 116
DEMIFER - Der SCE,GSE_LF,m,55-5		Male Aged 55-59 Labour Force Participation in policy scenar	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 117
DEMIFER - Der LF,f,40-44_rt		Female Aged 40-44 Labour Force Participation	2005	0602 Social Affairs - Labour market	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 118
DEMIFER - Der SCE,CME_LF,f,40-4		Female Aged 40-44 Labour Force Participation in policy scer	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 119
DEMIFER - Der SCE,LSE_LF,f,40-44		Female Aged 40-44 Labour Force Participation in policy scer	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 120
DEMIFER - Der SCE,EME_LF,f,40-4		Female Aged 40-44 Labour Force Participation in policy scer	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 121
DEMIFER - Der SCE,GSE_LF,f,40-4		Female Aged 40-44 Labour Force Participation in policy scer	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 122
DEMIFER - Der LF,m,20-24_rt		Male Aged 20-24 Labour Force Participation	2005	0602 Social Affairs - Labour market	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 123
DEMIFER - Der SCE,CME_LF,m,20-2		Male Aged 20-24 Labour Force Participation in policy scenar	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 124
DEMIFER - Der SCE,LSE_LF,m,20-2		Male Aged 20-24 Labour Force Participation in policy scenar	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 125
DEMIFER - Der SCE,EME_LF,m,20-2		Male Aged 20-24 Labour Force Participation in policy scenar	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 126
DEMIFER - Der SCE,GSE_LF,m,20-2		Male Aged 20-24 Labour Force Participation in policy scenar	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 127
DEMIFER - Der IMMIG_abs		Immigration from outside Europe	2050	0204 Demography - Migrations	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 128
DEMIFER - Der SCE,CME_IMMIG_ab		Immigration from outside Europe in policy scenario CME	2050	0204 Demography - Migrations	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 129
DEMIFER - Der SCE,LSE_IMMIG_ab		Immigration from outside Europe in policy scenario LSE'	2050	0204 Demography - Migrations	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 130
DEMIFER - Der SCE,EME_IMMIG_ab		Immigration from outside Europe in policy scenario EME	2050	0204 Demography - Migrations	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 131
DEMIFER - Der SCE,GSE_IMMIG_ab		Immigration from outside Europe in policy scenario GSE	2050	0204 Demography - Migrations	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 132
DEMIFER - Der EMIG_f,30-34_rt		Emigration Rate - Females Aged 30-34	2005	0204 Demography - Migrations	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 133
DEMIFER - Der SCE,CME_TFR_ix		Total Fertility Rate in policy scenario CME	2050	0202 Demography - Natural changes	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 134
DEMIFER - Der SCE,LSE_TFR_ix		Total Fertility Rate in policy scenario LSE'	2050	0202 Demography - Natural changes	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 135
DEMIFER - Der SCE,EME_TFR_ix		Total Fertility Rate in policy scenario EME	2050	0202 Demography - Natural changes	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 136
DEMIFER - Der SCE,GSE_TFR_ix		Total Fertility Rate in policy scenario GSE	2050	0202 Demography - Natural changes	NUTS	NUTS 2006	policy_scenario_NEAA	Regional data - 137
DEMIFER - Der OD_rt		Old Age Dependency Ratio	2005	0602 Social Affairs - Labour market	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 138
DEMIFER - Der SC,STQ_OD_rt		Old Age Dependency Ratio - STQ scenario	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 139
DEMIFER - Der SC,NMI_OD_rt		Old Age Dependency Ratio - NMI scenario	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 140
DEMIFER - Der SC,NEM_OD_rt		Old Age Dependency Ratio - NEM scenario	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 141
DEMIFER - Der EOD_rt		Economic Old Age Dependency Ratio	2005	0602 Social Affairs - Labour market	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 142
DEMIFER - Der SC,STQ_EOD_rt		Economic Old Age Dependency Ratio - STQ scenario	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 143
DEMIFER - Der SC,NMI_EOD_rt		Economic Old Age Dependency Ratio - NMI scenario	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 144
DEMIFER - Der SC,NEM_EOD_rt		Economic Old Age Dependency Ratio - NEM scenario	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 145
DEMIFER - Der LMD_rt		Labour Market Dependency Ratio	2005	0602 Social Affairs - Labour market	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 146
DEMIFER - Der SC,STQ_LMD_rt		Labour Market Dependency Ratio - STQ scenario	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 147
DEMIFER - Der SC,NMI_LMD_rt		Labour Market Dependency Ratio - NMI scenario	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 148
DEMIFER - Der SC,NEM_LMD_rt		Labour Market Dependency Ratio - NEM scenario	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 149
DEMIFER - Der VOD_rt		Very Old Age Dependency Ratio	2005	0602 Social Affairs - Labour market	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 150
DEMIFER - Der SC,STQ_VOD_rt		Very Old Age Dependency Ratio - STQ scenario	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 151
DEMIFER - Der SC,NMI_VOD_rt		Very Old Age Dependency Ratio - NMI scenario	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 152
DEMIFER - Der SC,NEM_VOD_rt		Very Old Age Dependency Ratio - NEM scenario	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 153
DEMIFER - Der SC,IMP_mig,pop_rt		Impact of Migration on Population	2050	0204 Demography - Migrations	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 154
DEMIFER - Der SC,IMP_mig,vod_rt		Impact of migration on Very Old Age Dependency Ratio	2050	0204 Demography - Migrations	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 155
DEMIFER - Der SC,IMP_mig,lf_rt		Impact of migration on Labour Force	2050	0204 Demography - Migrations	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 156
DEMIFER - Der POP_abs		Total population	2005	0201 Demography - Population struct	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 157
DEMIFER - Der SC,STQ_pop_abs		Total population - STQ scenario	2050	0201 Demography - Population struct	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 158
DEMIFER - Der SC,NMI_pop_abs		Total population - NMI scenario	2050	0201 Demography - Population struct	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 159
DEMIFER - Der SC,NEM_pop_abs		Total population - NEM scenario	2050	0201 Demography - Population struct	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 160
DEMIFER - Der SC,STQ_pop_ch		Change in population - STQ scenario	Between 2005 and 2050	0201 Demography - Population struct	NUTS	NUTS 2006	scenarios STQ NMI NEM	Regional data - 161

- Regional data (4)

Provider	Indicator - code	Indicator - name	Year(s)	Theme - Sub theme	Geo object	Geo version	Dataset title	Category
DEMIFER - Der SC,NMI_pop_ch		Change in population - NMI scenario	Between 2005 and 2050	0201 Demography - Population struct	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 162
DEMIFER - Der SC,NEM_pop_ch		Change in population - NEM scenario	Between 2005 and 2050	0201 Demography - Population struct	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 163
DEMIFER - Der LF_abs		Total labour force	2005	0602 Social Affairs - Labour market	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 164
DEMIFER - Der SC,STQ_if_abs		Total labour force - STQ scenario	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 165
DEMIFER - Der SC,NMI_if_abs		Total labour force - NMI scenario	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 166
DEMIFER - Der SC,NEM_if_abs		Total labour force - NEM scenario	2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 167
DEMIFER - Der SC,STQ_if_ch		Change in labour force - STQ scenario	Between 2005 and 2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 168
DEMIFER - Der SC,NMI_if_ch		Change in labour force - NMI scenario	Between 2005 and 2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 169
DEMIFER - Der SC,NEM_if_ch		Change in labour force - NEM scenario	Between 2005 and 2050	0602 Social Affairs - Labour market	NUTS	NUTS 2006	scenarios_STQ_NMI_NEM	Regional data - 170
DEMIFER - Der CBR		Crude Birth Rate	Between 2001 and 2005	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 171
DEMIFER - Der CDR		Crude Death Rate	Between 2001 and 2005	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 172
DEMIFER - Der TFR		Total Fertility Rate	2005	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 173
DEMIFER - Der e0_ttl		Life Expectancy at Birth	Between 2002 and 2004	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 174
DEMIFER - Der e0_m		Life Expectancy at Birth - Men	Between 2002 and 2004	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 175
DEMIFER - Der e0_f		Life Expectancy at Birth - Women	Between 2002 and 2004	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 176
DEMIFER - Der icluster		Cluster Analysis of Demographic Indicators	Between 2000 and 2006	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	TFR_LE	Regional data - 177
DEMIFER - Der LE_M_TRE		Life expectancies at birth for males	Between 2005 and 2050	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 178
DEMIFER - Der LE_F_TRE		Life expectancies at birth for females	Between 2005 and 2050	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 179
DEMIFER - Der LE_M_GSE		Life expectancies at birth for males, GSE scenario	Between 2045 and 2050	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 180
DEMIFER - Der LE_F_GSE		Life expectancies at birth for females, GSE scenario	Between 2045 and 2050	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 181
DEMIFER - Der LE_M_EME		Life expectancies at birth for males, EME scenario	Between 2045 and 2050	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 182
DEMIFER - Der LE_F_EME		Life expectancies at birth for females, EME scenario	Between 2045 and 2050	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 183
DEMIFER - Der LE_M_LSE		Life expectancies at birth for males, LSE scenario	Between 2045 and 2050	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 184
DEMIFER - Der LE_F_LSE		Life expectancies at birth for females, LSE scenario	Between 2045 and 2050	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 185
DEMIFER - Der LE_M_CME		Life expectancies at birth for males, CME scenario	Between 2045 and 2050	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 186
DEMIFER - Der LE_F_CME		Life expectancies at birth for females, CME scenario	Between 2045 and 2050	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 187
DEMIFER - Der SMR_M		Mortality ratios for males	Between 1992 and 2005	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 188
DEMIFER - Der SMR_F		Mortality ratios for females	Between 1992 and 2005	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 189
DEMIFER - Der SMR_GSE_M		Mortality ratios for males, GSE Scenario	Between 2045 and 2050	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 190
DEMIFER - Der SMR_GSE_F		Mortality ratios for females, GSE Scenario	Between 2045 and 2050	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 191
DEMIFER - Der SMR_EME_M		Mortality ratios for males, EME Scenario	Between 2045 and 2050	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 192
DEMIFER - Der SMR_EME_F		Mortality ratios for females, EME Scenario	Between 2045 and 2050	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 193
DEMIFER - Der SMR_LSE_M		Mortality ratios for males, LSE Scenario	Between 2045 and 2050	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 194
DEMIFER - Der SMR_LSE_F		Mortality ratios for females, LSE Scenario	Between 2045 and 2050	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 195
DEMIFER - Der SMR_CME_M		Mortality ratios for males, CME Scenario	Between 2045 and 2050	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 196
DEMIFER - Der SMR_CME_F		Mortality ratios for females, CME Scenario	Between 2045 and 2050	0202 Demography - Natural changes	NUTS	NUTS 2006	TFR_LE	Regional data - 197
Demography # pop_t		Total population	Between 1995 and 2005	0201 Demography - Population struct	NUTS	NUTS 2006	Demographic trends and migratic	Regional data - 198
Demography # pop_ch		Total population change	Between 1995 and 2005	0202 Demography - Natural changes	NUTS	NUTS 2006	Demographic trends and migratic	Regional data - 199
Demography # nat_ch		Natural population change	Between 1996 and 2005	0202 Demography - Natural changes	NUTS	NUTS 2006	Demographic trends and migratic	Regional data - 200
Demography # mig_ch		Migratory population change	Between 1996 and 2005	0204 Demography - Migrations	NUTS	NUTS 2006	Demographic trends and migratic	Regional data - 201
Demography # CI1_TOT		Core Indicator 1: Annual growth rate (population)	Between 1995 and 2005	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Demographic trends and migratic	Regional data - 202
Demography # CI2_MIG		Core Indicator 2: Annual net migration development	Between 1996 and 2005	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Demographic trends and migratic	Regional data - 203
Demography # CI3_NAT		Core Indicator 3: Annual natural population development	Between 1996 and 2005	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Demographic trends and migratic	Regional data - 204
Demography # CI4_TYPO		Core Indicator 4: Population development by components	Between 1996 and 2005	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Demographic trends and migratic	Regional data - 205
EDORA - Euro HOT_emp_rtp		Employed persons in hotel and restaurants	2006	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 206
EDORA - Euro URB-RUR_oeecd_orc		OECD rural /urban classification	2006	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 207
EDORA - Euro POP_rtc		Total population	Between 2001 and 2007	0201 Demography - Population struct	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 208
EDORA - Euro POP_0-14_rtp		Share of population aged 0-14	Between 2001 and 2007	0201 Demography - Population struct	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 209
EDORA - Euro POP_15-64_rtp		Share of population aged 15-64	Between 2001 and 2007	0201 Demography - Population struct	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 210
EDORA - Euro POP_65+_rtp		Share of population aged 65 years and more	Between 2001 and 2007	0201 Demography - Population struct	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 211
EDORA - Euro DEP_rtd		Dependency rate	Between 2001 and 2007	0201 Demography - Population struct	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 212
EDORA - Euro POP_ch_rtd		Total population change	Between 2001 and 2007	0202 Demography - Natural changes	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 213
EDORA - Euro NAT_rte		Natural population change	Between 2001 and 2006	0202 Demography - Natural changes	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 214
EDORA - Euro NAT_ch_rtd		Natural population change, evolution	Between 2001 and 2006	0202 Demography - Natural changes	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 215

- Regional data (5)

Provider	Indicator - code	Indicator - name	Year(s)	Theme - Sub theme	Geo object	Geo version	Dataset title	Category
EDORA - Euro	MIG_rte	Migration population change	Between 2001 and 2006	0204 Demography - Migrations	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 216
EDORA - Euro	MIG_ch_rtd	Migratory population change, evolution	Between 2001 and 2006	0204 Demography - Migrations	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 217
EDORA - Euro	EDUC_prim_rtc	Primary education level	2007	0601 Social Affairs - Education	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 218
EDORA - Euro	EDUC_sec_rtc	Secondary education level	2007	0601 Social Affairs - Education	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 219
EDORA - Euro	EDUC_tert_rtc	Tertiary education level	2007	0601 Social Affairs - Education	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 220
EDORA - Euro	EDUC_long_learn_rur	Life-Long Learning in Rural Areas	2000	0601 Social Affairs - Education	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 221
EDORA - Euro	EMP_rtp	Employment rate	2007	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 222
EDORA - Euro	EMP_15-64_m_rtp	Employment rate, males, 15-64 years	2007	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 223
EDORA - Euro	EMP_15-64_f_rtp	Employment rate, females, 15-64 years	2007	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 224
EDORA - Euro	EMP_15-24_rtp	Employment rate, 15-24 years	2007	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 225
EDORA - Euro	EMP_45+_rtp	Employment rate, 45+ years	2007	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 226
EDORA - Euro	EMP_45-54_rtp	Employment rate, 45-54 years	2007	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 227
EDORA - Euro	EMP_55-64_rtp	Employment rate, 55-64 years	2007	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 228
EDORA - Euro	EMP_prim_rtc	Employed persons in primary sector	2005	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 229
EDORA - Euro	EMP_sec_rtc	Employed persons in secondary sector	2005	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 230
EDORA - Euro	EMP_tert_rtc	Employed persons in tertiary sector	2005	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 231
EDORA - Euro	UMP_rtp	Unemployment rate	Between 2006 and 2007	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 232
EDORA - Euro	UMP_rtc	Unemployed persons	Between 2006 and 2007	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 233
EDORA - Euro	UMP_ch_rtd	Unemployment evolution	Between 2002 and 2007	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 234
EDORA - Euro	UMP_ch_15+_rtd	Evolution of unemployed persons aged 15+	Between 2000 and 2005	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 235
EDORA - Euro	UMP_ch_15-24_rtd	Evolution of unemployed persons aged 15-24	Between 2000 and 2005	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 236
EDORA - Euro	UMP_ch_25+_rtd	Evolution of unemployed persons aged 25+	Between 2000 and 2005	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 237
EDORA - Euro	UMP_ch_15+_m_rtd	Evolution of males unemployed aged 15+	Between 2000 and 2005	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 238
EDORA - Euro	UMP_15+_rtp	Unemployment rate of population aged 15+	Between 2006 and 2007	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 239
EDORA - Euro	UMP_15+_m_rtp	Unemployment rate of males aged 15+	Between 2006 and 2007	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 240
EDORA - Euro	UMP_15+_f_rtp	Unemployment rate of females aged 15+	Between 2006 and 2007	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 241
EDORA - Euro	UMP_15-24_rtp	Unemployment rate of population aged 15-24	Between 2006 and 2007	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 242
EDORA - Euro	UMP_25+_rtp	Unemployment rate of population aged 25+	Between 2006 and 2007	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 243
EDORA - Euro	UMP_lng_rtp	Long term unemployment rate	Between 2002 and 2007	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 244
EDORA - Euro	UMP_ch_lng_rtd	Evolution of long term unemployment rate	Between 2002 and 2007	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 245
EDORA - Euro	MIN_fir_rtp	Number of firms in mining and quarrying	2006	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 246
EDORA - Euro	MAN_fir_rtp	Number of firms in manufacturing	2006	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 247
EDORA - Euro	GAS_fir_rtp	Number of firms in electricity, gas and water supply	2006	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 248
EDORA - Euro	CON_fir_rtp	Number of firms in construction	2006	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 249
EDORA - Euro	RET_fir_rtp	Number of firms in wholesale and retail trade	2006	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 250
EDORA - Euro	HOT_fir_rtp	Number of firms in hotel and restaurants	2006	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 251
EDORA - Euro	TRA_fir_rtp	Number of firms in transport, storage and communication	2006	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 252
EDORA - Euro	RES_fir_rtp	Number of firms in real state, renting and business activities	2006	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 253
EDORA - Euro	MIN_emp_rtp	Employed persons in mining and quarrying	2006	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 254
EDORA - Euro	MAN_emp_rtp	Employed persons in manufacturing	2006	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 255
EDORA - Euro	GAS_emp_rtp	Employed persons in electricity, gas and water supply	2006	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 256
EDORA - Euro	CON_emp_rtp	Employed persons in construction	2006	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 257
EDORA - Euro	RET_emp_rtp	Employed persons in wholesale and retail trade	2006	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 258
EDORA - Euro	TRA_emp_rtp	Employed persons in transport, storage and communication	2006	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 259
EDORA - Euro	RES_emp_rtp	Employed persons in real state, renting and business activities	2006	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 260
EDORA - Euro	HTEC_emp_rtp	Employed persons in high and medium tech manufacturing and construction	2004	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 261
EDORA - Euro	HTEC_emp_ix	Employed persons in high and medium tech manufacturing and construction	2004	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 262
EDORA - Euro	FIR_web_rtp	Proportion of firms with own website regression	2002	0704 Economy - Research and innovation	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 263
EDORA - Euro	AREA_rte	Total land area	2000	9999 Cross-thematic and non-thematic	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 264
EDORA - Euro	DENS_ch_rtd	Density change	Between 2000 and 2006	0201 Demography - Population structure	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 265
EDORA - Euro	DENS_rde	Population density	Between 2006 and 2007	9901 Cross-thematic and non-thematic	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 266
EDORA - Euro	ACC_pop_car_rte	Daily population accessible by car	1999	0301 Transport - Accessibility	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 267
EDORA - Euro	BRO_acc_rtp	households with broadband access	2008	0203 Demography - Households	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 268
EDORA - Euro	INT_acc_rtp	households with internet at home	2008	0203 Demography - Households	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 269

- Regional data (6)

Provider	Indicator - code	Indicator - name	Year(s)	Theme - Sub theme	Geo object	Geo version	Dataset title	Category
EDORA - Euro	EDU_pre_rtp	Students with pre-primary education	Between 2005 and 2006	0601 Social Affairs - Education	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 270
EDORA - Euro	EDU_prm_rtp	Students with primary education	Between 2005 and 2006	0601 Social Affairs - Education	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 271
EDORA - Euro	EDU_sec_rtp	Students with secondary education	Between 2005 and 2006	0601 Social Affairs - Education	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 272
EDORA - Euro	EDU_ups_rtp	Students with upper secondary education	Between 2005 and 2006	0601 Social Affairs - Education	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 273
EDORA - Euro	EDU_pos_rtp	Students with post secondary education	Between 2005 and 2006	0601 Social Affairs - Education	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 274
EDORA - Euro	EDU_trt_rtp	Students with tertiary education	Between 2005 and 2006	0601 Social Affairs - Education	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 275
EDORA - Euro	HOS_bed_rtd	Beds in hospital	2005	0603 Social Affairs - Living conditions	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 276
EDORA - Euro	HLD_rtc	Total number of holdings	2005	0101 Agriculture and Fisheries - Farm	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 277
EDORA - Euro	HLD_2_rtc	Total number of holdings with less than 2 ESU	2005	0101 Agriculture and Fisheries - Farm	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 278
EDORA - Euro	HLD_2-100_rtc	Total number of holdings with 2 to 100 ESU	2005	0101 Agriculture and Fisheries - Farm	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 279
EDORA - Euro	HLD_100+_rtc	Total number of holdings with 100 ESU and over	2005	0101 Agriculture and Fisheries - Farm	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 280
EDORA - Euro	HLD_ch_rtd	Change in total number of holdings	Between 2000 and 2005	0101 Agriculture and Fisheries - Farm	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 281
EDORA - Euro	HLD_ch_2_rtd	Change in number of holdings with less than 2 ESU	Between 2000 and 2005	0101 Agriculture and Fisheries - Farm	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 282
EDORA - Euro	HLD_ch_2-100_rtd	Change in number of holdings with 2 to 100 ESU	Between 2000 and 2005	0101 Agriculture and Fisheries - Farm	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 283
EDORA - Euro	HLD_ch_100+_rtd	Change in number of holdings with 100 ESU and over	Between 2000 and 2005	0101 Agriculture and Fisheries - Farm	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 284
EDORA - Euro	HLD_ft_rtp	Holders working at full time	Between 2000 and 2005	0101 Agriculture and Fisheries - Farm	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 285
EDORA - Euro	HLD_ch_ft_rtd	Change of holders working at full time	Between 2000 and 2005	0101 Agriculture and Fisheries - Farm	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 286
EDORA - Euro	FRM_siz_rte	Economic farm size	2007	0101 Agriculture and Fisheries - Farm	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 287
EDORA - Euro	FRM_OGA_rtc	Farmers with OGA	2003	0101 Agriculture and Fisheries - Farm	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 288
EDORA - Euro	HLD_55+_rtp	Share of holders above 55 years	2007	0101 Agriculture and Fisheries - Farm	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 289
EDORA - Euro	HLD_35+_rtp	Share of holders under 35 years	2007	0101 Agriculture and Fisheries - Farm	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 290
EDORA - Euro	HLD_ch_35+_rtd	Change of the share of holders under 35 years	Between 2000 and 2005	0101 Agriculture and Fisheries - Farm	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 291
EDORA - Euro	HLD_ch_55+_rtd	Change of the share of holders above 55 years	Between 2000 and 2005	0101 Agriculture and Fisheries - Farm	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 292
EDORA - Euro	FRM_edu_prm_rtp	Share of farmers with basic and full education in agriculture	2000	0101 Agriculture and Fisheries - Farm	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 293
EDORA - Euro	GDP_eur_rte	GDP in Euros	2005	0701 Economy - Aggregated account	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 294
EDORA - Euro	GDP_cap_pps_rtd	GDP per capita in PPS	2005	0701 Economy - Aggregated account	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 295
EDORA - Euro	GDP_cap_eur_ix	GDP per capita in Euros as regard to the EU average	2005	0701 Economy - Aggregated account	NUTS	NUTS 2006	EDORA Country Profiles	Regional data - 296
EDORA - Euro	UMP_rtp	Unemployment rate	2008	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Future Perspective	Regional data - 297
EDORA - Euro	EMP_prim_priv_rtp	Share of primary sector employment in total private sector	2006	0701 Economy - Aggregated account	NUTS	NUTS 2006	EDORA Future Perspective	Regional data - 298
EDORA - Euro	GVA_prim_priv_rtp	Share of private GVA from primary sector	2006	0703 Economy - Production and cost	NUTS	NUTS 2006	EDORA Future Perspective	Regional data - 299
EDORA - Euro	EMP_sec_priv_rtp	Share of secondary sector employment in private sector	2006	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Future Perspective	Regional data - 300
EDORA - Euro	GVA_sec_priv_rtp	Share of private GVA from secondary sector	2006	0703 Economy - Production and cost	NUTS	NUTS 2006	EDORA Future Perspective	Regional data - 301
EDORA - Euro	FRM_siz_rtd	Average farm size	2005	0101 Agriculture and Fisheries - Farm	NUTS	NUTS 2006	EDORA Future Perspective	Regional data - 302
EDORA - Euro	HLD_2_rtp	Share of holdings with less than 2 ESU	2005	0101 Agriculture and Fisheries - Farm	NUTS	NUTS 2006	EDORA Future Perspective	Regional data - 303
EDORA - Euro	HLD_100+_rtp	Share of holdings with more than 2 ESU	2005	0101 Agriculture and Fisheries - Farm	NUTS	NUTS 2006	EDORA Future Perspective	Regional data - 304
EDORA - Euro	ACC_nat_rtc	Access to natural area	2008	0401 Energy and Environment - Energy	NUTS	NUTS 2006	EDORA Future Perspective	Regional data - 305
EDORA - Euro	BED_cap_rtd	Bed place per capita	Between 2006 and 2008	0703 Economy - Production and cost	NUTS	NUTS 2006	EDORA Future Perspective	Regional data - 306
EDORA - Euro	ECO_ACT_rtp	Economic activity rate	2008	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Future Perspective	Regional data - 307
EDORA - Euro	DR_rtd	Dependency rate	2006	0201 Demography - Population structure	NUTS	NUTS 2006	EDORA Future Perspective	Regional data - 308
EDORA - Euro	DR_65+_rtd	Old dependency rate	2006	0201 Demography - Population structure	NUTS	NUTS 2006	EDORA Future Perspective	Regional data - 309
EDORA - Euro	DR_15+_rtd	Young dependency rate	2006	0201 Demography - Population structure	NUTS	NUTS 2006	EDORA Future Perspective	Regional data - 310
EDORA - Euro	POP_65+_rtp	Share of population aged 65 years and more	2006	0201 Demography - Population structure	NUTS	NUTS 2006	EDORA Future Perspective	Regional data - 311
EDORA - Euro	NAT_incr_rtd	Natural increase rate	Between 2001 and 2005	0202 Demography - Natural changes	NUTS	NUTS 2006	EDORA Future Perspective	Regional data - 312
EDORA - Euro	MIG_rtd	Net migration rate	Between 2001 and 2005	0204 Demography - Migrations	NUTS	NUTS 2006	EDORA Future Perspective	Regional data - 313
EDORA - Euro	DENS_rde	Population density	2007	9901 Cross-thematic and non-thematic	NUTS	NUTS 2006	EDORA Future Perspective	Regional data - 314
EDORA - Euro	FP_EDORA_5_orc	Typology on rural economies	2009	9901 Cross-thematic and non-thematic	NUTS	NUTS 2006	EDORA Future Perspective	Regional data - 315
EDORA - Euro	GDP_cap_pps_rtd	GDP per capita in PPS	2006	0701 Economy - Aggregated account	NUTS	NUTS 2006	EDORA Future Perspective	Regional data - 316
EDORA - Euro	EMP_rtp	Employment rate	2006	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Future Perspective	Regional data - 317
EDORA - Euro	BED_cap_rtd	Bed place per capita	2007	0703 Economy - Production and cost	NUTS	NUTS 2006	EDORA Typology	Regional data - 318
EDORA - Euro	NGT_nres_cap_rtd	Nights spent by non residents per capita	2008	0703 Economy - Production and cost	NUTS	NUTS 2006	EDORA Typology	Regional data - 319
EDORA - Euro	NGT_cap_rtd	Nights spent per capita	2008	0703 Economy - Production and cost	NUTS	NUTS 2006	EDORA Typology	Regional data - 320
EDORA - Euro	ACC_nat_rtc	Access to natural area	2008	0401 Energy and Environment - Energy	NUTS	NUTS 2006	EDORA Typology	Regional data - 321
EDORA - Euro	HLD_oga_rtp	Share of holdings with other gainful activity	2005	0101 Agriculture and Fisheries - Farm	NUTS	NUTS 2006	EDORA Typology	Regional data - 322
EDORA - Euro	HLD_4+_rtp	Share of holdings with more than 4 ESU	2007	0101 Agriculture and Fisheries - Farm	NUTS	NUTS 2006	EDORA Typology	Regional data - 323

- Regional data (7)

Provider	Indicator - code	Indicator - name	Year(s)	Theme - Sub theme	Geo object	Geo version	Dataset title	Category
EDORA - Euro EMP_ce-gp_rtp		Ratio of employment from NACE CE to GP	2007	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Typology	Regional data - 324
EDORA - Euro GVA_ce-gk_rtp		Ratio of GVA from NACE CE to GK	2007	0703 Economy - Production and cost	NUTS	NUTS 2006	EDORA Typology	Regional data - 325
EDORA - Euro GDP_ch_rtd		Annual average GDP change	Between 1995 and 2006	0701 Economy - Aggregated account	NUTS	NUTS 2006	EDORA Typology	Regional data - 326
EDORA - Euro GDP_cap,pps_rtd		GDP per capita in PPS	2007	0701 Economy - Aggregated account	NUTS	NUTS 2006	EDORA Typology	Regional data - 327
EDORA - Euro MIG_rtd		Net migration rate	Between 2001 and 2005	0204 Demography - Migrations	NUTS	NUTS 2006	EDORA Typology	Regional data - 328
EDORA - Euro GVA_ce-gp_rtp		Ratio of GVA from NACE CE to GP	2007	0703 Economy - Production and cost	NUTS	NUTS 2006	EDORA Typology	Regional data - 329
EDORA - Euro GVA_cf-gp_rtp		Ratio of GVA from NACE CF to GP	2007	0703 Economy - Production and cost	NUTS	NUTS 2006	EDORA Typology	Regional data - 330
EDORA - Euro EMP_ce-gk_rtp		Ratio of employment from NACE CE to GK	2007	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Typology	Regional data - 331
EDORA - Euro GVA_prim_rtp		Share of GVA from primary sector	2006	0703 Economy - Production and cost	NUTS	NUTS 2006	EDORA Typology	Regional data - 332
EDORA - Euro EMP_prim_rtp		Share of employed persons in primary sector	2006	0701 Economy - Aggregated account	NUTS	NUTS 2006	EDORA Typology	Regional data - 333
EDORA - Euro EMP_prim_priv_rtp		Share of primary sector employment in total private sector	2006	0701 Economy - Aggregated account	NUTS	NUTS 2006	EDORA Typology	Regional data - 334
EDORA - Euro GVA_prim_priv_rtp		Share of private GVA from primary sector	2006	0703 Economy - Production and cost	NUTS	NUTS 2006	EDORA Typology	Regional data - 335
EDORA - Euro AWU_emp_priv_rtp		AWU as a share of total private employment	2007	0701 Economy - Aggregated account	NUTS	NUTS 2006	EDORA Typology	Regional data - 336
EDORA - Euro EMP_hot,cat_rtp		Share of employed persons in hotels and catering	2007	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Typology	Regional data - 337
EDORA - Euro NGT_res,cap_rtd		Nights spent by residents per capita	2008	0703 Economy - Production and cost	NUTS	NUTS 2006	EDORA Typology	Regional data - 338
EDORA - Euro UMP_rtp		Unemployment rate	2007	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Typology	Regional data - 339
EDORA - Euro EMP_ch_rtd		Annual average employment change	Between 1995 and 2006	0702 Economy - Employment	NUTS	NUTS 2006	EDORA Typology	Regional data - 340
EDORA - Euro CompType		Combining Urban-Rural Typology	2010	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	EDORA Typology Data	Regional data - 341
EDORA - Euro A-D Type		Performance typology	2010	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	EDORA Typology Data	Regional data - 342
EDORA - Euro SType		Structural Typology	2010	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	EDORA Typology Data	Regional data - 343
EDORA - Euro DTypeNo		Urban-Rural Typology (Dijkstra-Poelma Types)	2008	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	EDORA Typology Data	Regional data - 344
ESPON CLIMA EXP_HR		Change in annual mean number of days with heavy rainfall	Between 1961 and 2100	0402 Energy and Environment - Clima	NUTS	NUTS 2006	espon_climate_data	Regional data - 345
ESPON CLIMA EXP_SRO		Change in annual mean surface runoff	Between 1961 and 2100	0402 Energy and Environment - Clima	NUTS	NUTS 2006	espon_climate_data	Regional data - 346
ESPON CLIMA EXP_EVAP		Change in annual mean evaporation	Between 1961 and 2100	0402 Energy and Environment - Clima	NUTS	NUTS 2006	espon_climate_data	Regional data - 347
ESPON CLIMA EXP_SNC		Change in annual mean number of days with snow cover	Between 1961 and 2100	0402 Energy and Environment - Clima	NUTS	NUTS 2006	espon_climate_data	Regional data - 348
ESPON CLIMA EXP_TMEAN		Change in annual mean temperature	Between 1961 and 2100	0402 Energy and Environment - Clima	NUTS	NUTS 2006	espon_climate_data	Regional data - 349
ESPON CLIMA EXP_FD		Change in annual mean number of frost days	Between 1961 and 2100	0402 Energy and Environment - Clima	NUTS	NUTS 2006	espon_climate_data	Regional data - 350
ESPON CLIMA EXP_SD		Change in annual mean number of summer days	Between 1961 and 2100	0402 Energy and Environment - Clima	NUTS	NUTS 2006	espon_climate_data	Regional data - 351
ESPON CLIMA EXP_PW		Change in annual mean precipitation in winter months	Between 1961 and 2100	0402 Energy and Environment - Clima	NUTS	NUTS 2006	espon_climate_data	Regional data - 352
ESPON CLIMA EXP_PS		Change in annual mean precipitation in summer months	Between 1961 and 2100	0402 Energy and Environment - Clima	NUTS	NUTS 2006	espon_climate_data	Regional data - 353
ESPON Datab: age_0-4_t		0-4years, total	Between 1990 and 2000	0201 Demography - Population struct	NUTS	NUTS 1999	age_structure_new_cronos	Regional data - 354
ESPON Datab: age_35-39_t		35-39years, total	Between 1990 and 2000	0201 Demography - Population struct	NUTS	NUTS 1999	age_structure_new_cronos	Regional data - 355
ESPON Datab: age_40-44_t		40-44years, total	Between 1990 and 2000	0201 Demography - Population struct	NUTS	NUTS 1999	age_structure_new_cronos	Regional data - 356
ESPON Datab: age_45-49_t		45-49years, total	Between 1990 and 2000	0201 Demography - Population struct	NUTS	NUTS 1999	age_structure_new_cronos	Regional data - 357
ESPON Datab: age_50-54_t		50-54years, total	Between 1990 and 2000	0201 Demography - Population struct	NUTS	NUTS 1999	age_structure_new_cronos	Regional data - 358
ESPON Datab: age_55-59_t		55-59years, total	Between 1990 and 2000	0201 Demography - Population struct	NUTS	NUTS 1999	age_structure_new_cronos	Regional data - 359
ESPON Datab: age_60-64_t		60-64years, total	Between 1990 and 2000	0201 Demography - Population struct	NUTS	NUTS 1999	age_structure_new_cronos	Regional data - 360
ESPON Datab: age_65-69_t		65-69years, total	Between 1990 and 2000	0201 Demography - Population struct	NUTS	NUTS 1999	age_structure_new_cronos	Regional data - 361
ESPON Datab: age_70-74_t		70-74years, total	Between 1990 and 2000	0201 Demography - Population struct	NUTS	NUTS 1999	age_structure_new_cronos	Regional data - 362
ESPON Datab: age_75-79_t		75-79years, total	Between 1990 and 2000	0201 Demography - Population struct	NUTS	NUTS 1999	age_structure_new_cronos	Regional data - 363
ESPON Datab: age_80-84_t		80-84years, total	Between 1990 and 2000	0201 Demography - Population struct	NUTS	NUTS 1999	age_structure_new_cronos	Regional data - 364
ESPON Datab: age_85-89_t		85-89years, total	Between 1990 and 2000	0201 Demography - Population struct	NUTS	NUTS 1999	age_structure_new_cronos	Regional data - 365
ESPON Datab: age_90+_t		90+ years, total	Between 1990 and 2000	0201 Demography - Population struct	NUTS	NUTS 1999	age_structure_new_cronos	Regional data - 366
ESPON Datab: age_85+_t		85+ years, total	Between 1990 and 2000	0201 Demography - Population struct	NUTS	NUTS 1999	age_structure_new_cronos	Regional data - 367
ESPON Datab: age_70+_t		70+ years, total	Between 1990 and 2000	0201 Demography - Population struct	NUTS	NUTS 1999	age_structure_new_cronos	Regional data - 368
ESPON Datab: pop_t		Total population	Between 1990 and 2000	0201 Demography - Population struct	NUTS	NUTS 1999	age_structure_new_cronos	Regional data - 369
ESPON Datab: age_30-34_t		30-34years, total	Between 1990 and 2000	0201 Demography - Population struct	NUTS	NUTS 1999	age_structure_new_cronos	Regional data - 370
ESPON Datab: age_5-9_t		5-9years, total	Between 1990 and 2000	0201 Demography - Population struct	NUTS	NUTS 1999	age_structure_new_cronos	Regional data - 371
ESPON Datab: age_10-14_t		10-14years, total	Between 1990 and 2000	0201 Demography - Population struct	NUTS	NUTS 1999	age_structure_new_cronos	Regional data - 372
ESPON Datab: age_15-19_t		15-19years, total	Between 1990 and 2000	0201 Demography - Population struct	NUTS	NUTS 1999	age_structure_new_cronos	Regional data - 373
ESPON Datab: age_20-24_t		20-24years, total	Between 1990 and 2000	0201 Demography - Population struct	NUTS	NUTS 1999	age_structure_new_cronos	Regional data - 374
ESPON Datab: age_25-29_t		25-29years, total	Between 1990 and 2000	0201 Demography - Population struct	NUTS	NUTS 1999	age_structure_new_cronos	Regional data - 375
ESPON Datab: pop_t		Total population	Between 2005 and 2030	0201 Demography - Population struct	NUTS	NUTS 2003	Basic indicators	Regional data - 376
ESPON Datab: gdp_eur		GDP in euro	Between 2000 and 2006	0701 Economy - Aggregated account	NUTS	NUTS 2006	Basic indicators_delivery june 21	Regional data - 377

- Regional data (8)

Provider	Indicator - code	Indicator - name	Year(s)	Theme - Sub theme	Geo object	Geo version	Dataset title	Category
ESPON Datab:	pop_t	Total population	Between 2000 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 378	
ESPON Datab:	age_85+_f	80+years, females	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 379	
ESPON Datab:	age_80-84_f	80-84years, females	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 380	
ESPON Datab:	age_75-79_f	75-79years, females	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 381	
ESPON Datab:	age_70-74_f	70-74years, females	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 382	
ESPON Datab:	age_65-69_f	65-69years, females	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 383	
ESPON Datab:	age_60-64_f	60-64years, females	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 384	
ESPON Datab:	age_55-59_f	55-59years, females	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 385	
ESPON Datab:	age_50-54_f	50-54years, females	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 386	
ESPON Datab:	age_45-49_f	45-49years, females	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 387	
ESPON Datab:	age_40-44_f	40-44years, females	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 388	
ESPON Datab:	age_35-39_f	35-39years, females	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 389	
ESPON Datab:	age_30-34_f	30-34years, females	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 390	
ESPON Datab:	age_25-29_f	25-29years, females	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 391	
ESPON Datab:	gdp_pps	GDP in PPS	Between 2000 and 2006	0701 Economy - Aggregated account	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 392	
ESPON Datab:	age_15-19_f	15-19years, females	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 393	
ESPON Datab:	age_10-14_f	10-14years, females	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 394	
ESPON Datab:	age_5-9_f	5-9years, females	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 395	
ESPON Datab:	age_0-4_f	0-4years, females	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 396	
ESPON Datab:	age_85+_m	80+years, males	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 397	
ESPON Datab:	age_80-84_m	80-84years, males	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 398	
ESPON Datab:	age_75-79_m	75-79years, males	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 399	
ESPON Datab:	age_70-74_m	70-74years, males	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 400	
ESPON Datab:	age_65-69_m	65-69years, males	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 401	
ESPON Datab:	age_60-64_m	60-64years, males	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 402	
ESPON Datab:	age_55-59_m	55-59years, males	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 403	
ESPON Datab:	age_50-54_m	50-54years, males	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 404	
ESPON Datab:	age_45-49_m	45-49years, males	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 405	
ESPON Datab:	age_40-44_m	40-44years, males	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 406	
ESPON Datab:	age_35-39_m	35-39years, males	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 407	
ESPON Datab:	age_30-34_m	30-34years, males	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 408	
ESPON Datab:	age_25-29_m	25-29years, males	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 409	
ESPON Datab:	age_20-24_m	20-24years, males	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 410	
ESPON Datab:	age_15-19_m	15-19years, males	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 411	
ESPON Datab:	age_10-14_m	10-14years, males	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 412	
ESPON Datab:	age_5-9_m	5-9years, males	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 413	
ESPON Datab:	age_0-4_m	0-4years, males	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 414	
ESPON Datab:	age_20-24_f	20-24years, females	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 415	
ESPON Datab:	unemp	Unemployment	Between 2000 and 2007	0702 Economy - Employment	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 416	
ESPON Datab:	activ	Active population	Between 2000 and 2007	0702 Economy - Employment	NUTS	NUTS 2006	Basic indicators, delivery june 21 Regional data - 417	
ESPON Datab:	GDP_person_empl	GDP (Purchasing Power Parities) per person employed, EU2	2004	0701 Economy - Aggregated account	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s Regional data - 418	
ESPON Datab:	GDP_head_UE27=11	GDP (Purchasing Power Parities) per capita, EU27=100	2004	0701 Economy - Aggregated account	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s Regional data - 419	
ESPON Datab:	Emp_pop_fem(15-64)	Employed rate, female, 15-64 years	2005	0702 Economy - Employment	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s Regional data - 420	
ESPON Datab:	Emp_pop(55-64)	Employment rate, 55-64 years	2005	0702 Economy - Employment	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s Regional data - 421	
ESPON Datab:	Pop_age_>65years	Share of population above the retirement age	2004	0201 Demography - Population struct	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s Regional data - 422	
ESPON Datab:	Pop_age_15-64year	Share of population in the working age	2004	0201 Demography - Population struct	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s Regional data - 423	
ESPON Datab:	Pop_age_<15years	Share of young people	2004	0201 Demography - Population struct	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s Regional data - 424	
ESPON Datab:	R_D_exp_B_ent	Expenditure on research and development	2004	0704 Economy - Research and innov	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s Regional data - 425	
ESPON Datab:	R_D_exp	Expenditure in research and development of GDP	2004	0704 Economy - Research and innov	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s Regional data - 426	
ESPON Datab:	Econ_Lisb_ind	Economic Lisbon Indicators	Between 2004 and 2005	0704 Economy - Research and innov	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s Regional data - 427	
ESPON Datab:	Pop_den	Population density	2004	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s Regional data - 428	
ESPON Datab:	Pop_grow_aver	average annual population growth	Between 1995 and 2004	0202 Demography - Natural changes	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s Regional data - 429	
ESPON Datab:	Pop_tot	Total population	2004	0201 Demography - Population struct	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s Regional data - 430	
ESPON Datab:	Emp_pop(15-64)	Employment rate, 15-64 years	2005	0702 Economy - Employment	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s Regional data - 431	

- Regional data (9)

Provider	Indicator - code	Indicator - name	Year(s)	Theme - Sub theme	Geo object	Geo version	Dataset title	Category
ESPON Datab:	Unemp_young	Youth unemployment rate	2005	0702 Economy - Employment	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s	Regional data - 432
ESPON Datab:	Unemp_W	Female unemployment rate	2005	0702 Economy - Employment	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s	Regional data - 433
ESPON Datab:	Unemp_LT	Long-term unemployment rate	2005	0702 Economy - Employment	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s	Regional data - 434
ESPON Datab:	Unemp	Unemployment rate	2005	0702 Economy - Employment	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s	Regional data - 435
ESPON Datab:	Emp_serv	Share of employed persons in services	2005	0702 Economy - Employment	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s	Regional data - 436
ESPON Datab:	Emp_ind	Share of employed persons in industry	2005	0702 Economy - Employment	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s	Regional data - 437
ESPON Datab:	Emp_agri	Share of employed persons in agriculture	2005	0702 Economy - Employment	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s	Regional data - 438
ESPON Datab:	Growth_GDP	GDP Growth	Between 1995 and 2004	0701 Economy - Aggregated account	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s	Regional data - 439
ESPON Datab:	Hig_educ_att_pop2	High educational level, 25-59 years	2005	0601 Social Affairs - Education	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s	Regional data - 440
ESPON Datab:	Med_educ_att_pop	Medium educational level, 25-59 years	2005	0601 Social Affairs - Education	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s	Regional data - 441
ESPON Datab:	Low_educ_att_pop	Low educational level, 25-59 years	2005	0601 Social Affairs - Education	NUTS	NUTS 2003, NUTS 2006	Fourth report on economic and s	Regional data - 442
ESPON Datab:	reg_aland_CLC_111	Continuous urban fabric	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 443
ESPON Datab:	reg_aland_CLC_2	Agricultural areas	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 444
ESPON Datab:	reg_aland_CLC_3	Forest and semi natural areas	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 445
ESPON Datab:	reg_aland_CLC_4	Wetlands	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 446
ESPON Datab:	reg_aland_CLC_5	Water bodies	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 447
ESPON Datab:	reg_aland_CLC_11	Urban fabric	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 448
ESPON Datab:	reg_aland_CLC_12	Industrial, commercial and transport units	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 449
ESPON Datab:	reg_aland_CLC_13	Mine, dump and construction sites	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 450
ESPON Datab:	reg_aland_CLC_14	Artificial, non-agricultural vegetated areas	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 451
ESPON Datab:	reg_aland_CLC_21	Arable land	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 452
ESPON Datab:	reg_aland_CLC_22	Permanent crops	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 453
ESPON Datab:	reg_aland_CLC_23	Pastures	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 454
ESPON Datab:	reg_aland_CLC_24	Heterogeneous agricultural areas	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 455
ESPON Datab:	reg_aland_CLC_31	Forests	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 456
ESPON Datab:	reg_aland_CLC_32	Scrub and/or herbaceous vegetation associations	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 457
ESPON Datab:	reg_aland_CLC_33	Open spaces with little or no vegetation	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 458
ESPON Datab:	reg_aland_CLC_41	Inland wetlands	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 459
ESPON Datab:	reg_aland_CLC_42	Maritime wetlands	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 460
ESPON Datab:	reg_aland_CLC_51	Inland waters	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 461
ESPON Datab:	reg_aland_CLC_52	Marine waters	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 462
ESPON Datab:	reg_aland_CLC_523	Seas and oceans	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 463
ESPON Datab:	reg_aland_no_data	Not provided	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 464
ESPON Datab:	reg_aland_CLC_524	Estuaries	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 465
ESPON Datab:	reg_aland_CLC_521	Coastal lagoons	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 466
ESPON Datab:	reg_aland_CLC_512	Water bodies	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 467
ESPON Datab:	reg_aland_CLC_511	Water courses	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 468
ESPON Datab:	reg_aland_CLC_423	Intertidal flats	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 469
ESPON Datab:	reg_aland_CLC_422	Salines	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 470
ESPON Datab:	reg_aland_CLC_421	Salt marshes	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 471
ESPON Datab:	reg_aland_CLC_412	Peat bogs	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 472
ESPON Datab:	reg_aland_CLC_411	Inland marshes	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 473
ESPON Datab:	reg_aland_CLC_334	Glaciers and perpetual snow	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 474
ESPON Datab:	reg_aland_CLC_334	Burnt areas	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 475
ESPON Datab:	reg_aland_CLC_333	Sparsely vegetated areas	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 476
ESPON Datab:	reg_aland_CLC_332	Bare rocks	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 477
ESPON Datab:	reg_aland_CLC_331	Beaches, dunes, sands	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 478
ESPON Datab:	reg_aland_CLC_324	Transitional woodland-shrub	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 479
ESPON Datab:	reg_aland_CLC_323	Sclerophyllous vegetation	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 480
ESPON Datab:	reg_aland_CLC_322	Moors and heathland	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 481
ESPON Datab:	reg_aland_CLC_321	Natural grasslands	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 482
ESPON Datab:	reg_aland_CLC_312	Mixed forest	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 483
ESPON Datab:	reg_aland_CLC_312	Coniferous forest	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 484
ESPON Datab:	reg_aland_CLC_311	Broad-leaved forest	1990	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 1990	Regional data - 485

- Regional data (10)

Provider	Indicator - code	Indicator - name	Year(s)	Theme - Sub theme	Geo object	Geo version	Dataset title	Category
ESPON Datab: reg_aland_CLC_244	Agro-forestry areas	1990 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 1990	Regional data - 486			
ESPON Datab: reg_aland_CLC_243	Land principally occupied by agriculture, with significant are	1990 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 1990	Regional data - 487			
ESPON Datab: reg_aland_CLC_242	Complex cultivation patterns	1990 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 1990	Regional data - 488			
ESPON Datab: reg_aland_CLC_241	Annual crops associated with permanent crops	1990 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 1990	Regional data - 489			
ESPON Datab: reg_aland_CLC_231	Pastures	1990 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 1990	Regional data - 490			
ESPON Datab: reg_aland_CLC_223	Olive groves	1990 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 1990	Regional data - 491			
ESPON Datab: reg_aland_CLC_222	Fruit trees and berry plantations	1990 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 1990	Regional data - 492			
ESPON Datab: reg_aland_CLC_221	Vineyards	1990 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 1990	Regional data - 493			
ESPON Datab: reg_aland_CLC_213	Rice fields	1990 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 1990	Regional data - 494			
ESPON Datab: reg_aland_CLC_212	Permanently irrigated land	1990 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 1990	Regional data - 495			
ESPON Datab: reg_aland_CLC_211	Non-irrigated arable land	1990 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 1990	Regional data - 496			
ESPON Datab: reg_aland_CLC_142	Sport and leisure facilities	1990 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 1990	Regional data - 497			
ESPON Datab: reg_aland_CLC_141	Green urban areas	1990 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 1990	Regional data - 498			
ESPON Datab: reg_aland_CLC_133	Construction sites	1990 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 1990	Regional data - 499			
ESPON Datab: reg_aland_CLC_132	Dump sites	1990 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 1990	Regional data - 500			
ESPON Datab: reg_aland_CLC_131	Mineral extraction sites	1990 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 1990	Regional data - 501			
ESPON Datab: reg_aland_CLC_124	Airports	1990 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 1990	Regional data - 502			
ESPON Datab: reg_aland_CLC_123	Port areas	1990 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 1990	Regional data - 503			
ESPON Datab: reg_aland_CLC_122	Road and rail networks and associated land	1990 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 1990	Regional data - 504			
ESPON Datab: reg_aland_CLC_121	Industrial or commercial units	1990 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 1990	Regional data - 505			
ESPON Datab: reg_aland_CLC_112	Discontinuous urban fabric	1990 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 1990	Regional data - 506			
ESPON Datab: reg_aland_CLC_1	Artificial surfaces	1990 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 1990	Regional data - 507			
ESPON Datab: reg_aland_CLC_121	Industrial or commercial units	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 508			
ESPON Datab: reg_aland_CLC_112	Discontinuous urban fabric	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 509			
ESPON Datab: reg_aland_CLC_111	Continuous urban fabric	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 510			
ESPON Datab: reg_aland_CLC_52	Marine waters	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 511			
ESPON Datab: reg_aland_CLC_51	Inland waters	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 512			
ESPON Datab: reg_aland_CLC_42	Maritime wetlands	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 513			
ESPON Datab: reg_aland_CLC_41	Inland wetlands	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 514			
ESPON Datab: reg_aland_CLC_33	Open spaces with little or no vegetation	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 515			
ESPON Datab: reg_aland_CLC_32	Scrub and/or herbaceous vegetation associations	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 516			
ESPON Datab: reg_aland_CLC_31	Forests	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 517			
ESPON Datab: reg_aland_CLC_24	Heterogeneous agricultural areas	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 518			
ESPON Datab: reg_aland_CLC_23	Pastures	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 519			
ESPON Datab: reg_aland_CLC_22	Permanent crops	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 520			
ESPON Datab: reg_aland_CLC_21	Arable land	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 521			
ESPON Datab: reg_aland_CLC_14	Artificial, non-agricultural vegetated areas	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 522			
ESPON Datab: reg_aland_CLC_13	Mine, dump and construction sites	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 523			
ESPON Datab: reg_aland_CLC_12	Industrial, commercial and transport units	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 524			
ESPON Datab: reg_aland_CLC_11	Urban fabric	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 525			
ESPON Datab: reg_aland_CLC_5	Water bodies	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 526			
ESPON Datab: reg_aland_CLC_4	Wetlands	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 527			
ESPON Datab: reg_aland_CLC_3	Forest and semi natural areas	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 528			
ESPON Datab: reg_aland_CLC_2	Agricultural areas	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 529			
ESPON Datab: reg_aland_CLC_1	Artificial surfaces	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 530			
ESPON Datab: reg_aland_CLC_523	Seas and oceans	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 531			
ESPON Datab: reg_aland_no_data	Not provided	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 532			
ESPON Datab: reg_aland_CLC_522	Estuaries	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 533			
ESPON Datab: reg_aland_CLC_521	Coastal lagoons	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 534			
ESPON Datab: reg_aland_CLC_512	Water bodies	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 535			
ESPON Datab: reg_aland_CLC_511	Water courses	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 536			
ESPON Datab: reg_aland_CLC_423	Intertidal flats	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 537			
ESPON Datab: reg_aland_CLC_422	Salines	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 538			
ESPON Datab: reg_aland_CLC_421	Salt marshes	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 539			



- Regional data (11)

Provider	Indicator - code	Indicator - name	Year(s)	Theme - Sub theme	Geo object	Geo version	Dataset title	Category
ESPOL Datab: reg_aland_CLC_412	Peat bogs	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 540			
ESPOL Datab: reg_aland_CLC_411	Inland marshes	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 541			
ESPOL Datab: reg_aland_CLC_335	Glaciers and perpetual snow	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 542			
ESPOL Datab: reg_aland_CLC_334	Burnt areas	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 543			
ESPOL Datab: reg_aland_CLC_333	Sparsely vegetated areas	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 544			
ESPOL Datab: reg_aland_CLC_332	Bare rocks	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 545			
ESPOL Datab: reg_aland_CLC_331	Beaches, dunes, sands	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 546			
ESPOL Datab: reg_aland_CLC_324	Transitional woodland-shrub	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 547			
ESPOL Datab: reg_aland_CLC_323	Sclerophyllous vegetation	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 548			
ESPOL Datab: reg_aland_CLC_322	Moors and heathland	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 549			
ESPOL Datab: reg_aland_CLC_321	Natural grasslands	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 550			
ESPOL Datab: reg_aland_CLC_313	Mixed forest	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 551			
ESPOL Datab: reg_aland_CLC_312	Coniferous forest	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 552			
ESPOL Datab: reg_aland_CLC_311	Broad-leaved forest	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 553			
ESPOL Datab: reg_aland_CLC_244	Agro-forestry areas	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 554			
ESPOL Datab: reg_aland_CLC_243	Land principally occupied by agriculture, with significant are	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 555			
ESPOL Datab: reg_aland_CLC_242	Complex cultivation patterns	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 556			
ESPOL Datab: reg_aland_CLC_241	Annual crops associated with permanent crops	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 557			
ESPOL Datab: reg_aland_CLC_231	Pastures	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 558			
ESPOL Datab: reg_aland_CLC_223	Olive groves	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 559			
ESPOL Datab: reg_aland_CLC_222	Fruit trees and berry plantations	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 560			
ESPOL Datab: reg_aland_CLC_221	Vineyards	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 561			
ESPOL Datab: reg_aland_CLC_213	Rice fields	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 562			
ESPOL Datab: reg_aland_CLC_212	Permanently irrigated land	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 563			
ESPOL Datab: reg_aland_CLC_211	Non-irrigated arable land	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 564			
ESPOL Datab: reg_aland_CLC_142	Sport and leisure facilities	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 565			
ESPOL Datab: reg_aland_CLC_141	Green urban areas	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 566			
ESPOL Datab: reg_aland_CLC_133	Construction sites	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 567			
ESPOL Datab: reg_aland_CLC_132	Dump sites	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 568			
ESPOL Datab: reg_aland_CLC_131	Mineral extraction sites	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 569			
ESPOL Datab: reg_aland_CLC_124	Airports	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 570			
ESPOL Datab: reg_aland_CLC_123	Port areas	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 571			
ESPOL Datab: reg_aland_CLC_122	Road and rail networks and associated land	2000 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2000	Regional data - 572			
ESPOL Datab: reg_aland_CLC_122	Road and rail networks and associated land	2006 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2006	Regional data - 573			
ESPOL Datab: reg_aland_CLC_112	Discontinuous urban fabric	2006 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2006	Regional data - 574			
ESPOL Datab: reg_aland_CLC_111	Continuous urban fabric	2006 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2006	Regional data - 575			
ESPOL Datab: reg_aland_CLC_52	Marine waters	2006 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2006	Regional data - 576			
ESPOL Datab: reg_aland_CLC_51	Inland waters	2006 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2006	Regional data - 577			
ESPOL Datab: reg_aland_CLC_42	Maritime wetlands	2006 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2006	Regional data - 578			
ESPOL Datab: reg_aland_CLC_41	Inland wetlands	2006 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2006	Regional data - 579			
ESPOL Datab: reg_aland_CLC_33	Open spaces with little or no vegetation	2006 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2006	Regional data - 580			
ESPOL Datab: reg_aland_CLC_32	Scrub and/or herbaceous vegetation associations	2006 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2006	Regional data - 581			
ESPOL Datab: reg_aland_CLC_31	Forests	2006 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2006	Regional data - 582			
ESPOL Datab: reg_aland_CLC_24	Heterogeneous agricultural areas	2006 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2006	Regional data - 583			
ESPOL Datab: reg_aland_CLC_23	Pastures	2006 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2006	Regional data - 584			
ESPOL Datab: reg_aland_CLC_22	Permanent crops	2006 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2006	Regional data - 585			
ESPOL Datab: reg_aland_CLC_21	Arable land	2006 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2006	Regional data - 586			
ESPOL Datab: reg_aland_CLC_14	Artificial, non-agricultural vegetated areas	2006 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2006	Regional data - 587			
ESPOL Datab: reg_aland_CLC_13	Mine, dump and construction sites	2006 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2006	Regional data - 588			
ESPOL Datab: reg_aland_CLC_12	Industrial, commercial and transport units	2006 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2006	Regional data - 589			
ESPOL Datab: reg_aland_CLC_11	Urban fabric	2006 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2006	Regional data - 590			
ESPOL Datab: reg_aland_CLC_5	Water bodies	2006 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2006	Regional data - 591			
ESPOL Datab: reg_aland_CLC_4	Wetlands	2006 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2006	Regional data - 592			
ESPOL Datab: req_aland_CLC_3	Forest and semi natural areas	2006 0501 Land Use - Land use and land c NUTS	NUTS 2006	Land use data, 2006	Regional data - 593			

- Regional data (12)

Provider	Indicator - code	Indicator - name	Year(s)	Theme - Sub theme	Geo object	Geo version	Dataset title	Category
ESPN Datab:	reg_aland_CLC_2	Agricultural areas	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 594
ESPN Datab:	reg_aland_CLC_1	Artificial surfaces	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 595
ESPN Datab:	reg_aland_CLC_324	Transitional woodland-shrub	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 596
ESPN Datab:	reg_aland_CLC_331	Beaches, dunes, sands	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 597
ESPN Datab:	reg_aland_CLC_332	Bare rocks	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 598
ESPN Datab:	reg_aland_CLC_333	Sparsely vegetated areas	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 599
ESPN Datab:	reg_aland_CLC_334	Burnt areas	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 600
ESPN Datab:	reg_aland_CLC_335	Glaciers and perpetual snow	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 601
ESPN Datab:	reg_aland_CLC_411	Inland marshes	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 602
ESPN Datab:	reg_aland_CLC_412	Peat bogs	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 603
ESPN Datab:	reg_aland_CLC_421	Salt marshes	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 604
ESPN Datab:	reg_aland_CLC_422	Salines	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 605
ESPN Datab:	reg_aland_CLC_423	Intertidal flats	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 606
ESPN Datab:	reg_aland_CLC_511	Water courses	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 607
ESPN Datab:	reg_aland_CLC_512	Water bodies	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 608
ESPN Datab:	reg_aland_CLC_521	Coastal lagoons	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 609
ESPN Datab:	reg_aland_CLC_522	Estuaries	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 610
ESPN Datab:	reg_aland_no_data	Not provided	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 611
ESPN Datab:	reg_aland_CLC_523	Seas and oceans	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 612
ESPN Datab:	reg_aland_CLC_124	Airports	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 613
ESPN Datab:	reg_aland_CLC_123	Port areas	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 614
ESPN Datab:	reg_aland_CLC_121	Industrial or commercial units	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 615
ESPN Datab:	reg_aland_CLC_323	Sclerophyllous vegetation	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 616
ESPN Datab:	reg_aland_CLC_322	Moors and heathland	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 617
ESPN Datab:	reg_aland_CLC_321	Natural grasslands	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 618
ESPN Datab:	reg_aland_CLC_313	Mixed forest	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 619
ESPN Datab:	reg_aland_CLC_312	Coniferous forest	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 620
ESPN Datab:	reg_aland_CLC_311	Broad-leaved forest	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 621
ESPN Datab:	reg_aland_CLC_244	Agro-forestry areas	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 622
ESPN Datab:	reg_aland_CLC_243	Land principally occupied by agriculture, with significant are	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 623
ESPN Datab:	reg_aland_CLC_242	Complex cultivation patterns	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 624
ESPN Datab:	reg_aland_CLC_241	Annual crops associated with permanent crops	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 625
ESPN Datab:	reg_aland_CLC_231	Pastures	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 626
ESPN Datab:	reg_aland_CLC_223	Olive groves	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 627
ESPN Datab:	reg_aland_CLC_222	Fruit trees and berry plantations	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 628
ESPN Datab:	reg_aland_CLC_221	Vineyards	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 629
ESPN Datab:	reg_aland_CLC_213	Rice fields	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 630
ESPN Datab:	reg_aland_CLC_212	Permanently irrigated land	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 631
ESPN Datab:	reg_aland_CLC_211	Non-irrigated arable land	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 632
ESPN Datab:	reg_aland_CLC_142	Sport and leisure facilities	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 633
ESPN Datab:	reg_aland_CLC_141	Green urban areas	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 634
ESPN Datab:	reg_aland_CLC_133	Construction sites	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 635
ESPN Datab:	reg_aland_CLC_132	Dump sites	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 636
ESPN Datab:	reg_aland_CLC_131	Mineral extraction sites	2006	0501 Land Use - Land use and land c	NUTS	NUTS 2006	Land use data, 2006	Regional data - 637
ESPN Datab:	pop_t	Total population	Between 1970 and 2001	0201 Demography - Population struct	NUTS	NUTS 1999	Pop_new_cronos	Regional data - 638
ESPN Datab:	pop_f	Female population	Between 1970 and 2001	0201 Demography - Population struct	NUTS	NUTS 1999	Pop_new_cronos	Regional data - 639
ESPN Datab:	pop_m	Male population	Between 1970 and 2001	0201 Demography - Population struct	NUTS	NUTS 1999	Pop_new_cronos	Regional data - 640
ESPN Datab:	GDP_head_EU15=100	GDP (Purchasing Power Parities) per capita, EU15=100	Between 1987 and 1998	0701 Economy - Aggregated account	NUTS	NUTS 1999, NUTS 2003, I	Second report on economic and	Regional data - 641
ESPN Datab:	Emp_agri	Share of employed persons in agriculture	1999	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003, I	Second report on economic and	Regional data - 642
ESPN Datab:	Emp_ind	Share of employed persons in industry	1999	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003, I	Second report on economic and	Regional data - 643
ESPN Datab:	Emp_serv	Share of employed persons in services	1999	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003, I	Second report on economic and	Regional data - 644
ESPN Datab:	Eur_Patent_appli	European patent applications	Between 1996 and 1998	0704 Economy - Research and innov	NUTS	NUTS 1999, NUTS 2003, I	Second report on economic and	Regional data - 645
ESPN Datab:	Unemp	Unemployment rate	Between 1989 and 1999	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003, I	Second report on economic and	Regional data - 646
ESPN Datab:	Unemp_LT	Long-term unemployment rate	1999	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003, I	Second report on economic and	Regional data - 647

## • Regional data (13)

Provider	Indicator - code	Indicator - name	Year(s)	Theme - Sub theme	Geo object	Geo version	Dataset title	Category
ESPON Datab:	Unemp_W	Female unemployment rate	1999	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003,	Second report on economic and	Regional data - 648
ESPON Datab:	Unemp_young	Youth unemployment rate	1999	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003,	Second report on economic and	Regional data - 649
ESPON Datab:	Emp_pop(15-64)	Employment rate, 15-64 years	1999	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003,	Second report on economic and	Regional data - 650
ESPON Datab:	Emp_pop_fem(15-64)	Employed rate, female, 15-64 years	1999	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003,	Second report on economic and	Regional data - 651
ESPON Datab:	Emp_pop_male(15-64)	Employed rate, male, 15-64 years	1999	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003,	Second report on economic and	Regional data - 652
ESPON Datab:	Pop_tot	Total population	1997	0201 Demography - Population struct	NUTS	NUTS 1999, NUTS 2003,	Second report on economic and	Regional data - 653
ESPON Datab:	Pop_den	Population density	1997	9901 Cross-thematic and non-themat	NUTS	NUTS 1999, NUTS 2003,	Second report on economic and	Regional data - 654
ESPON Datab:	Pop_age_<15years	Share of young people	1997	0201 Demography - Population struct	NUTS	NUTS 1999, NUTS 2003,	Second report on economic and	Regional data - 655
ESPON Datab:	Pop_age_15-64years	Share of population in the working age	1997	0201 Demography - Population struct	NUTS	NUTS 1999, NUTS 2003,	Second report on economic and	Regional data - 656
ESPON Datab:	Pop_age_>65years	Share of population above the retirement age	1997	0201 Demography - Population struct	NUTS	NUTS 1999, NUTS 2003,	Second report on economic and	Regional data - 657
ESPON Datab:	Low_educ_att_pop	Low educational level, 25-59 years	1999	0601 Social Affairs - Education	NUTS	NUTS 1999, NUTS 2003,	Second report on economic and	Regional data - 658
ESPON Datab:	Med_educ_att_pop	Medium educational level, 25-59 years	1999	0601 Social Affairs - Education	NUTS	NUTS 1999, NUTS 2003,	Second report on economic and	Regional data - 659
ESPON Datab:	Hig_educ_att_pop2	High educational level, 25-59 years	1999	0601 Social Affairs - Education	NUTS	NUTS 1999, NUTS 2003,	Second report on economic and	Regional data - 660
ESPON Datab:	GDP_head_UE25=11	GDP (Purchasing Power Parities) per capita, EU25=100	2001	0701 Economy - Aggregated account	NUTS	NUTS 1999, NUTS 2003,	Third report on economic and so	Regional data - 661
ESPON Datab:	Unemp_young	Youth unemployment rate	2002	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003,	Third report on economic and so	Regional data - 662
ESPON Datab:	Unemp_W	Female unemployment rate	2002	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003,	Third report on economic and so	Regional data - 663
ESPON Datab:	Unemp_LT	Long-term unemployment rate	2002	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003,	Third report on economic and so	Regional data - 664
ESPON Datab:	Unemp	Unemployment rate	Between 1992 and 2002	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003,	Third report on economic and so	Regional data - 665
ESPON Datab:	Eur_Patent_appli	European patent applications	Between 1999 and 2001	0704 Economy - Research and innov	NUTS	NUTS 2003, NUTS 2006	Third report on economic and so	Regional data - 666
ESPON Datab:	Emp_serv	Share of employed persons in services	2002	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003,	Third report on economic and so	Regional data - 667
ESPON Datab:	Emp_ind	Share of employed persons in industry	2002	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003,	Third report on economic and so	Regional data - 668
ESPON Datab:	Emp_agri	Share of employed persons in agriculture	2002	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003,	Third report on economic and so	Regional data - 669
ESPON Datab:	Growth_GDP	GDP Growth	Between 1995 and 2001	0701 Economy - Aggregated account	NUTS	NUTS 1999, NUTS 2003,	Third report on economic and so	Regional data - 670
ESPON Datab:	Emp_pop(15-64)	Employment rate, 15-64 years	2002	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003,	Third report on economic and so	Regional data - 671
ESPON Datab:	GDP_head_EU15=11	GDP (Purchasing Power Parities) per capita, EU15=100	Between 2000 and 2001	0701 Economy - Aggregated account	NUTS	NUTS 1999, NUTS 2003,	Third report on economic and so	Regional data - 672
ESPON Datab:	Hig_educ_att_pop2	High educational level, 25-59 years	2002	0601 Social Affairs - Education	NUTS	NUTS 1999, NUTS 2003,	Third report on economic and so	Regional data - 673
ESPON Datab:	Med_educ_att_pop	Medium educational level, 25-59 years	2002	0601 Social Affairs - Education	NUTS	NUTS 1999, NUTS 2003,	Third report on economic and so	Regional data - 674
ESPON Datab:	Low_educ_att_pop	Low educational level, 25-59 years	2002	0601 Social Affairs - Education	NUTS	NUTS 1999, NUTS 2003,	Third report on economic and so	Regional data - 675
ESPON Datab:	Pop_age_>65years	Share of population above the retirement age	2000	0201 Demography - Population struct	NUTS	NUTS 1999, NUTS 2003,	Third report on economic and so	Regional data - 676
ESPON Datab:	Pop_age_15-64years	Share of population in the working age	2000	0201 Demography - Population struct	NUTS	NUTS 1999, NUTS 2003,	Third report on economic and so	Regional data - 677
ESPON Datab:	Pop_age_<15years	Share of young people	2000	0201 Demography - Population struct	NUTS	NUTS 1999, NUTS 2003,	Third report on economic and so	Regional data - 678
ESPON Datab:	Pop_den	Population density	2001	9901 Cross-thematic and non-themat	NUTS	NUTS 1999, NUTS 2003,	Third report on economic and so	Regional data - 679
ESPON Datab:	Pop_tot	Total population	2001	0201 Demography - Population struct	NUTS	NUTS 1999, NUTS 2003,	Third report on economic and so	Regional data - 680
ESPON Datab:	Emp_pop_male(15-64)	Employed rate, male, 15-64 years	2002	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003,	Third report on economic and so	Regional data - 681
ESPON Datab:	Emp_pop_fem(15-64)	Employed rate, female, 15-64 years	2002	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003,	Third report on economic and so	Regional data - 682
Lisbon Strateg	ESP2013_TYPOLISE	Change of Composit Lisbon performance benchmark	Between 2000 and 2006	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Lisbon strategy performance	Regional data - 683
Lisbon Strateg	ESP2013_TYPOLISE	Composit Lisbon performance benchmark	Between 2000 and 2006	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Lisbon strategy performance	Regional data - 684
DEMIFER	D_2064	Share of population aged 20-64 years	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Age Structure Data	Regional data - 685
DEMIFER	D_2039	Share of population aged 20-39 years	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Age Structure Data	Regional data - 686
DEMIFER	CH_80+	Change in Population Aged 80+	Between 2001 and 2005	0202 Demography - Natural changes	NUTS	NUTS 2006	Age Structure Data	Regional data - 687
DEMIFER	CH_75+	Change in Population Aged 75+	Between 2000 and 2007	0202 Demography - Natural changes	NUTS	NUTS 2006	Age Structure Data	Regional data - 688
DEMIFER	CH_65+	Change in Population Aged 65+	Between 2001 and 2005	0202 Demography - Natural changes	NUTS	NUTS 2006	Age Structure Data	Regional data - 689
DEMIFER	CH_5064	Change in Population Aged 50-64	Between 2001 and 2005	0202 Demography - Natural changes	NUTS	NUTS 2006	Age Structure Data	Regional data - 690
DEMIFER	CH_2064	Change in Population Aged 20-64	Between 2000 and 2007	0202 Demography - Natural changes	NUTS	NUTS 2006	Age Structure Data	Regional data - 691
DEMIFER	CH_2039	Change in Population Aged 20-39	Between 2001 and 2005	0202 Demography - Natural changes	NUTS	NUTS 2006	Age Structure Data	Regional data - 692
DEMIFER	DR_OLD_CH	Change in Old Age Dependency Ratio	Between 2001 and 2005	0201 Demography - Population struct	NUTS	NUTS 2006	Age Structure Data	Regional data - 693
DEMIFER	DR_OLD	Old Age Dependency Ratio	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Age Structure Data	Regional data - 694
DEMIFER	RCLUSTER	Cluster analysis of demographic growth	Between 2000 and 2007	0202 Demography - Natural changes	NUTS	NUTS 2006	Age Structure Data	Regional data - 695
DEMIFER	DR	Total Dependency Ratio	2005	0602 Social Affairs - Labour market	NUTS	NUTS 2006	Age Structure Data	Regional data - 696
DEMIFER	SEXR_2029	Sex Ratio at Age 20-29 Years	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Age Structure Data	Regional data - 697
DEMIFER	PASR	Parent Support Ratio	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Age Structure Data	Regional data - 698
DEMIFER	LF_REPL	Labour Force Replacement Ratio	2005	0602 Social Affairs - Labour market	NUTS	NUTS 2006	Age Structure Data	Regional data - 699
DEMIFER	DR_YOUNG	Young Age Dependency Ratio	2005	0602 Social Affairs - Labour market	NUTS	NUTS 2006	Age Structure Data	Regional data - 700
DEMIFER	D_80+	Share of population aged 80+ years	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Aqe Structure Data	Regional data - 701

- Regional data (14)

Provider	Indicator - code	Indicator - name	Year(s)	Theme - Sub theme	Geo object	Geo version	Dataset title	Category
DEMIFER	D_65+	Share of population aged 65+ years	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Age Structure Data	Regional data - 702
DEMIFER	D_65+_AVE	Average share of Population Aged 65+ years	Between 2000 and 2007	0201 Demography - Population struct	NUTS	NUTS 2006	Age Structure Data	Regional data - 703
DEMIFER	D_5064	Share of population aged 50-64 years	2005	0201 Demography - Population struct	NUTS	NUTS 2006	Age Structure Data	Regional data - 704
ReRISK - Regi	tmin_ju	Minimum temperature, june	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk maximum temperatures d	Regional data - 705
ReRISK - Regi	tmin_de	Minimum temperature, december	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk maximum temperatures d	Regional data - 706
ReRISK - Regi	tmin_no	Minimum temperature, november	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk maximum temperatures d	Regional data - 707
ReRISK - Regi	tmin_oc	Minimum temperature, october	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk maximum temperatures d	Regional data - 708
ReRISK - Regi	tmin_se	Minimum temperature, september	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk maximum temperatures d	Regional data - 709
ReRISK - Regi	tmin_au	Minimum temperature, august	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk maximum temperatures d	Regional data - 710
ReRISK - Regi	tmin_jl	Minimum temperature, july	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk maximum temperatures d	Regional data - 711
ReRISK - Regi	tmin_may	Minimum temperature, may	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk maximum temperatures d	Regional data - 712
ReRISK - Regi	tmin_ap	Minimum temperature, april	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk maximum temperatures d	Regional data - 713
ReRISK - Regi	tmin_mar	Minimum temperature, march	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk maximum temperatures d	Regional data - 714
ReRISK - Regi	tmin_fe	Minimum temperature, february	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk maximum temperatures d	Regional data - 715
ReRISK - Regi	tmin_ja	Minimum temperature, january	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk maximum temperatures d	Regional data - 716
ReRISK - Regi	tmean_ja	Mean temperature, january	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk median temperatures dat	Regional data - 717
ReRISK - Regi	tmean_de	Mean temperature, december	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk median temperatures dat	Regional data - 718
ReRISK - Regi	tmean_no	Mean temperature, november	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk median temperatures dat	Regional data - 719
ReRISK - Regi	tmean_oc	Mean temperature, october	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk median temperatures dat	Regional data - 720
ReRISK - Regi	tmean_se	Mean temperature, september	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk median temperatures dat	Regional data - 721
ReRISK - Regi	tmean_au	Mean temperature, august	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk median temperatures dat	Regional data - 722
ReRISK - Regi	tmean_jl	Mean temperature, july	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk median temperatures dat	Regional data - 723
ReRISK - Regi	tmean_ju	Mean temperature, june	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk median temperatures dat	Regional data - 724
ReRISK - Regi	tmean_may	Mean temperature, may	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk median temperatures dat	Regional data - 725
ReRISK - Regi	tmean_ap	Mean temperature, april	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk median temperatures dat	Regional data - 726
ReRISK - Regi	tmean_mar	Mean temperature, march	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk median temperatures dat	Regional data - 727
ReRISK - Regi	tmean_fe	Mean temperature, february	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk median temperatures dat	Regional data - 728
ReRISK - Regi	tmax_au	Maximum temperature, august	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk minimum temperatures d	Regional data - 729
ReRISK - Regi	tmax_ja	Maximum temperature, january	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk minimum temperatures d	Regional data - 730
ReRISK - Regi	tmax_fe	Maximum temperature, february	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk minimum temperatures d	Regional data - 731
ReRISK - Regi	tmax_mar	Maximum temperature, march	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk minimum temperatures d	Regional data - 732
ReRISK - Regi	tmax_ap	Maximum temperature, april	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk minimum temperatures d	Regional data - 733
ReRISK - Regi	tmax_may	Maximum temperature, may	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk minimum temperatures d	Regional data - 734
ReRISK - Regi	tmax_ju	Maximum temperature, june	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk minimum temperatures d	Regional data - 735
ReRISK - Regi	tmax_jl	Maximum temperature, july	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk minimum temperatures d	Regional data - 736
ReRISK - Regi	tmax_se	Maximum temperature, september	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk minimum temperatures d	Regional data - 737
ReRISK - Regi	tmax_oc	Maximum temperature, october	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk minimum temperatures d	Regional data - 738
ReRISK - Regi	tmax_no	Maximum temperature, november	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk minimum temperatures d	Regional data - 739
ReRISK - Regi	tmax_de	Maximum temperature, december	Between 1994 and 2008	0402 Energy and Environment - Clima	NUTS	NUTS 2006	ReRisk minimum temperatures d	Regional data - 740
ReRISK - Regi	%_air_pass_trasp	% of national air passenger transport	2005	0302 Transport - Flows	NUTS	NUTS 2006	ReRisk_air_passenger_data	Regional data - 741
ReRISK - Regi	Air travel	passengers embarked and disembarked /total regional popul	2005	0302 Transport - Flows	NUTS	NUTS 2006	ReRisk_air_passenger_data	Regional data - 742
ReRISK - Regi	pass_thsnd	Air passengers	2005	0302 Transport - Flows	NUTS	NUTS 2006	ReRisk_air_passenger_data	Regional data - 743
ReRISK - Regi	age_0-14_t	0-14years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 744
ReRISK - Regi	age_15-64_t	15-64years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 745
ReRISK - Regi	age_65+_t	65+ years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 746
ReRISK - Regi	age_35-39_t	35-39years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 747
ReRISK - Regi	age_40-44_t	40-44years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 748
ReRISK - Regi	age_45-49_t	45-49years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 749
ReRISK - Regi	age_50-54_t	50-54years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 750
ReRISK - Regi	age_55-59_t	55-59years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 751
ReRISK - Regi	age_65-69_t	65-69years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 752
ReRISK - Regi	age_70-74_t	70-74years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 753
ReRISK - Regi	age_75-79_t	75-79years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 754
ReRISK - Regi	age_80-84_t	80-84years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 755
ReRISK - Regi	area	Total land area	Between 1990 and 2007	9999 Cross-thematic and non-themat	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 755

- Regional data (15)

Provider	Indicator - code	Indicator - name	Year(s)	Theme - Sub theme	Geo object	Geo version	Dataset title	Category
ReRISK - Regi	age_60-64_t	60-64years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 756
ReRISK - Regi	age_65-69_t	65-69years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 757
ReRISK - Regi	age_70-74_t	70-74years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 758
ReRISK - Regi	age_75-79_t	75-79years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 759
ReRISK - Regi	age_80-84_t	80-84years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 760
ReRISK - Regi	age_90-94_t	90-94years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 761
ReRISK - Regi	age_dep_rt	Age dependency ratio	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 762
ReRISK - Regi	age_0-4_t	0-4years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 763
ReRISK - Regi	age_5-9_t	5-9years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 764
ReRISK - Regi	age_10-14_t	10-14years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 765
ReRISK - Regi	age_15-19_t	15-19years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 766
ReRISK - Regi	age_20-24_t	20-24years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 767
ReRISK - Regi	age_25-29_t	25-29years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 768
ReRISK - Regi	age_30-34_t	30-34years, total	Between 2005 and 2006	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 769
ReRISK - Regi	pop_t	Total population	Between 2005 and 2008	0201 Demography - Population struct	NUTS	NUTS 2006	ReRisk_area_pop_data	Regional data - 770
ReRISK - Regi	work_oth	Working in another region	2005	0204 Demography - Migrations	NUTS	NUTS 2006	ReRisk_ecosoc_data	Regional data - 771
ReRISK - Regi	work_in	Working in the same region	2005	0702 Economy - Employment	NUTS	NUTS 2006	ReRisk_ecosoc_data	Regional data - 772
ReRISK - Regi	work_oth_in_rt	Working in another region/working in the same region	2005	0702 Economy - Employment	NUTS	NUTS 2006	ReRisk_ecosoc_data	Regional data - 773
ReRISK - Regi	long_unemp_rt	Long term unemployment rate	2007	0702 Economy - Employment	NUTS	NUTS 2006	ReRisk_ecosoc_data	Regional data - 774
ReRISK - Regi	inc_PPCS_rt	Disposable income	Between 2004 and 2005	0603 Social Affairs - Living condition	NUTS	NUTS 2006	ReRisk_ecosoc_data	Regional data - 775
ReRISK - Regi	act_15+_rt	Economic activity rate, 15 years and over	2005	0702 Economy - Employment	NUTS	NUTS 2006	ReRisk_ecosoc_data	Regional data - 776
ReRISK - Regi	inc_PPCS_inh_rt	Disposable income per inhabitant	Between 2004 and 2005	0603 Social Affairs - Living condition	NUTS	NUTS 2006	ReRisk_ecosoc_data	Regional data - 777
ReRISK - Regi	pers_employ_I64	Employment in post and telecommunications	2005	0702 Economy - Employment	NUTS	NUTS 2003, NUTS 2006	ReRisk_employment_transport_c	Regional data - 778
ReRISK - Regi	empI_transp	Employment in transport sector (% of total regional employ	2005	0702 Economy - Employment	NUTS	NUTS 2003, NUTS 2006	ReRisk_employment_transport_c	Regional data - 779
ReRISK - Regi	pers_employ_I	Employment in Transport, storage and communication	2005	0702 Economy - Employment	NUTS	NUTS 2003, NUTS 2006	ReRisk_employment_transport_c	Regional data - 780
ReRISK - Regi	pers_employ_I60	Employment in Land transport, transport via pipelines	2005	0702 Economy - Employment	NUTS	NUTS 2003, NUTS 2006	ReRisk_employment_transport_c	Regional data - 781
ReRISK - Regi	pers_employ_I61	Employment i Water transport	2005	0702 Economy - Employment	NUTS	NUTS 2003, NUTS 2006	ReRisk_employment_transport_c	Regional data - 782
ReRISK - Regi	pers_employ_I62	Employment in Air transport	2005	0702 Economy - Employment	NUTS	NUTS 2003, NUTS 2006	ReRisk_employment_transport_c	Regional data - 783
ReRISK - Regi	pers_employ_I63	Employment in Supporting and auxiliary transport activities; a	2005	0702 Economy - Employment	NUTS	NUTS 2003, NUTS 2006	ReRisk_employment_transport_c	Regional data - 784
ReRISK - Regi	fuel_cost_%_GDP	Fuel costs as % of GDP	2005	0401 Energy and Environment - Ener	NUTS	NUTS 2006	ReRisk_fuelcosts_data	Regional data - 785
ReRISK - Regi	PV_pot	PV potential	2005	0401 Energy and Environment - Ener	NUTS	NUTS 2006	ReRisk_PV_output_data	Regional data - 786
ReRISK - Regi	GVA_energ_int_ind	Gross value added in energy intensive industries	2005	0401 Energy and Environment - Ener	NUTS	NUTS 2006	ReRisk_regional_competitiveness	Regional data - 787
ReRISK - Regi	GVA_ind	Gross value added of industry	2005	0701 Economy - Aggregated account	NUTS	NUTS 2006	ReRisk_regional_competitiveness	Regional data - 788
ReRISK - Regi	%_employ_ind_high	% of employment in industries with high energy spending	2005	0401 Energy and Environment - Ener	NUTS	NUTS 2006	ReRisk_regional_competitiveness	Regional data - 789
ReRISK - Regi	%_employ_ind_high	% of employment in industries with high energy purchases (	2005	0401 Energy and Environment - Ener	NUTS	NUTS 2006	ReRisk_regional_competitiveness	Regional data - 790
ReRISK - Regi	pers_employ	Total employment	2005	0702 Economy - Employment	NUTS	NUTS 2006	ReRisk_regional_competitiveness	Regional data - 791
ReRISK - Regi	employ_ind_high_en	Employment in industries with high energy purchases	2005	0702 Economy - Employment	NUTS	NUTS 2006	ReRisk_regional_competitiveness	Regional data - 792
ReRISK - Regi	%_employ_ind_high	% of employment in industries with high energy purchases	2005	0401 Energy and Environment - Ener	NUTS	NUTS 2006	ReRisk_regional_competitiveness	Regional data - 793
ReRISK - Regi	%_GVA_ene_int_ind	% of GVA depending of energy intensive industries	2005	0401 Energy and Environment - Ener	NUTS	NUTS 2006	ReRisk_regional_competitiveness	Regional data - 794
ReRISK - Regi	tot_GVA	Total Gross value added	2005	0701 Economy - Aggregated account	NUTS	NUTS 2006	ReRisk_regional_competitiveness	Regional data - 795
ReRISK - Regi	win_ener_pot	Wind energy potential	2005	0401 Energy and Environment - Ener	NUTS	NUTS 2006	ReRisk_wind_eneq_potential_dc	Regional data - 796
Telecommunic	Intrnt_infra_evnl	Internet infrastructure evolution	Between 2006 and 2008	0704 Economy - Research and innov	NUTS	NUTS 2006	internet_infra_data	Regional data - 797
Telecommunic	Intrnt_infra	Internet infrastructure	Between 2006 and 2008	0704 Economy - Research and innov	NUTS	NUTS 2006	internet_infra_data	Regional data - 798
Telecommunic	Intrnt_usage	Internet usage	Between 2006 and 2008	0704 Economy - Research and innov	NUTS	NUTS 2003, NUTS 2006	internet_usage_data	Regional data - 799
Telecommunic	Intrnt_usage_evnl	Internet usage evolution	Between 2006 and 2008	0704 Economy - Research and innov	NUTS	NUTS 2006	internet_usage_data	Regional data - 800
TIPTAP: Territ	TIM_Q1	Environmental quality, Territorial Impact	Between 2000 and 2004	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Common Agricultural Policies im	Regional data - 801
TIPTAP: Territ	TIM_Q2	Community viability, Territorial Impact	Between 2000 and 2004	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Common Agricultural Policies im	Regional data - 802
TIPTAP: Territ	TIM_Q3	CO2 emissions, Territorial Impact	Between 2000 and 2004	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Common Agricultural Policies im	Regional data - 803
TIPTAP: Territ	TIM_Q4	Risk of soil erosion, Territorial Impact	Between 2000 and 2004	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Common Agricultural Policies im	Regional data - 804
TIPTAP: Territ	TIM_I1	Landscape diversity, Territorial impact	Between 2000 and 2004	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Common Agricultural Policies im	Regional data - 805
TIPTAP: Territ	TIM_I2	Community identity, Territorial impact	Between 2000 and 2004	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Common Agricultural Policies im	Regional data - 806
TIPTAP: Territ	TIM_I3	Heritage products, Territorial impact	Between 2000 and 2004	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Common Agricultural Policies im	Regional data - 807
TIPTAP: Territ	TIM_TE	Territorial efficiency, Territorial impact	Between 2000 and 2004	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Common Agricultural Policies im	Regional data - 808
TIPTAP: Territ	TIM_TQ	Territorial quality, Territorial impact	Between 2000 and 2004	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Common Agricultural Policies im	Regional data - 809



- Regional data (17)

Provider	Indicator - code	Indicator - name	Year(s)	Theme - Sub theme	Geo object	Geo version	Dataset title	Category
TIPTAP: Territc	TIM_E3	Tourism diversification, Territorial Impact	Between 2000 and 2004	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Common Agricultural Policies imp	Regional data - 864
TIPTAP: Territc	TIM_SI_SUM	Summative impact, Territorial impact	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 865
TIPTAP: Territc	S_I3_Road_N2_n	Regional integration, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 866
TIPTAP: Territc	S_I2_Ext_no_3h	Exposure to external visitors, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 867
TIPTAP: Territc	S_I1_HCI_dens	Landscape fragmentation Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 868
TIPTAP: Territc	S_Q4_GDP_3h	Market opportunities, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 869
TIPTAP: Territc	S_Q3_Traffic_seg	Safety, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 870
TIPTAP: Territc	S_Q2_CO2/km2	Emissions, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 871
TIPTAP: Territc	S_Q1_Fre_PT	Traffic passing through, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 872
TIPTAP: Territc	S_E4_Cong_Level	Congestion, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 873
TIPTAP: Territc	S_E3_GDP_CAP	Economic growth, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 874
TIPTAP: Territc	S_E2_Prod_air	Productivity of airports, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 875
TIPTAP: Territc	S_E1_PROD	Productivity of inland infrastructure, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 876
TIPTAP: Territc	V_I3_Road_N2_n	Regional integration, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 877
TIPTAP: Territc	V_I2_Ext_no_3h	Exposure to external visitors, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 878
TIPTAP: Territc	V_I1_HCI_dens	Landscape fragmentation Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 879
TIPTAP: Territc	V_Q4_GDP_3h	Market opportunities, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 880
TIPTAP: Territc	V_Q3_Traffic_seg	Safety, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 881
TIPTAP: Territc	V_Q2_CO2/km2	Emission, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 882
TIPTAP: Territc	V_Q1_Fre_PT	Traffic passing through, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 883
TIPTAP: Territc	V_E4_Cong_level	Congestion, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 884
TIPTAP: Territc	V_E3_GDP_CAP	Economic growth, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 885
TIPTAP: Territc	V_E2_Prod_air	Productivity of airports, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 886
TIPTAP: Territc	V_E1_PROD	Productivity of inland infrastructure, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 887
TIPTAP: Territc	D_I3_Road_N2_n	Regional integration, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 888
TIPTAP: Territc	D_I2_Ext_no_3h	Exposure to external visitors, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 889
TIPTAP: Territc	D_I1_HCI_dens	Landscape fragmentation, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 890
TIPTAP: Territc	D_Q4_GDP_3h	Market opportunities, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 891
TIPTAP: Territc	D_Q3_Traffic_seg	Congestion, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 892
TIPTAP: Territc	D_Q2_CO2/km2	Emissions, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 893
TIPTAP: Territc	D_Q1_Fre_PT	Traffic passing through, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 894
TIPTAP: Territc	D_E4_Cong_Level	Congestion, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 895
TIPTAP: Territc	TIM_E3_GDP_CAP	Economic growth	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 896
TIPTAP: Territc	TIM_E4_Cong_Level	Congestion costs	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 897
TIPTAP: Territc	TIM_Q1_Fre_PT	Traffic freight passing through	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 898
TIPTAP: Territc	TIM_Q2_CO2/km2	CO2 Emissions by road traffic	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 899
TIPTAP: Territc	TIM_Q3_Traffic_seg	Safety	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 900
TIPTAP: Territc	TIM_Q4_GDP_3h	Market opportunities	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 901
TIPTAP: Territc	TIM_I1_HCI_dens	Landscape fragmentation	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 902
TIPTAP: Territc	D_E3_GDP_CAP	Economic growth, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 903
TIPTAP: Territc	D_E2_Prod_air	Productivity of airports Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 904
TIPTAP: Territc	D_E1_PROD	Productivity of inland infrastructure, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 905
TIPTAP: Territc	TIM_TL_IDENTITY	Territorial identity, Territorial impact	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 906
TIPTAP: Territc	TIM_TQ_QUALITY	Territorial quality, Territorial impact	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 907
TIPTAP: Territc	TIM_TE_EFFICIENCY	Territorial efficiency, Territorial impact	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 908
TIPTAP: Territc	TIM_I3_Road_N2_n	Regional integration	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 909
TIPTAP: Territc	TIM_I2_Ext_no_3h	Exposure to external visitors	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 910
TIPTAP: Territc	PIM_E1_PROD	Productivity of inland infrastructure	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 911
TIPTAP: Territc	PIM_E2_Prod_air	Productivity of airports (in relation to external trips)	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 912
TIPTAP: Territc	PIM_E3_GDP_CAP	Economic growth	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 913
TIPTAP: Territc	PIM_E4_Cong_Level	Congestion costs	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 914
TIPTAP: Territc	PIM_Q1_Fre_PT	Traffic freight passing through	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 915
TIPTAP: Territc	PIM_Q2_CO2/km2	CO2 Emissions by road traffic	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 916
TIPTAP: Territc	PIM_Q3_Traffic_seg	Safety	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 917

- Regional data (18)

Provider	Indicator - code	Indicator - name	Year(s)	Theme - Sub theme	Geo object	Geo version	Dataset title	Category
TIPTAP: Territ	PIM_Q4_GDP_3h	Market opportunities	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 918
TIPTAP: Territ	PIM_I1_HCI_dens	Landscape fragmentation	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 919
TIPTAP: Territ	PIM_I2_Ext_no_3h	Exposure to external visitors	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 920
TIPTAP: Territ	PIM_I3_Road_N2_n	Regional integration	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 921
TIPTAP: Territ	TIM_E1_PROD	Productivity of inland infrastructure	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 922
TIPTAP: Territ	TIM_E2_Prod_air	Productivity of airports (in relation to external trips)	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 923
TIPTAP: Territ	D_Q1_Fre_PT	Traffic passing through, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 924
TIPTAP: Territ	D_E4_Cong_level	Congestion, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 925
TIPTAP: Territ	D_E3_GDP_CAP	Economic growth, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 926
TIPTAP: Territ	D_E2_Prod_air	Productivity of airports Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 927
TIPTAP: Territ	D_E1_PROD	Productivity of inland infrastructure, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 928
TIPTAP: Territ	TIM_SI_SUM	Summative impact, Territorial impact	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 929
TIPTAP: Territ	TIM_TI_IDENTITY	Territorial identity, Territorial impact	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 930
TIPTAP: Territ	TIM_TQ_QUALITY	Territorial quality, Territorial impact	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 931
TIPTAP: Territ	TIM_TE_EFFICIENCY	Territorial efficiency, Territorial impact	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 932
TIPTAP: Territ	TIM_I3_Road_N2_n	Regional integration	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 933
TIPTAP: Territ	TIM_I2_Ext_no_3h	Exposure to external visitors	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 934
TIPTAP: Territ	TIM_I1_HCI_dens	Landscape fragmentation	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 935
TIPTAP: Territ	PIM_Q3_Traffic_seg	Safety	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 936
TIPTAP: Territ	PIM_Q2_CO2/km2	CO2 Emissions by road traffic	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 937
TIPTAP: Territ	PIM_Q1_Fre_PT	Traffic freight passing through	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 938
TIPTAP: Territ	PIM_E4_Cong_level	Congestion costs	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 939
TIPTAP: Territ	PIM_I2_Ext_no_3h	Exposure to external visitors	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 940
TIPTAP: Territ	PIM_I1_HCI_dens	Landscape fragmentation	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 941
TIPTAP: Territ	PIM_Q4_GDP_3h	Market opportunities	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 942
TIPTAP: Territ	TIM_Q4_GDP_3h	Market opportunities	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 943
TIPTAP: Territ	TIM_Q3_Traffic_seg	Safety	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 944
TIPTAP: Territ	TIM_Q2_CO2/km2	CO2 Emissions by road traffic	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 945
TIPTAP: Territ	TIM_Q1_Fre_PT	Traffic freight passing through	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 946
TIPTAP: Territ	TIM_E4_Cong_level	Congestion costs	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 947
TIPTAP: Territ	TIM_E3_GDP_CAP	Economic growth	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 948
TIPTAP: Territ	TIM_E2_Prod_air	Productivity of airports (in relation to external trips)	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 949
TIPTAP: Territ	TIM_E1_PROD	Productivity of inland infrastructure	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 950
TIPTAP: Territ	PIM_I3_Road_N2_n	Regional integration	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 951
TIPTAP: Territ	PIM_E1_PROD	Productivity of inland infrastructure	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 952
TIPTAP: Territ	PIM_E2_Prod_air	Productivity of airports (in relation to external trips)	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 953
TIPTAP: Territ	PIM_E3_GDP_CAP	Economic growth	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 954
TIPTAP: Territ	S_I3_Road_N2_n	Regional integration, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 955
TIPTAP: Territ	S_I2_Ext_no_3h	Exposure to external visitors, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 956
TIPTAP: Territ	S_I1_HCI_dens	Landscape fragmentation Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 957
TIPTAP: Territ	S_Q4_GDP_3h	Market opportunities, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 958
TIPTAP: Territ	S_Q3_Traffic_seg	Safety, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 959
TIPTAP: Territ	S_Q2_CO2/km2	Emissions, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 960
TIPTAP: Territ	S_Q1_Fre_PT	Traffic passing through, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 961
TIPTAP: Territ	S_E4_Cong_level	Congestion, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 962
TIPTAP: Territ	S_E3_GDP_CAP	Economic growth, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 963
TIPTAP: Territ	S_E2_Prod_air	Productivity of airports, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 964
TIPTAP: Territ	S_E1_PROD	Productivity of inland infrastructure, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 965
TIPTAP: Territ	V_I3_Road_N2_n	Regional integration, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 966
TIPTAP: Territ	V_I2_Ext_no_3h	Exposure to external visitors, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 967
TIPTAP: Territ	V_I1_HCI_dens	Landscape fragmentation Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 968
TIPTAP: Territ	V_Q4_GDP_3h	Market opportunities, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 969
TIPTAP: Territ	V_Q3_Traffic_seg	Safety, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 970
TIPTAP: Territ	V_Q2_CO2/km2	Emission, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 971



- Regional data (19)

Provider	Indicator - code	Indicator - name	Year(s)	Theme - Sub theme	Geo object	Geo version	Dataset title	Category
TIPTAP: Territc	V_Q1_Fre_PT	Traffic passing through, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 972
TIPTAP: Territc	V_E4_Cong_level	Congestion, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 973
TIPTAP: Territc	V_E3_GDP_CAP	Economic growth, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 974
TIPTAP: Territc	V_E2_Prod_air	Productivity of airports, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 975
TIPTAP: Territc	V_E1_PROD	Productivity of inland infrastructure, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 976
TIPTAP: Territc	D_I3_Road_N2_n	Regional integration, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 977
TIPTAP: Territc	D_I2_Ext_no_3h	Exposure to external visitors, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 978
TIPTAP: Territc	D_I1_HCI_dens	Landscape fragmentation, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 979
TIPTAP: Territc	D_Q4_GDP_3h	Market opportunities, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 980
TIPTAP: Territc	D_Q3_Traffic_seg	Congestion, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 981
TIPTAP: Territc	D_Q2_CO2/km2	Emissions, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2003, NUTS 2006	Policy impact assessment, scen	Regional data - 982
TIPTAP: Territc	D_Q2_CO2/km2	Emissions, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 983
TIPTAP: Territc	PIM_E1_PROD	Productivity of inland infrastructure	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 984
TIPTAP: Territc	PIM_E2_Prod_air	Productivity of airports (in relation to external trips)	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 985
TIPTAP: Territc	PIM_E3_GDP_CAP	Economic growth	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 986
TIPTAP: Territc	PIM_E4_Cong_level	Congestion costs	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 987
TIPTAP: Territc	PIM_Q1_Fre_PT	Traffic freight passing through	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 988
TIPTAP: Territc	PIM_Q2_CO2/km2	CO2 Emissions by road traffic	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 989
TIPTAP: Territc	PIM_Q3_Traffic_seg	Safety	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 990
TIPTAP: Territc	PIM_Q4_GDP_3h	Market opportunities	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 991
TIPTAP: Territc	PIM_I1_HCI_dens	Landscape fragmentation	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 992
TIPTAP: Territc	PIM_I2_Ext_no_3h	Exposure to external visitors	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 993
TIPTAP: Territc	PIM_I3_Road_N2_n	Regional integration	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 994
TIPTAP: Territc	TIM_E1_PROD	Productivity of inland infrastructure	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 995
TIPTAP: Territc	TIM_E2_Prod_air	Productivity of airports (in relation to external trips)	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 996
TIPTAP: Territc	TIM_E3_GDP_CAP	Economic growth	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 997
TIPTAP: Territc	TIM_E4_Cong_level	Congestion costs	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 998
TIPTAP: Territc	TIM_Q1_Fre_PT	Traffic freight passing through	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 999
TIPTAP: Territc	TIM_Q2_CO2/km2	CO2 Emissions by road traffic	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1000
TIPTAP: Territc	TIM_Q3_Traffic_seg	Safety	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1001
TIPTAP: Territc	TIM_Q4_GDP_3h	Market opportunities	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1002
TIPTAP: Territc	TIM_I1_HCI_dens	Landscape fragmentation	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1003
TIPTAP: Territc	TIM_I2_Ext_no_3h	Exposure to external visitors	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1004
TIPTAP: Territc	TIM_I3_Road_N2_n	Regional integration	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1005
TIPTAP: Territc	TIM_TE_EFFICIENCY	Territorial efficiency, Territorial impact	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1006
TIPTAP: Territc	TIM_TQ_QUALITY	Territorial quality, Territorial impact	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1007
TIPTAP: Territc	TIM_TI_IDENTITY	Territorial identity, Territorial impact	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1008
TIPTAP: Territc	TIM_SI_SUM	Summative impact, Territorial impact	2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1009
TIPTAP: Territc	D_E1_PROD	Productivity of inland infrastructure, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1010
TIPTAP: Territc	D_E2_Prod_air	Productivity of airports Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1011
TIPTAP: Territc	D_E3_GDP_CAP	Economic growth, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1012
TIPTAP: Territc	D_E4_Cong_level	Congestion, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1013
TIPTAP: Territc	D_Q1_Fre_PT	Traffic passing through, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1014
TIPTAP: Territc	D_Q3_Traffic_seg	Congestion, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1015
TIPTAP: Territc	D_Q4_GDP_3h	Market opportunities, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1016
TIPTAP: Territc	D_I1_HCI_dens	Landscape fragmentation, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1017
TIPTAP: Territc	D_I2_Ext_no_3h	Exposure to external visitors, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1018
TIPTAP: Territc	D_I3_Road_N2_n	Regional integration, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1019
TIPTAP: Territc	V_E1_PROD	Productivity of inland infrastructure, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1020
TIPTAP: Territc	V_E2_Prod_air	Productivity of airports, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1021
TIPTAP: Territc	V_E3_GDP_CAP	Economic growth, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1022
TIPTAP: Territc	V_E4_Cong_level	Congestion, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1023
TIPTAP: Territc	V_Q1_Fre_PT	Traffic passing through, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1024
TIPTAP: Territc	V_Q2_CO2/km2	Emission, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1025

## • Regional data (20)

Provider	Indicator - code	Indicator - name	Year(s)	Theme - Sub theme	Geo object	Geo version	Dataset title	Category
TIPTAP: Territ	V_Q3_Traffic_seg	Safety, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1026
TIPTAP: Territ	V_Q4_GDP_3h	Market opportunities, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1027
TIPTAP: Territ	V_I1_HCI_dens	Landscape fragmentation Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1028
TIPTAP: Territ	V_I2_Ext_no_3h	Exposure to external visitors, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1029
TIPTAP: Territ	V_I3_Road_N2_n	Regional integration, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1030
TIPTAP: Territ	S_E1_PROD	Productivity of inland infrastructure, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1031
TIPTAP: Territ	S_E2_Prod_air	Productivity of airports, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1032
TIPTAP: Territ	S_E3_GDP_CAP	Economic growth, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1033
TIPTAP: Territ	S_E4_Cong_level	Congestion, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1034
TIPTAP: Territ	S_Q1_Fre_PT	Traffic passing through, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1035
TIPTAP: Territ	S_Q2_CO2/km2	Emissions, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1036
TIPTAP: Territ	S_Q3_Traffic_seg	Safety, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1037
TIPTAP: Territ	S_Q4_GDP_3h	Market opportunities, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1038
TIPTAP: Territ	S_I1_HCI_dens	Landscape fragmentation Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1039
TIPTAP: Territ	S_I2_Ext_no_3h	Exposure to external visitors, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1040
TIPTAP: Territ	S_I3_Road_N2_n	Regional integration, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Policy impact assessment, scen	Regional data - 1041
TIPTAP: Territ	D_E3_GDP_CAP	Economic growth, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1042
TIPTAP: Territ	D_Q1_Fre_PT	Traffic passing through, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1043
TIPTAP: Territ	D_Q2_CO2/km2	Emissions, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1044
TIPTAP: Territ	S_Q4_GDP_3h	Market opportunities, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1045
TIPTAP: Territ	D_Q3_Traffic_seg	Congestion, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1046
TIPTAP: Territ	D_Q4_GDP_3h	Market opportunities, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1047
TIPTAP: Territ	D_I1_HCI_dens	Landscape fragmentation, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1048
TIPTAP: Territ	D_I2_Ext_no_3h	Exposure to external visitors, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1049
TIPTAP: Territ	D_I3_Road_N2_n	Regional integration, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1050
TIPTAP: Territ	V_E1_PROD	Productivity of inland infrastructure, Vunlerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1051
TIPTAP: Territ	V_E2_Prod_air	Productivity of airports, Vunlerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1052
TIPTAP: Territ	V_E3_GDP_CAP	Economic growth, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1053
TIPTAP: Territ	V_E4_Cong_level	Congestion, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1054
TIPTAP: Territ	V_Q1_Fre_PT	Traffic passing through, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1055
TIPTAP: Territ	V_Q2_CO2/km2	Emission, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1056
TIPTAP: Territ	V_Q3_Traffic_seg	Safety, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1057
TIPTAP: Territ	V_Q4_GDP_3h	Market opportunities, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1058
TIPTAP: Territ	V_I1_HCI_dens	Landscape fragmentation Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1059
TIPTAP: Territ	V_I2_Ext_no_3h	Exposure to external visitors, Vulnerability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1060
TIPTAP: Territ	V_I3_Road_N2_n	Regional integration, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1061
TIPTAP: Territ	S_E1_PROD	Productivity of inland infrastructure, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1062
TIPTAP: Territ	S_E2_Prod_air	Productivity of airports, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1063
TIPTAP: Territ	S_E3_GDP_CAP	Economic growth, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1064
TIPTAP: Territ	S_Q3_Traffic_seg	Safety, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1065
TIPTAP: Territ	S_Q2_CO2/km2	Emissions, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1066
TIPTAP: Territ	S_Q1_Fre_PT	Traffic passing through, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1067
TIPTAP: Territ	S_E4_Cong_level	Congestion, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1068
TIPTAP: Territ	TIM_I2_Ext_no_3h	Exposure to external visitors	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1069
TIPTAP: Territ	TIM_I3_Road_N2_n	Regional integration	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1070
TIPTAP: Territ	TIM_TE_EFFICIENCY	Territorial efficiency, Territorial impact	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1071
TIPTAP: Territ	TIM_TQ_QUALITY	Territorial quality, Territorial impact	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1072
TIPTAP: Territ	TIM_TI_IDENTITY	Territorial identity, Territorial impact	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1073
TIPTAP: Territ	TIM_SI_SUM	Summative impact, Territorial impact	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1074
TIPTAP: Territ	D_E1_PROD	Productivity of inland infrastructure, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1075
TIPTAP: Territ	D_E2_Prod_air	Productivity of airports, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1076
TIPTAP: Territ	D_E4_Cong_level	Congestion, Desiderability	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1077
TIPTAP: Territ	PIM_E1_PROD	Productivity of inland infrastructure	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1078
TIPTAP: Territ	S_I3_Road_N2_n	Regional integration, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1079

- Regional data (21)

Provider	Indicator - code	Indicator - name	Year(s)	Theme - Sub theme	Geo object	Geo version	Dataset title	Category
TIPTAP: Territ	S_I2_Ext_no_3h	Exposure to external visitors, Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1080
TIPTAP: Territ	S_I1_HCI_dens	Landscape fragmentation Sensitivity	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1081
TIPTAP: Territ	PIM_E2_Prod_air	Productivity of airports (in relation to external trips)	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1082
TIPTAP: Territ	PIM_E3_GDP_CAP	Economic growth	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1083
TIPTAP: Territ	PIM_E4_Cong_level	Congestion costs	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1084
TIPTAP: Territ	PIM_Q1_Fre_PT	Traffic freight passing through	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1085
TIPTAP: Territ	PIM_Q2_CO2/km2	CO2 Emissions by road traffic	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1086
TIPTAP: Territ	PIM_Q3_Traffic_sec	Safety	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1087
TIPTAP: Territ	PIM_Q4_GDP_3h	Market opportunities	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1088
TIPTAP: Territ	PIM_I1_HCI_dens	Landscape fragmentation	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1089
TIPTAP: Territ	PIM_I2_Ext_no_3h	Exposure to external visitors	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1090
TIPTAP: Territ	PIM_I3_Road_N2_n	Regional integration	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1091
TIPTAP: Territ	TIM_E1_PROD	Productivity of inland infrastructure	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1092
TIPTAP: Territ	TIM_E2_Prod_air	Productivity of airports (in relation to external trips)	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1093
TIPTAP: Territ	TIM_E3_GDP_CAP	Economic growth	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1094
TIPTAP: Territ	TIM_E4_Cong_level	Congestion costs	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1095
TIPTAP: Territ	TIM_Q1_Fre_PT	Traffic freight passing through	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1096
TIPTAP: Territ	TIM_Q2_CO2/km2	CO2 Emissions by road traffic	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1097
TIPTAP: Territ	TIM_Q3_Traffic_sec	Safety	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1098
TIPTAP: Territ	TIM_Q4_GDP_3h	Market opportunities	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1099
TIPTAP: Territ	TIM_I1_HCI_dens	Landscape fragmentation	Between 2005 and 2030	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	Transport policy impact assessr	Regional data - 1100
Typology Com	typ_urbur	Typology on urban-rural regions	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	ESPON CU Typology Compilation	Regional data - 1101
Typology Com	typ_metro	Typology on metropolitan regions	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	ESPON CU Typology Compilation	Regional data - 1102
Typology Com	typ_border	Border regions	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	ESPON CU Typology Compilation	Regional data - 1103
Typology Com	typ_island	Typology on island regions	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	ESPON CU Typology Compilation	Regional data - 1104
Typology Com	typ_sparsely	Sparsely populated regions	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	ESPON CU Typology Compilation	Regional data - 1105
Typology Com	typ_outermost	Outermost regions	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	ESPON CU Typology Compilation	Regional data - 1106
Typology Com	typ_mountains	Typology on mountainous regions	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	ESPON CU Typology Compilation	Regional data - 1107
Typology Com	typ_coastal	Typology on coastal regions	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	ESPON CU Typology Compilation	Regional data - 1108
Typology Com	typ_indtrans	Typology on regions in industrial transition	2009	9901 Cross-thematic and non-themat	NUTS	NUTS 2006	ESPON CU Typology Compilation	Regional data - 1109
Typology Com	pop_t	Total population	2003	0201 Demography - Population struct	NUTS	NUTS 2003	ESPON_2006_00_area_nuts_2-	Regional data - 1110
Typology Com	RDBES	FuE Business Enterprise Sector, personnell	2003	0704 Economy - Research and innov	NUTS	NUTS 2003	ESPON_2006_00_area_nuts_2-	Regional data - 1111
Typology Com	Rdex	Total intramural R&D expenditure	2002	0704 Economy - Research and innov	NUTS	NUTS 2003	ESPON_2006_00_area_nuts_2-	Regional data - 1112
Typology Com	Patpe	Patent applications to the EPO per persons employed	2002	0704 Economy - Research and innov	NUTS	NUTS 2003	ESPON_2006_00_area_nuts_2-	Regional data - 1113
Typology Com	DevGDPeh	Development of GDP per capita (Euros)	Between 1998 and 2002	0701 Economy - Aggregated account	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_00_area_nuts_2-	Regional data - 1114
Typology Com	DevGDPph	Development of GDP in PPS per capita	Between 1998 and 2002	0701 Economy - Aggregated account	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_00_area_nuts_2-	Regional data - 1115
Typology Com	GDPe	GDP per capita (Euros)	2002	0701 Economy - Aggregated account	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_00_area_nuts_2-	Regional data - 1116
Typology Com	GDPh	GDP per capita (PPS)	2002	0701 Economy - Aggregated account	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_00_area_nuts_2-	Regional data - 1117
Typology Com	DevURI	Development of unemployment rate, young population	Between 1998 and 2001	0702 Economy - Employment	NUTS	NUTS 1999	ESPON_2006_00_area_nuts_2-	Regional data - 1118
Typology Com	DevURm	Development of unemployment rate, male	Between 1998 and 2004	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_00_area_nuts_2-	Regional data - 1119
Typology Com	DevURf	Development of unemployment rate, female	Between 1998 and 2004	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_00_area_nuts_2-	Regional data - 1120
Typology Com	DevUR	Development of unemployment rate	Between 1998 and 2004	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_00_area_nuts_2-	Regional data - 1121
Typology Com	UrI	Young unemployment rate	Between 2001 and 2004	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_00_area_nuts_2-	Regional data - 1122
Typology Com	Urm	Unemployment rate male	Between 2001 and 2004	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_00_area_nuts_2-	Regional data - 1123
Typology Com	Urf	Unemployment rate female	Between 2001 and 2004	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_00_area_nuts_2-	Regional data - 1124
Typology Com	UR	Unemployment rate	Between 2001 and 2004	0702 Economy - Employment	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_00_area_nuts_2-	Regional data - 1125
Typology Com	EmN65	Share of employed persons, national, > 65 years	2002	0702 Economy - Employment	NUTS	NUTS 1999	ESPON_2006_00_area_nuts_2-	Regional data - 1126
Typology Com	EmN25	Share of employed persons, national, < 25 years	2002	0702 Economy - Employment	NUTS	NUTS 1999	ESPON_2006_00_area_nuts_2-	Regional data - 1127
Typology Com	EmN	Persons employed, national	2002	0702 Economy - Employment	NUTS	NUTS 1999	ESPON_2006_00_area_nuts_2-	Regional data - 1128
Typology Com	PESer	Share of persons employed in services	2001	0702 Economy - Employment	NUTS	NUTS 1999	ESPON_2006_00_area_nuts_2-	Regional data - 1129
Typology Com	PEIn	Share of persons employed in industry	2001	0702 Economy - Employment	NUTS	NUTS 1999	ESPON_2006_00_area_nuts_2-	Regional data - 1130
Typology Com	PEAg	Share of persons employed in agriculture	2001	0702 Economy - Employment	NUTS	NUTS 1999	ESPON_2006_00_area_nuts_2-	Regional data - 1131
Typology Com	Pesf	Share of persons employed, females	2001	0702 Economy - Employment	NUTS	NUTS 1999	ESPON_2006_00_area_nuts_2-	Regional data - 1132
Typology Com	Popf14	Share of female population < 14 years	2003	0201 Demography - Population struct	NUTS	NUTS 2003	ESPON_2006_00_area_nuts_2-	Regional data - 1133

- Regional data (22)

Provider	Indicator - code	Indicator - name	Year(s)	Theme - Sub theme	Geo object	Geo version	Dataset title	Category
Typology Com Pop75		Share of high aged population (> 75 years)	2003	0201 Demography - Population struct	NUTS	NUTS 2003	ESPON_2006_00_area_nuts_2-1	Regional data - 1134
Typology Com Pop65		Share of population > 65 years	2003	0201 Demography - Population struct	NUTS	NUTS 2003	ESPON_2006_00_area_nuts_2-1	Regional data - 1135
Typology Com Pop14		Share of population < 14 years	2003	0201 Demography - Population struct	NUTS	NUTS 2003	ESPON_2006_00_area_nuts_2-1	Regional data - 1136
Typology Com Popm		Share of male population	2003	0201 Demography - Population struct	NUTS	NUTS 2003	ESPON_2006_00_area_nuts_2-1	Regional data - 1137
Typology Com Popf		Share of female population	2003	0201 Demography - Population struct	NUTS	NUTS 2003	ESPON_2006_00_area_nuts_2-1	Regional data - 1138
Typology Com Popdens		Population density	2002	9901 Cross-thematic and non-themat	NUTS	NUTS 2003	ESPON_2006_00_area_nuts_2-1	Regional data - 1139
Typology Com Area		Total land area	2003	9999 Cross-thematic and non-themat	NUTS	NUTS 2003	ESPON_2006_00_area_nuts_2-1	Regional data - 1140
Typology Com Pesm		Share of persons employed, males	2001	0702 Economy - Employment	NUTS	NUTS 1999	ESPON_2006_00_area_nuts_2-1	Regional data - 1141
Typology Com Em		Persons employed	Between 2001 and 2002	0702 Economy - Employment	NUTS	NUTS 1999	ESPON_2006_00_area_nuts_2-1	Regional data - 1142
Typology Com ACPs25		Share of active population < 25 years	2001	0602 Social Affairs - Labour market	NUTS	NUTS 1999	ESPON_2006_00_area_nuts_2-1	Regional data - 1143
Typology Com ACPT		Active population	2001	0702 Economy - Employment	NUTS	NUTS 1999	ESPON_2006_00_area_nuts_2-1	Regional data - 1144
Typology Com Dmpop		Development of male population	Between 1995 and 2003	0201 Demography - Population struct	NUTS	NUTS 2003	ESPON_2006_00_area_nuts_2-1	Regional data - 1145
Typology Com Dfpop		Development of female population	Between 1995 and 2003	0201 Demography - Population struct	NUTS	NUTS 2003	ESPON_2006_00_area_nuts_2-1	Regional data - 1146
Typology Com Dpop		Development of total population	Between 1995 and 2003	0201 Demography - Population struct	NUTS	NUTS 2003	ESPON_2006_00_area_nuts_2-1	Regional data - 1147
Typology Com Popm75		Share of high aged males (> 75 years)	2003	0201 Demography - Population struct	NUTS	NUTS 2003	ESPON_2006_00_area_nuts_2-1	Regional data - 1148
Typology Com Popm65		Share of male population > 65 years	2003	0201 Demography - Population struct	NUTS	NUTS 2003	ESPON_2006_00_area_nuts_2-1	Regional data - 1149
Typology Com Popm14		Share of male population < 14 years	2003	0201 Demography - Population struct	NUTS	NUTS 2003	ESPON_2006_00_area_nuts_2-1	Regional data - 1150
Typology Com Popf75		Share of high aged females (> 75 years)	2003	0201 Demography - Population struct	NUTS	NUTS 2003	ESPON_2006_00_area_nuts_2-1	Regional data - 1151
Typology Com Popf65		Share of female population > 65 years	2003	0201 Demography - Population struct	NUTS	NUTS 2003	ESPON_2006_00_area_nuts_2-1	Regional data - 1152
Typology Com RCECLEc		RCE - classified economy	2002	9901 Cross-thematic and non-themat	NUTS	NUTS 2003	ESPON_2006_01_regional_class	Regional data - 1153
Typology Com RCECLLi		RCE - classified lisbon performance	2002	9901 Cross-thematic and non-themat	NUTS	NUTS 2003	ESPON_2006_01_regional_class	Regional data - 1154
Typology Com RCECLLa		RCE - classified labour market	2003	9901 Cross-thematic and non-themat	NUTS	NUTS 2003	ESPON_2006_01_regional_class	Regional data - 1155
Typology Com RCECLDe		RCE - classified demography	2002	9901 Cross-thematic and non-themat	NUTS	NUTS 2003	ESPON_2006_01_regional_class	Regional data - 1156
Typology Com RCECLNa		RCE - classified naturalness	2000	9901 Cross-thematic and non-themat	NUTS	NUTS 2003	ESPON_2006_01_regional_class	Regional data - 1157
Typology Com RCECLNH		RCE - classified natural hazards	2002	9901 Cross-thematic and non-themat	NUTS	NUTS 2003	ESPON_2006_01_regional_class	Regional data - 1158
Typology Com RCECLTH		RCE - classified technological hazards	2002	9901 Cross-thematic and non-themat	NUTS	NUTS 2003	ESPON_2006_01_regional_class	Regional data - 1159
Typology Com RCECLAc		RCE - classified accessibility	2001	9901 Cross-thematic and non-themat	NUTS	NUTS 2003	ESPON_2006_01_regional_class	Regional data - 1160
Typology Com N2aN3d		Share of NUTS 2 area comprising NUTS 3 regions with popu	Between 1995 and 1999	0201 Demography - Population struct	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1161
Typology Com TCMMB		Typology crossing mobility and migratory balances	Between 1996 and 1999	9901 Cross-thematic and non-themat	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1162
Typology Com IMB		Internal migratory balance	Between 1996 and 1999	0204 Demography - Migrations	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1163
Typology Com TMB		Total migratory balance	Between 1996 and 1999	0204 Demography - Migrations	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1164
Typology Com EMB		External migratory balance	Between 1996 and 1999	0204 Demography - Migrations	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1165
Typology Com lmo		Internal mobility by region	Between 1996 and 1999	0204 Demography - Migrations	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1166
Typology Com MoRN		Mobility by region relative to national mobility	Between 1996 and 1999	0204 Demography - Migrations	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1167
Typology Com P65s		Share (%) of population in the ages 65+	Between 1990 and 1999	0201 Demography - Population struct	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1168
Typology Com Pctpd		Total population development	Between 1996 and 1999	0201 Demography - Population struct	NUTS	NUTS 1995, NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1169
Typology Com Pcnpd		Natural population development	Between 1996 and 1999	0202 Demography - Natural changes	NUTS	NUTS 1995, NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1170
Typology Com Pcnm		Net migration	Between 1996 and 1999	0204 Demography - Migrations	NUTS	NUTS 1995, NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1171
Typology Com TMBac		Typology of migratory balances by age classes	Between 1995 and 2000	9901 Cross-thematic and non-themat	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1172
Typology Com DepR		Dependency rate	Between 1995 and 1999	0201 Demography - Population struct	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1173
Typology Com N2PopN3d		Share of NUTS 2 average population living in NUTS 3 regions	Between 1995 and 1999	0201 Demography - Population struct	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1174
Typology Com AM65		Population with 65 and more years (Model A)	Between 2000 and 2050	0201 Demography - Population struct	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1175
Typology Com B0M65		Population with 65 and more years (Model B0)	Between 2000 and 2050	0201 Demography - Population struct	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1176
Typology Com B1M65		Population with 65 and more years (Model B1)	Between 2000 and 2050	0201 Demography - Population struct	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1177
Typology Com B2M65		Population with 65 and more years (Model B2)	Between 2000 and 2050	0201 Demography - Population struct	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1178
Typology Com B3M65		Population with 65 and more years (Model B3)	Between 2000 and 2050	0201 Demography - Population struct	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1179
Typology Com SIAI		Average score on indirect "ageing"/ "depopulating" indicators	2000	9901 Cross-thematic and non-themat	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1180
Typology Com SIAIG		Average score on indirect "ageing"/ "depopulating" indicators	2000	9901 Cross-thematic and non-themat	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1181
Typology Com NTFR		National Total Fertility Rates	Between 1999 and 2000	0202 Demography - Natural changes	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1182
Typology Com APoG		Ageing Population (4 groups)	2000	0201 Demography - Population struct	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1183
Typology Com ALFG		Ageing "Labour Force" (4 groups)	2000	0602 Social Affairs - Labour market	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1184
Typology Com LFRG		"Labour Force" Replacement (4 groups)	2000	0602 Social Affairs - Labour market	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1185
Typology Com PADG		Post-Active Dependency (4 groups)	2000	0602 Social Affairs - Labour market	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1186
Typology Com ApYG		Aged People vs. Youth (4 groups)	2000	0201 Demography - Population struct	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1187

- Regional data (23)

Provider	Indicator - code	Indicator - name	Year(s)	Theme - Sub theme	Geo object	Geo version	Dataset title	Category
Typology Com SchG		Share of children (4 groups)	2000	0201 Demography - Population struct	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1188
Typology Com CNGPG		Changes in Natural Growth Potential (4 groups)	2000	0202 Demography - Natural changes	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1189
Typology Com APol		Ageing Population (indexes)	2000	0201 Demography - Population struct	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1190
Typology Com ALFI		Ageing "Labour Force" (indexes)	2000	0602 Social Affairs - Labour market	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1191
Typology Com PADI		Post-Active Dependency (indexes)	2000	0602 Social Affairs - Labour market	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1192
Typology Com ApYI		Aged People vs. Youth (indexes)	2000	0201 Demography - Population struct	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1193
Typology Com SchI		Share of children (indexes)	2000	0201 Demography - Population struct	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1194
Typology Com CNGPI		Changes in Natural Growth Potential (indexes)	Between 2000 and 2020	0202 Demography - Natural changes	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1195
Typology Com LFRl		"Labour Force" Replacement Index	2000	0602 Social Affairs - Labour market	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1196
Typology Com TypRA		Type of rural area	2000	9901 Cross-thematic and non-themat	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1197
Typology Com RelDepQ		Relative depopulation, quartiles	Between 1990 and 2000	9901 Cross-thematic and non-themat	NUTS	NUTS 1999	ESPON_2006_02_population-daf	Regional data - 1198
Typology Com PopC		Population change	Between 1990 and 2000	0201 Demography - Population struct	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_02_population-daf	Regional data - 1199
Typology Com DGP		GDP per capita in PPS	1999	0701 Economy - Aggregated account	NUTS	NUTS 1999	ESPON_2006_04_wealth_and_r	Regional data - 1200
Typology Com DGPI0		GDP per capita in PPS- Medium deviation	1999	0701 Economy - Aggregated account	NUTS	NUTS 1999	ESPON_2006_04_wealth_and_r	Regional data - 1201
Typology Com DGPIE		GDP per capita in PPS- Global deviation (EU15)	1999	0701 Economy - Aggregated account	NUTS	NUTS 1999	ESPON_2006_04_wealth_and_r	Regional data - 1202
Typology Com DGPIn		GDP per capita in PPS- Local deviation	1999	0701 Economy - Aggregated account	NUTS	NUTS 1999	ESPON_2006_04_wealth_and_r	Regional data - 1203
Typology Com NCA		Number of commercial airports	2001	0303 Transport - Infrastructures	NUTS	NUTS 1999	ESPON_2006_06_transport_nut	Regional data - 1204
Typology Com TMRP2		Accessibility time to market by rail and road	1997	0301 Transport - Accessibility	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_06_transport_nut	Regional data - 1205
Typology Com TMRaP2		Accessibility time to market by rail	1997	0301 Transport - Accessibility	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_06_transport_nut	Regional data - 1206
Typology Com TMRoP2		Accessibility time to market by road	1997	0301 Transport - Accessibility	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_06_transport_nut	Regional data - 1207
Typology Com AcME		Potential accessibility, multimodal	2001	0301 Transport - Accessibility	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_06_transport_nut	Regional data - 1208
Typology Com AcRoE		Potential accessibility by road	2001	0301 Transport - Accessibility	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_06_transport_nut	Regional data - 1209
Typology Com AcRaE		Potential accessibility by rail	2001	0301 Transport - Accessibility	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_06_transport_nut	Regional data - 1210
Typology Com AcAE		Potential accessibility by air	2001	0301 Transport - Accessibility	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_06_transport_nut	Regional data - 1211
Typology Com DMAC		Daily market accessible by car in terms of GDP	2000	0301 Transport - Accessibility	NUTS	NUTS 1999	ESPON_2006_06_transport_nut	Regional data - 1212
Typology Com DPC		Daily population accessible by car	1999	0301 Transport - Accessibility	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_06_transport_nut	Regional data - 1213
Typology Com CM		Time to the nearest motorway access, by car of the capital or	2001	0301 Transport - Accessibility	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_06_transport_nut	Regional data - 1214
Typology Com CCS		Connectivity to commercial seaports by car of the capital or	2001	0301 Transport - Accessibility	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_06_transport_nut	Regional data - 1215
Typology Com CCA		Connectivity to commercial airports by car of the capital or c	2001	0301 Transport - Accessibility	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_06_transport_nut	Regional data - 1216
Typology Com TCAPy		Traffic in commercial airports	2001	0302 Transport - Flows	NUTS	NUTS 1999	ESPON_2006_06_transport_nut	Regional data - 1217
Typology Com LR		Length of railway network	2001	0303 Transport - Infrastructures	NUTS	NUTS 1999	ESPON_2006_06_transport_nut	Regional data - 1218
Typology Com Lro		Length of road network	2001	0303 Transport - Infrastructures	NUTS	NUTS 1999	ESPON_2006_06_transport_nut	Regional data - 1219
Typology Com NCS		Number of commercial seaports	2001	0303 Transport - Infrastructures	NUTS	NUTS 1999	ESPON_2006_06_transport_nut	Regional data - 1220
Typology Com PFw		Share of firms with own website	2002	0704 Economy - Research and innov.	NUTS	NUTS 1999	ESPON_2006_09_communicatio	Regional data - 1221
Typology Com IUI		Share of Internet users	2002	0704 Economy - Research and innov.	NUTS	NUTS 1999	ESPON_2006_09_communicatio	Regional data - 1222
Typology Com PCL2		Share of permanent crops	1996	0101 Agriculture and Fisheries - Farn	NUTS	NUTS 1999	ESPON_2006_11_land_use_nut	Regional data - 1223
Typology Com ArSc		Artificial surfaces per 1000 inhabitants	1996	0502 Land use - urban land use attrit	NUTS	NUTS 1999	ESPON_2006_11_land_use_nut	Regional data - 1224
Typology Com ArSu		Share of artificial surfaces	1996	0502 Land use - urban land use attrit	NUTS	NUTS 1999	ESPON_2006_11_land_use_nut	Regional data - 1225
Typology Com UFL2		Share of urban fabric	1996	9901 Cross-thematic and non-themat	NUTS	NUTS 1999	ESPON_2006_11_land_use_nut	Regional data - 1226
Typology Com ALL2		Share of arable land	1996	0503 Land use - rural land use attrit	NUTS	NUTS 1999	ESPON_2006_11_land_use_nut	Regional data - 1227
Typology Com ArSg		Artificial surfaces per Gross Domestic Product	1996	0502 Land use - urban land use attrit	NUTS	NUTS 1999	ESPON_2006_11_land_use_nut	Regional data - 1228
Typology Com FoFhaz		Forest fire hazard	Between 1997 and 2003	0403 Energy and Environment - Vulne	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_12_environment_l	Regional data - 1229
Typology Com smwh		Sum of all weighted hazard values	2004	0403 Energy and Environment - Vulne	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_12_environment_l	Regional data - 1230
Typology Com ClaOilSum		Classification of Oil-SUM values	2002	0401 Energy and Environment - Energ	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_12_environment_l	Regional data - 1231
Typology Com RadAc		Potential risk of radioactive contamination on NUTS3 regions	2003	0403 Energy and Environment - Vulne	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_12_environment_l	Regional data - 1232
Typology Com CPHP		Chemical plants hazard potential	Between 2001 and 2004	0403 Energy and Environment - Vulne	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_12_environment_l	Regional data - 1233
Typology Com ATHP		Air traffics hazard potential	Between 1996 and 2003	0403 Energy and Environment - Vulne	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_12_environment_l	Regional data - 1234
Typology Com WinTroSto		Approximate probability of having winter storms and for trop	2004	0403 Energy and Environment - Vulne	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_12_environment_l	Regional data - 1235
Typology Com VolErub		Volcanic eruptions during the last 10 000 years	2003	0403 Energy and Environment - Vulne	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_12_environment_l	Regional data - 1236
Typology Com Tsunam		Occurrence of tsunami	2003	0403 Energy and Environment - Vulne	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_12_environment_l	Regional data - 1237
Typology Com StormS		Occurrence of storm surges	2004	0403 Energy and Environment - Vulne	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_12_environment_l	Regional data - 1238
Typology Com Lsloc		Occurrence of landslides	2004	0403 Energy and Environment - Vulne	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_12_environment_l	Regional data - 1239
Typology Com Floo		Regional flood hazard potential	Between 1987 and 2002	0403 Energy and Environment - Vulne	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_12_environment_l	Regional data - 1240
Typology Com HazCl		Extreme temperatures	Between 1961 and 1990	0403 Energy and Environment - Vulne	NUTS	NUTS 1999, NUTS 2003	ESPON_2006_12_environment_l	Regional data - 1241

- Regional data (24)

Provider	Indicator - code	Indicator - name	Year(s)	Theme - Sub theme	Geo object	Geo version	Dataset title	Category
Typology Com HAZC		Regional earthquake hazard potential	1998	0403 Energy and Environment - Vulne	NUTS	1999, NUTS 2003	ESPON_2006_12_environment_1	Regional data - 1242
Typology Com LSD		Large scale droughts in Europe	Between 1904 and 1995	0403 Energy and Environment - Vulne	NUTS	1999, NUTS 2003	ESPON_2006_12_environment_1	Regional data - 1243
Typology Com OSA		Occurrence of snow avalanches	2004	0403 Energy and Environment - Vulne	NUTS	1999, NUTS 2003	ESPON_2006_12_environment_1	Regional data - 1244
Typology Com FH65		Share of farm holders aged <65	1997	0101 Agriculture and Fisheries - Farn	NUTS	1999	ESPON_2006_13_agriculture_nu	Regional data - 1245
Typology Com FALL		Share of UAA that is fallow	2001	0101 Agriculture and Fisheries - Farn	NUTS	1999	ESPON_2006_13_agriculture_nu	Regional data - 1246
Typology Com ARAB		Share of UAA which is arable	2001	0101 Agriculture and Fisheries - Farn	NUTS	1999	ESPON_2006_13_agriculture_nu	Regional data - 1247
Typology Com SFCT		Cohesion Fund expenditure (T)	Between 1994 and 1999	9901 Cross-thematic and non-themat	NUTS	1999	ESPON_2006_18_public_sector	Regional data - 1248
Typology Com SFR		Structural Fund expenditure (R)	Between 1994 and 1999	9901 Cross-thematic and non-themat	NUTS	1999	ESPON_2006_18_public_sector	Regional data - 1249
Typology Com SFT		All Structural and Cohesion Fund expenditure	Between 1994 and 1999	9901 Cross-thematic and non-themat	NUTS	1999	ESPON_2006_18_public_sector	Regional data - 1250
Typology Com APAA		Average Annual Pre-Accession-Aid spending	Between 1998 and 2000	9901 Cross-thematic and non-themat	NUTS	1999, NUTS 2003	ESPON_2006_18_public_sector	Regional data - 1251
Typology Com TPAA		Total Pre-Accession-Aid spending	Between 1998 and 2000	9901 Cross-thematic and non-themat	NUTS	1999	ESPON_2006_18_public_sector	Regional data - 1252
Typology Com PRAu		Percentage of regional Pre-Accession-Aid addressing urban	Between 1998 and 2000	9901 Cross-thematic and non-themat	NUTS	1999	ESPON_2006_18_public_sector	Regional data - 1253
Typology Com SFCE		Cohesion Fund expenditure (E)	Between 1994 and 1999	9901 Cross-thematic and non-themat	NUTS	1999	ESPON_2006_18_public_sector	Regional data - 1254
Typology Com PRAI		Percentage of regional Pre-Accession-Aid addressing labour	Between 1998 and 2000	9901 Cross-thematic and non-themat	NUTS	1999	ESPON_2006_18_public_sector	Regional data - 1255
Typology Com PRAi		Percentage of regional Pre-Accession-Aid addressing institu	Between 1998 and 2000	9901 Cross-thematic and non-themat	NUTS	1999	ESPON_2006_18_public_sector	Regional data - 1256
Typology Com PRAp		Percentage of regional Pre-Accession-Aid addressing poten	Between 1998 and 2000	9901 Cross-thematic and non-themat	NUTS	1999	ESPON_2006_18_public_sector	Regional data - 1257
Typology Com PRAg		Percentage of regional Pre-Accession-Aid addressing geogr	Between 1998 and 2000	9901 Cross-thematic and non-themat	NUTS	1999	ESPON_2006_18_public_sector	Regional data - 1258
Typology Com PRAe		Percentage of regional Pre-Accession-Aid addressing enviri	Between 1998 and 2000	9901 Cross-thematic and non-themat	NUTS	1999	ESPON_2006_18_public_sector	Regional data - 1259
Typology Com PRAc		Percentage of regional Pre-Accession-Aid addressing capit	Between 1998 and 2000	9901 Cross-thematic and non-themat	NUTS	1999	ESPON_2006_18_public_sector	Regional data - 1260
Typology Com PRAr		Percentage of regional Pre-Accession-Aid addressing regio	Between 1998 and 2000	9901 Cross-thematic and non-themat	NUTS	1999	ESPON_2006_18_public_sector	Regional data - 1261
Typology Com SFA		Structural Fund expenditure (A)	Between 1994 and 1999	9901 Cross-thematic and non-themat	NUTS	1999	ESPON_2006_18_public_sector	Regional data - 1262
Typology Com SFS		Structural Fund expenditure (S)	Between 1994 and 1999	9901 Cross-thematic and non-themat	NUTS	1999	ESPON_2006_18_public_sector	Regional data - 1263

## • Local data (1)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
METROBORDER	country_belong	Country belonging of the LAU2 unit	2010	9902 Non/Cross-thematic data - Geograp	LAU2	2010	PART OF ESPON	LOCAL DATA - 1
METROBORDER	pop_t_rtc	Total population	2000	0201 Demography - Population structure	LAU2	2010	PART OF ESPON	LOCAL DATA - 2
METROBORDER	pop_t_rtc	Total population	2006	0201 Demography - Population structure	LAU2	2010	PART OF ESPON	LOCAL DATA - 3
METROBORDER	pop_nres_rtc	Non-resident population	2000	0201 Demography - Population structure	LAU2	2010	PART OF ESPON	LOCAL DATA - 4
METROBORDER	pop_nres_rtc	Non-resident population	2006	0201 Demography - Population structure	LAU2	2010	PART OF ESPON	LOCAL DATA - 5
METROBORDER	pop_nres_growth_rtp	Average annual growth of non-residents population	2000-2006	0201 Demography - Population structure	LAU2	2010	PART OF ESPON	LOCAL DATA - 6
METROBORDER	cb_commuters_origin	Origin of the commuters	2010	0302 Transport - Passengers and Goods	Cross border metropolitan region	2010	PART OF ESPON	LOCAL DATA - 7
METROBORDER	cb_commuters_dest	Destination of the commuters	2010	0302 Transport - Passengers and Goods	Cross border metropolitan region	2010	PART OF ESPON	LOCAL DATA - 8
METROBORDER	cb_commuters	Number of cross-border commuters	2000	0302 Transport - Passengers and Goods	Cross border metropolitan region	2010	PART OF ESPON	LOCAL DATA - 9
METROBORDER	cb_commuters	Number of cross-border commuters	2006	0302 Transport - Passengers and Goods	Cross border metropolitan region	2010	PART OF ESPON	LOCAL DATA - 10
METROBORDER	metroborder_belong	Cross border metropolitan region belonging of the LAU2 unit	2010	9902 Non/cross-thematic data - Geograp	Matrix - LAU2*LAU2	2010	PART OF ESPON	LOCAL DATA - 11
METROBORDER	Distance	Euclidian distance between the two cities	2010	0301 Transport - Transport Infrastructure	Matrix - LAU2*LAU2	2010	PART OF ESPON	LOCAL DATA - 12
METROBORDER	T_BUS_MIN	Average time (bus)	2010	0301 Transport - Transport Infrastructure	Matrix - LAU2*LAU2	2010	PART OF ESPON	LOCAL DATA - 13
METROBORDER	NB_BUS_MIN	Number of links (bus)	2010	0301 Transport - Transport Infrastructure	Matrix - LAU2*LAU2	2010	PART OF ESPON	LOCAL DATA - 14
METROBORDER	T_TRAIN_MIN	Average time (train)	2010	0301 Transport - Transport Infrastructure	Matrix - LAU2*LAU2	2010	PART OF ESPON	LOCAL DATA - 15
METROBORDER	NB_TRAIN	Number of links (train)	2010	0301 Transport - Transport Infrastructure	Matrix - LAU2*LAU2	2010	PART OF ESPON	LOCAL DATA - 16
METROBORDER	T_BOAT	Average time (boat)	2010	0301 Transport - Transport Infrastructure	Matrix - LAU2*LAU2	2010	PART OF ESPON	LOCAL DATA - 17
METROBORDER	NB_TRAIN	Number of links (boat)	2010	0301 Transport - Transport Infrastructure	Matrix - LAU2*LAU2	2010	PART OF ESPON	LOCAL DATA - 18
METROBORDER	T_MIN	Minimum time	2010	0301 Transport - Transport Infrastructure	Matrix - LAU2*LAU2	2010	PART OF ESPON	LOCAL DATA - 19
METROBORDER	Av_Speed_M	Average speed	2010	0301 Transport - Transport Infrastructure	Matrix - LAU2*LAU2	2010	PART OF ESPON	LOCAL DATA - 20
METROBORDER	Av_Speed_H	Average speed	2010	0301 Transport - Transport Infrastructure	Matrix - LAU2*LAU2	2010	PART OF ESPON	LOCAL DATA - 21
NUTS5 Database	D-NS_1a	Total population and by sex (total) by municipality	1981	0201 Demography - Population structure	NUTS5	1997	ESPON 29	LOCAL DATA - 22
NUTS5 Database	D-NS_1a	Total population and by sex (total) by municipality	1991	0201 Demography - Population structure	NUTS5	1997	ESPON 29	LOCAL DATA - 23
NUTS5 Database	D-NS_1a	Total population and by sex (total) by municipality	2001	0201 Demography - Population structure	NUTS5	1997	ESPON 29	LOCAL DATA - 24
NUTS5 Database	D-NS_1b	Total population and by sex (female) by municipality	1981	0201 Demography - Population structure	NUTS5	1997	ESPON 29	LOCAL DATA - 25
NUTS5 Database	D-NS_1b	Total population and by sex (female) by municipality	1991	0201 Demography - Population structure	NUTS5	1997	ESPON 29	LOCAL DATA - 26
NUTS5 Database	D-NS_1b	Total population and by sex (female) by municipality	2001	0201 Demography - Population structure	NUTS5	1997	ESPON 29	LOCAL DATA - 27
NUTS5 Database	D-NS_1c	Total population and by sex (male) by municipality	1981	0201 Demography - Population structure	NUTS5	1997	ESPON 29	LOCAL DATA - 28
NUTS5 Database	D-NS_1c	Total population and by sex (male) by municipality	1991	0201 Demography - Population structure	NUTS5	1997	ESPON 29	LOCAL DATA - 29
NUTS5 Database	D-NS_1c	Total population and by sex (male) by municipality	2001	0201 Demography - Population structure	NUTS5	1997	ESPON 29	LOCAL DATA - 30
NUTS5 Database	D-NS_2a	Number of inhabitants over 60 by municipality	1981	0201 Demography - Population structure	NUTS5	1997	ESPON 29	LOCAL DATA - 31
NUTS5 Database	D-NS_2a	Number of inhabitants over 60 by municipality	1991	0201 Demography - Population structure	NUTS5	1997	ESPON 29	LOCAL DATA - 32
NUTS5 Database	D-NS_2a	Number of inhabitants over 60 by municipality	2001	0201 Demography - Population structure	NUTS5	1997	ESPON 29	LOCAL DATA - 33
NUTS5 Database	D-NS_2b	Number of inhabitants under 15 by municipality	1981	0201 Demography - Population structure	NUTS5	1997	ESPON 29	LOCAL DATA - 34
NUTS5 Database	D-NS_2b	Number of inhabitants under 15 by municipality	1991	0201 Demography - Population structure	NUTS5	1997	ESPON 29	LOCAL DATA - 35
NUTS5 Database	D-NS_2b	Number of inhabitants under 15 by municipality	2001	0201 Demography - Population structure	NUTS5	1997	ESPON 29	LOCAL DATA - 36
NUTS5 Database	D-NS_3x	Number of births per year (annual average 81-91 and 91-2001)	1981-1991; 1991-2001	0202 Demography - Natural changes	NUTS5	1997	ESPON 29	LOCAL DATA - 37
NUTS5 Database	D-NS_3y	Number of deaths per year (annual average 81-91 and 91-2001)	1981-1991; 1991-2001	0202 Demography - Natural changes	NUTS5	1997	ESPON 29	LOCAL DATA - 38
NUTS5 Database	D-NS_4x	Number of immigrants (annual average 81-91 & 91-2001)	1981-1991; 1991-2001	0204 Demography - Migrations	NUTS5	1997	ESPON 29	LOCAL DATA - 39
NUTS5 Database	D-NS_4y	Number of emigrants (annual average 81-91 & 91-2001)	1981-1991; 1991-2001	0204 Demography - Migrations	NUTS5	1997	ESPON 29	LOCAL DATA - 40
NUTS5 Database	D-NS_4a	Net migration	1981-2001	0204 Demography - Migrations	NUTS5	1997	ESPON 29	LOCAL DATA - 41
NUTS5 Database	D-NS_5	Total number of foreign people by municipality	1981	0204 Demography - Migrations	NUTS5	1997	ESPON 29	LOCAL DATA - 42
NUTS5 Database	D-NS_5	Total number of foreign people by municipality	1991	0204 Demography - Migrations	NUTS5	1997	ESPON 29	LOCAL DATA - 43
NUTS5 Database	D-NS_5	Total number of foreign people by municipality	2001	0204 Demography - Migrations	NUTS5	1997	ESPON 29	LOCAL DATA - 44
NUTS5 Database	E-NS_1a	Total number active people (total) by municipality	2001	0602 Social affairs - Labour market	NUTS5	1997	ESPON 29	LOCAL DATA - 45
NUTS5 Database	E-NS_1b	Total number active people (female) by municipality	2001	0602 Social affairs - Labour market	NUTS5	1997	ESPON 29	LOCAL DATA - 46
NUTS5 Database	E-NS_1c	Total number active people (male) by municipality	2001	0602 Social affairs - Labour market	NUTS5	1997	ESPON 29	LOCAL DATA - 47
NUTS5 Database	E-NS_2a	Total number of employees in the agricultural sector	1981	0702 Economy - Employment	NUTS5	1997	ESPON 29	LOCAL DATA - 48
NUTS5 Database	E-NS_2a	Total number of employees in the agricultural sector	1991	0702 Economy - Employment	NUTS5	1997	ESPON 29	LOCAL DATA - 49
NUTS5 Database	E-NS_2a	Total number of employees in the agricultural sector	2001	0702 Economy - Employment	NUTS5	1997	ESPON 29	LOCAL DATA - 50
NUTS5 Database	E-NS_2b	Total number of employees in the industry sector	1981	0702 Economy - Employment	NUTS5	1997	ESPON 29	LOCAL DATA - 51
NUTS5 Database	E-NS_2b	Total number of employees in the industry sector	1991	0702 Economy - Employment	NUTS5	1997	ESPON 29	LOCAL DATA - 52

## • Local data (2)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
NUTS5 Database	E-NS_2c	Total number of employees in the services sector	1991	0702 Economy - Employment	NUTS5	1997	ESPON 29	LOCAL DATA - 55
NUTS5 Database	E-NS_2c	Total number of employees in the services sector	2001	0702 Economy - Employment	NUTS5	1997	ESPON 29	LOCAL DATA - 56
NUTS5 Database	E-NS_4a	Total number of unemployed people (total)	2001	0702 Economy - Employment	NUTS5	1997	ESPON 29	LOCAL DATA - 57
NUTS5 Database	E-NS_4b	Total number of unemployed people (female)	2001	0702 Economy - Employment	NUTS5	1997	ESPON 29	LOCAL DATA - 58
NUTS5 Database	E-NS_4c	Total number of unemployed people (male)	2001	0702 Economy - Employment	NUTS5	1997	ESPON 29	LOCAL DATA - 59
NUTS5 Database	A-NS_1	Number of farms by municipality	2001	0101 Agriculture & Fisheries - Farm structure	NUTS5	1997	PART OF ESPON	LOCAL DATA - 60
NUTS5 Database	A-NS_3a	Total livestock (livestock unit equivalents) by municipality	2001	0102 Agriculture & Fisheries - Livestock	NUTS5	1997	PART OF ESPON	LOCAL DATA - 61
NUTS5 Database	A-NS_3b	Cattle by municipality	2001	0102 Agriculture & Fisheries - Livestock	NUTS5	1997	PART OF ESPON	LOCAL DATA - 62
NUTS5 Database	A-NS_3c	Pigs by municipality	2001	0102 Agriculture & Fisheries - Livestock	NUTS5	1997	PART OF ESPON	LOCAL DATA - 63
NUTS5 Database	A-NS_3d	Poultry by municipality	2001	0102 Agriculture & Fisheries - Livestock	NUTS5	1997	PART OF ESPON	LOCAL DATA - 64
NUTS5 Database	A-NS_3e	Sheep by municipality	2001	0102 Agriculture & Fisheries - Livestock	NUTS5	1997	PART OF ESPON	LOCAL DATA - 65
NUTS5 Database	A-NS_8	Utilisable agricultural area by municipality (in sq. km)	2001	0503 Land use - Rural land use attributes	NUTS5	1997	PART OF ESPON	LOCAL DATA - 66
NUTS5 Database	E-NS_5	Total number of unemployed people under -25 years by municipality	2001	0702 Economy - Employment	NUTS5	1997	PART OF ESPON	LOCAL DATA - 67
NUTS5 Database	E-NS_6	Total number of long-term unemployed (i.e. longer than 1 year)	2001	0702 Economy - Employment	NUTS5	1997	PART OF ESPON	LOCAL DATA - 68
NUTS5 Database	E-NS_11	Level of qualification: number of people with highest education	2001	0601 Social affairs - Education	NUTS5	1997	PART OF ESPON	LOCAL DATA - 69
NUTS5 Database	E-NS_18	Disposable income per capita	2001	0602 Social affairs - Labour market	NUTS5	1997	PART OF ESPON	LOCAL DATA - 70
NUTS5 Database	E-NS_22	Commuting patterns: In- /out-commuting municipality	2001	0204 Demography - Migrations	NUTS5	1997	PART OF ESPON	LOCAL DATA - 71
NUTS5 Database	E-NS_23	Total number of active people (incl. self-employed & job-seekers)	2001	0602 Social affairs - Labour market	NUTS5	1997	PART OF ESPON	LOCAL DATA - 72
NUTS5 Database	I-NS_5	Total number of overnight stays of tourists by municipality	2001	0604 Social affairs - Culture	NUTS5	1997	PART OF ESPON	LOCAL DATA - 73
NUTS5 Database	I-NS_12	Number of doctors by municipality	2001	0603 Social affairs- Living conditions	NUTS5	1997	PART OF ESPON	LOCAL DATA - 74
NUTS5 Database	I-NS_18	Number of secondary by municipality	2001	0603 Social affairs- Living conditions	NUTS5	1997	PART OF ESPON	LOCAL DATA - 75
NUTS5 Database	A-GIS_5	Total forest area by municipality	2001	0503 Land use - Rural land use attributes	NUTS5	1997	PART OF ESPON	LOCAL DATA - 76
NUTS5 Database	A-GIS_6	Share of forest area by municipality	2001	0503 Land use - Rural land use attributes	NUTS5	1997	PART OF ESPON	LOCAL DATA - 77
NUTS5 Database	A-GIS_7	Total arable land by municipality	2001	0503 Land use - Rural land use attributes	NUTS5	1997	PART OF ESPON	LOCAL DATA - 78
NUTS5 Database	A-GIS_8	Share of arable land by municipality	2001	0503 Land use - Rural land use attributes	NUTS5	1997	PART OF ESPON	LOCAL DATA - 79
NUTS5 Database	EN-GIS_7a	Degree of urbanisation by municipality	2001	0502 Land use - Urban land use attributes	NUTS5	1997	PART OF ESPON	LOCAL DATA - 80
NUTS5 Database	EN-GIS_20a	Climatic Contrast Index by municipality	2001	0402 Energy & Environment - Climate change	NUTS5	1997	PART OF ESPON	LOCAL DATA - 81
NUTS5 Database	G-GIS_1a	Longitude of municipality centre	2001	3902 Non/Cross-thematic data - Geographical	NUTS5	1997	PART OF ESPON	LOCAL DATA - 82
NUTS5 Database	G-GIS_1b	Latitude of municipality centre	2001	3902 Non/Cross-thematic data - Geographical	NUTS5	1997	PART OF ESPON	LOCAL DATA - 83
NUTS5 Database	G-GIS_2a	Mean altitude by municipality	2001	3902 Non/Cross-thematic data - Geographical	NUTS5	1997	PART OF ESPON	LOCAL DATA - 84
NUTS5 Database	G-GIS_2b	Minimum altitude by municipality	2001	3902 Non/Cross-thematic data - Geographical	NUTS5	1997	PART OF ESPON	LOCAL DATA - 85
NUTS5 Database	G-GIS_2c	Maximum altitude by municipality	2001	3902 Non/Cross-thematic data - Geographical	NUTS5	1997	PART OF ESPON	LOCAL DATA - 86
NUTS5 Database	G-GIS_3	Ratio minimum/maximum elevation by municipality	2001	3902 Non/Cross-thematic data - Geographical	NUTS5	1997	PART OF ESPON	LOCAL DATA - 87
NUTS5 Database	G-GIS_5	Standard deviation of elevation by municipality	2001	3902 Non/Cross-thematic data - Geographical	NUTS5	1997	PART OF ESPON	LOCAL DATA - 88
NUTS5 Database	G-GIS_7a	Mean slope by municipality	2001	3902 Non/Cross-thematic data - Geographical	NUTS5	1997	PART OF ESPON	LOCAL DATA - 89
NUTS5 Database	G-GIS_7b	Steepest slope by municipality	2001	3902 Non/Cross-thematic data - Geographical	NUTS5	1997	PART OF ESPON	LOCAL DATA - 90
NUTS5 Database	G-GIS_8a	Total municipality area	2001	3902 Non/Cross-thematic data - Geographical	NUTS5	1997	PART OF ESPON	LOCAL DATA - 91
NUTS5 Database	G-GIS_10b	Average minimum temperature by municipality	2001	0402 Energy & Environment - Climate change	NUTS5	1997	PART OF ESPON	LOCAL DATA - 92
NUTS5 Database	G-GIS_10d	Average maximum temperature by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 93
NUTS5 Database	G-GIS_17a	Land use by coniferous forest by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 94
NUTS5 Database	G-GIS_17b	Land use by deciduous forest by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 95
NUTS5 Database	G-GIS_17c	Land use by mixed forest by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 96
NUTS5 Database	G-GIS_17d	Land use by grassland by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 97
NUTS5 Database	G-GIS_17e	Land use by rainfed arable land by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 98
NUTS5 Database	G-GIS_17f	Land use by irrigated arable land by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 99
NUTS5 Database	G-GIS_17g	Land use by permanent crops by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 100
NUTS5 Database	G-GIS_17h	Land use by scrubland by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 101
NUTS5 Database	G-GIS_17i	Land use by barren land by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 102
NUTS5 Database	G-GIS_17j	Land use by permanent ice and snow by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 103
NUTS5 Database	G-GIS_17k	Land use by wetlands by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 104
NUTS5 Database	G-GIS_17l	Land use by inland waters by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 105
NUTS5 Database	G-GIS_17m	Land use by urban areas by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 106



## • Local data (3)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
NUTS5 Database	G-GIS_19a	Typology of municipalities according to main land use	2001	9901 Non/cross thematic data - Integrativ	NUTS5	1997	PART OF ESPON	LOCAL DATA - 107
NUTS5 Database	G-GIS_22a	Airline distance from municipality centres to nearest city	2001	0301 Transport - Accessibility	NUTS5	1997	PART OF ESPON	LOCAL DATA - 108
NUTS5 Database	G-GIS_22b	Average airline distance to the next three cities	2001	0301 Transport - Accessibility	NUTS5	1997	PART OF ESPON	LOCAL DATA - 109
NUTS5 Database	G-GIS_23a	Airline distance from municipality centres to capital city	2001	0301 Transport - Accessibility	NUTS5	1997	PART OF ESPON	LOCAL DATA - 110
NUTS5 Database	G-GIS_30a	Airline distance to next hospital by municipality	2001	0603 Social affairs - Living conditions	NUTS5	1997	PART OF ESPON	LOCAL DATA - 111
NUTS5 Database	G-GIS_30b	Airline distance to next hospital by municipality	2001	0603 Social affairs - Living conditions	NUTS5	1997	PART OF ESPON	LOCAL DATA - 112
NUTS5 Database	G-GIS_30c	Airline distance to third next hospital by municipality	2001	0603 Social affairs - Living conditions	NUTS5	1997	PART OF ESPON	LOCAL DATA - 113
NUTS5 Database	G-GIS_30d	Average airline distance to next 3 hospital by municipality	2001	0603 Social affairs - Living conditions	NUTS5	1997	PART OF ESPON	LOCAL DATA - 114
NUTS5 Database	G-GIS_30e	Average airline distance to all hospital by municipality	2001	0603 Social affairs - Living conditions	NUTS5	1997	PART OF ESPON	LOCAL DATA - 115
NUTS5 Database	G-GIS_32b	Distance to nearest university by municipality	2001	0603 Social affairs - Living conditions	NUTS5	1997	PART OF ESPON	LOCAL DATA - 116
NUTS5 Database	G-GIS_32d	Potential accessibility to universities by municipality	2001	0603 Social affairs - Living conditions	NUTS5	1997	PART OF ESPON	LOCAL DATA - 117
NUTS5 Database	G-GIS_32e	Travel time to nearest university by municipality	2001	0603 Social affairs - Living conditions	NUTS5	1997	PART OF ESPON	LOCAL DATA - 118
NUTS5 Database	G-GIS_36a	Share of coniferous forests by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 119
NUTS5 Database	G-GIS_36b	Share of deciduous forests by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 120
NUTS5 Database	G-GIS_36c	Share of mixed forests by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 121
NUTS5 Database	G-GIS_36d	Share of grassland by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 122
NUTS5 Database	G-GIS_36e	Share of rainfed arable land by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 123
NUTS5 Database	G-GIS_36f	Share of irrigated arable land by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 124
NUTS5 Database	G-GIS_36g	Share of permanent crops by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 125
NUTS5 Database	G-GIS_36h	Share of scrubland by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 126
NUTS5 Database	G-GIS_36i	Share of barren land by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 127
NUTS5 Database	G-GIS_36j	Share of permanent ice and snow by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 128
NUTS5 Database	G-GIS_36k	Share of wetlands by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 129
NUTS5 Database	G-GIS_36l	Share of inland waters by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 130
NUTS5 Database	G-GIS_36m	Share of urban areas by municipality	2001	0501 Land use - Land Use and land cover t	NUTS5	1997	PART OF ESPON	LOCAL DATA - 131
NUTS5 Database	I-GIS_16a	Number of ports by municipality	2001	0303 Transport - Infrastructure	NUTS5	1997	PART OF ESPON	LOCAL DATA - 132
NUTS5 Database	I-GIS_17a	Number of airports by municipality	2001	0303 Transport - Infrastructure	NUTS5	1997	PART OF ESPON	LOCAL DATA - 133
NUTS5 Database	I-GIS_17c	Car travel time to nearest airport by municipality	2001	0303 Transport - Infrastructure	NUTS5	1997	PART OF ESPON	LOCAL DATA - 134
NUTS5 Database	I-GIS_17f	Potential accessibility to airports by municipality	2001	0301 Transport - Accessibility	NUTS5	1997	PART OF ESPON	LOCAL DATA - 135
NUTS5 Database	I-GIS_18	Peripherality by car to population by municipalities (EU average)	2001	0301 Transport - Accessibility	NUTS5	1997	PART OF ESPON	LOCAL DATA - 136
NUTS5 Database	I-GIS_20	Peripherality by car to population by municipalities (nat. averages)	2001	0301 Transport - Accessibility	NUTS5	1997	PART OF ESPON	LOCAL DATA - 137
NUTS5 Database	I-GIS_30a	No of people living within 1h car driving time by municipality	2001	0301 Transport - Accessibility	NUTS5	1997	PART OF ESPON	LOCAL DATA - 138
NUTS5 Database	I-GIS_31a	No of people living within 30 min driving time by municipality	2001	0301 Transport - Accessibility	NUTS5	1997	PART OF ESPON	LOCAL DATA - 139
NUTS5 Database	I-GIS_32a	Length of motorways by municipality	2001	0303 Transport - Infrastructure	NUTS5	1997	PART OF ESPON	LOCAL DATA - 140
NUTS5 Database	I-GIS_32b	Length of dual-carriageways by municipality	2001	0303 Transport - Infrastructure	NUTS5	1997	PART OF ESPON	LOCAL DATA - 141
NUTS5 Database	I-GIS_32c	Length of other trunk roads by municipality	2001	0303 Transport - Infrastructure	NUTS5	1997	PART OF ESPON	LOCAL DATA - 142
NUTS5 Database	I-GIS_33a	Density of motorways by municipality	2001	0303 Transport - Infrastructure	NUTS5	1997	PART OF ESPON	LOCAL DATA - 143
NUTS5 Database	I-GIS_33b	Density of dual-carriageways by municipality	2001	0303 Transport - Infrastructure	NUTS5	1997	PART OF ESPON	LOCAL DATA - 144
NUTS5 Database	I-GIS_34	Density of other trunk roads by municipality by municipality	2001	0303 Transport - Infrastructure	NUTS5	1997	PART OF ESPON	LOCAL DATA - 145
NUTS5 Database	I-GIS_35a	Length of existing and planned TEN road network by municipality (in km)	2001	0303 Transport - Infrastructure	NUTS5	1997	PART OF ESPON	LOCAL DATA - 146
NUTS5 Database	I-GIS_36a	Length of existing TEN road network by municipality (in km)	2001	0303 Transport - Infrastructure	NUTS5	1997	PART OF ESPON	LOCAL DATA - 147
NUTS5 Database	I-GIS_37a	Length of planned TEN road network by municipality (in km)	2001	0303 Transport - Infrastructure	NUTS5	1997	PART OF ESPON	LOCAL DATA - 148
NUTS5 Database	I-GIS_38a	Density of navigable waterways by municipality	2001	0303 Transport - Infrastructure	NUTS5	1997	PART OF ESPON	LOCAL DATA - 149
NUTS5 Database	I-GIS_39a	Density of existing TEN road network by municipality	2001	0303 Transport - Infrastructure	NUTS5	1997	PART OF ESPON	LOCAL DATA - 150
NUTS5 Database	I-GIS_40a	Density of planned TEN road network by municipality	2001	0303 Transport - Infrastructure	NUTS5	1997	PART OF ESPON	LOCAL DATA - 151
NUTS5 Database	I-GIS_41a	Density of existing & planned TEN road network by municipality	2001	0303 Transport - Infrastructure	NUTS5	1997	PART OF ESPON	LOCAL DATA - 152
SS-LR	rgdpme2025_ref	real GDP in euro (2025 reference scenario)	2025	9901 Non/cross thematic data - Integrativ	NUTS3	2006	PART OF ESPON	LOCAL DATA - 153
SS-LR	r_rgdp0525_ref	annual growth rate of real GDP (2005-2025 reference scenario)	2005-2025	9901 Non/cross thematic data - Integrativ	NUTS3	2006	PART OF ESPON	LOCAL DATA - 154
SS-LR	diff_r_rgdp0525_ref	differential growth rate of real GDP (2005-2025 reference scenario)	2005-2025	9901 Non/cross thematic data - Integrativ	NUTS3	2006	PART OF ESPON	LOCAL DATA - 155
SS-LR	rgdpme2025_pro	real GDP in euro (2025 proactive scenario)	2025	9901 Non/cross thematic data - Integrativ	NUTS3	2006	PART OF ESPON	LOCAL DATA - 156
SS-LR	r_rgdp0525_pro	annual growth rate of real GDP (2005-2025 reference scenario)	2005-2025	9901 Non/cross thematic data - Integrativ	NUTS3	2006	PART OF ESPON	LOCAL DATA - 157
SS-LR	diff_r_rgdp0525_pro	differential growth rate of real GDP (2005-2025 proactive scenario)	2005-2025	9901 Non/cross thematic data - Integrativ	NUTS3	2006	PART OF ESPON	LOCAL DATA - 158

- Local data (4)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
SS-LR	rgdpme2025_def	real GDP in euro (2025 defensive scenario)	2005-2025	9901 Non/cross thematic data - Integrativ	NUTS3	2006	PART OF ESPON	LOCAL DATA - 159
SS-LR	r_rgdp0525_def	annual growth rate of real GDP (2005-2025 def scenario)	2005-2025	9901 Non/cross thematic data - Integrativ	NUTS3	2006	PART OF ESPON	LOCAL DATA - 160
SS-LR	diff_r_rgdp0525_def	differential growth rate of real GDP (2005-2025 defensive scenario)	2005-2025	9901 Non/cross thematic data - Integrativ	NUTS3	2006	PART OF ESPON	LOCAL DATA - 161
SS-LR	dif_pro_ref	difference between proactive and reference scenario	2005-2025	9901 Non/cross thematic data - Integrativ	NUTS3	2006	PART OF ESPON	LOCAL DATA - 162
SS-LR	dif_rea_def	difference between defensive and reference scenario	2005-2025	9901 Non/cross thematic data - Integrativ	NUTS3	2006	PART OF ESPON	LOCAL DATA - 163
SS-LR	area_km2	total area	2003	9902 Non/Cross-thematic data - Geograf	NUTS3	2006	PART OF ESPON	LOCAL DATA - 164
SS-LR	ufi296n3	share of urban fabric	2001	0502 Land Use - Urban land use attributes	NUTS3	2006	PART OF ESPON	LOCAL DATA - 165
SS-LR	tourism2001	tourism infrastructure	1996	0604 Social affairs - Culture	NUTS3	2006	PART OF ESPON	LOCAL DATA - 166
SS-LR	MB9699N23	migratory balance	2001	0204 Demography - Migrations	NUTS3	2006	PART OF ESPON	LOCAL DATA - 167
SS-LR	people_less20_01	young people	2001	0201 Demography - Population structure	NUTS3	2006	PART OF ESPON	LOCAL DATA - 168
SS-LR	isco7_all01	craft and related trade workers	2001	0602 Social affairs - Labour market	NUTS3	2006	PART OF ESPON	LOCAL DATA - 169
SS-LR	empserv_all01	service employment	2001	0702 Economy - Employment	NUTS3	2006	PART OF ESPON	LOCAL DATA - 170
SS-LR	EU_election	European Elections, number of valid votes	1994	XXXXX - Other	NUTS3	2006	PART OF ESPON	LOCAL DATA - 171
SS-LR	EU_election	European Elections, number of valid votes	1999	XXXXX - Other	NUTS3	2006	PART OF ESPON	LOCAL DATA - 172
SS-LR	LRo01N3	Length of road network (km)	2001	0303 Transport - Infrastructure	NUTS3	2006	PART OF ESPON	LOCAL DATA - 173
SS-LR	LR01N3	Length of railway network, km	2001	0303 Transport - Infrastructure	NUTS3	2006	PART OF ESPON	LOCAL DATA - 174
SS-LR	CM01N3	Time to the nearest motorway access, by car of the capital or centroid	2001	0301 Transport - Accessibility	NUTS3	2006	PART OF ESPON	LOCAL DATA - 175
SS-LR	AcAiE01N3	Potential accessibility air, ESPON space = 100	2001	0301 Transport - Accessibility	NUTS3	2006	PART OF ESPON	LOCAL DATA - 176
SS-LR	AcRaE01N3	Potential accessibility rail, ESPON space = 100	2001	0301 Transport - Accessibility	NUTS3	2006	PART OF ESPON	LOCAL DATA - 177
SS-LR	AcRoE01N3	Potential accessibility road, ESPON space = 100	2001	0301 Transport - Accessibility	NUTS3	2006	PART OF ESPON	LOCAL DATA - 178
SS-LR	AcME01N3	Potential accessibility multimodal, ESPON space = 100	2001	0301 Transport - Accessibility	NUTS3	2006	PART OF ESPON	LOCAL DATA - 179
SS-LR	isco2_all01	professionals	2001	0702 Economy - Employment	NUTS3	2006	PART OF ESPON	LOCAL DATA - 180
SS-LR	isced5_6_01	tertiary education	2001	0601 Social affairs - Education	NUTS3	2006	PART OF ESPON	LOCAL DATA - 181
SS-LR	population	Total population	1995	0201 Demography - Population structure	NUTS3	2006	PART OF ESPON	LOCAL DATA - 182
SS-LR	population	Total population	1996	0201 Demography - Population structure	NUTS3	2006	PART OF ESPON	LOCAL DATA - 183
SS-LR	population	Total population	1997	0201 Demography - Population structure	NUTS3	2006	PART OF ESPON	LOCAL DATA - 184
SS-LR	population	Total population	1998	0201 Demography - Population structure	NUTS3	2006	PART OF ESPON	LOCAL DATA - 185
SS-LR	population	Total population	1999	0201 Demography - Population structure	NUTS3	2006	PART OF ESPON	LOCAL DATA - 186
SS-LR	population	Total population	2000	0201 Demography - Population structure	NUTS3	2006	PART OF ESPON	LOCAL DATA - 187
SS-LR	population	Total population	2001	0201 Demography - Population structure	NUTS3	2006	PART OF ESPON	LOCAL DATA - 188
SS-LR	population	Total population	2002	0201 Demography - Population structure	NUTS3	2006	PART OF ESPON	LOCAL DATA - 189
SS-LR	population	Total population	2003	0201 Demography - Population structure	NUTS3	2006	PART OF ESPON	LOCAL DATA - 190
SS-LR	population	Total population	2004	0201 Demography - Population structure	NUTS3	2006	PART OF ESPON	LOCAL DATA - 191
SS-LR	population	Total population	2005	0201 Demography - Population structure	NUTS3	2006	PART OF ESPON	LOCAL DATA - 192
SS-LR	population	Total population	2006	0201 Demography - Population structure	NUTS3	2006	PART OF ESPON	LOCAL DATA - 193
SS-LR	population	Total population	2007	0201 Demography - Population structure	NUTS3	2006	PART OF ESPON	LOCAL DATA - 194
SS-LR	GDP_eur	GDP in euro	2001	0701 Economy - Aggregated accounts	NUTS3	2006	PART OF ESPON	LOCAL DATA - 195
SS-LR	GDP_eur	GDP in euro	2005	0701 Economy - Aggregated accounts	NUTS3	2006	PART OF ESPON	LOCAL DATA - 196
CAEE	NACE93_1_abs	Total locally based-jobs in NACE 1	1991	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 197
CAEE	NACE93_2_abs	Total locally based-jobs in NACE 2	1991	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 198
CAEE	NACE93_3_abs	Total locally based-jobs in NACE 3	1991	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 199
CAEE	NACE93_4_abs	Total locally based-jobs in NACE 4	1991	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 200
CAEE	NACE93_5_abs	Total locally based-jobs in NACE 5	1991	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 201
CAEE	NACE93_6_abs	Total locally based-jobs in NACE 6	1991	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 202
CAEE	NACE93_7_abs	Total locally based-jobs in NACE 7	1991	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 203
CAEE	NACE93_8_abs	Total locally based-jobs in NACE 8	1991	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 204
CAEE	NACE93_9_abs	Total locally based-jobs in NACE 9	1991	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 205
CAEE	NACE93_10_abs	Total locally based-jobs in NACE 10	1991	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 206
CAEE	NACE93_11_abs	Total locally based-jobs in NACE 11	1991	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 207
CAEE	NACE93_12_abs	Total locally based-jobs in NACE 12	1991	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 208
CAEE	NACE93_13_abs	Total locally based-jobs in NACE 13	1991	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 209
CAEE	NACE93_14_abs	Total locally based-jobs in NACE 14	1991	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 210























- Local data (15)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
CAEE	NACE93_45_abs	Total locally based-jobs in NACE 45	2001	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 732
CAEE	NACE93_46_abs	Total locally based-jobs in NACE 46	2001	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 733
CAEE	NACE93_47_abs	Total locally based-jobs in NACE 47	2001	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 734
CAEE	NACE93_48_abs	Total locally based-jobs in NACE 48	2001	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 735
CAEE	NACE93_49_abs	Total locally based-jobs in NACE 49	2001	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 736
CAEE	NACE93_50_abs	Total locally based-jobs in NACE 50	2001	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 737
CAEE	NACE93_51_abs	Total locally based-jobs in NACE 51	2001	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 738
CAEE	NACE93_52_abs	Total locally based-jobs in NACE 52	2001	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 739
CAEE	NACE93_53_abs	Total locally based-jobs in NACE 53	2001	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 740
CAEE	NACE93_54_abs	Total locally based-jobs in NACE 54	2001	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 741
CAEE	NACE93_55_abs	Total locally based-jobs in NACE 55	2001	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 742
CAEE	NACE93_56_abs	Total locally based-jobs in NACE 56	2001	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 743
CAEE	NACE93_57_abs	Total locally based-jobs in NACE 57	2001	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 744
CAEE	NACE93_58_abs	Total locally based-jobs in NACE 58	2001	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 745
CAEE	NACE93_59_abs	Total locally based-jobs in NACE 59	2001	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 746
CAEE	NACE93_60_abs	Total locally based-jobs in NACE 60	2001	0702 Economy - Employment	LAU2	2010	PART OF ESPON	LOCAL DATA - 747
CAEE	emp_t_abs	Employed persons	2003	0702 Economy - Employment	Lower Layer Super Output Area	2010	PART OF ESPON	LOCAL DATA - 748
CAEE	emp_t_abs	Employed persons	2008	0702 Economy - Employment	Lower Layer Super Output Area	2010	PART OF ESPON	LOCAL DATA - 749
CAEE	emp_creative_abs	Employed persons in the Creative / Digital / New Media Sector	2003	0702 Economy - Employment	Lower Layer Super Output Area	2010	PART OF ESPON	LOCAL DATA - 750
CAEE	emp_t_abs	Employed persons	2008	0702 Economy - Employment	Lower Layer Super Output Area	2010	PART OF ESPON	LOCAL DATA - 751
CAEE	emp_ing_abs	Employed persons in the Engineering Sector	2003	0702 Economy - Employment	Lower Layer Super Output Area	2010	PART OF ESPON	LOCAL DATA - 752
CAEE	emp_t_abs	Employed persons	2008	0702 Economy - Employment	Lower Layer Super Output Area	2010	PART OF ESPON	LOCAL DATA - 753
CAEE	emp_fp_abs	Employed persons in the financial and professional services sector	2003	0702 Economy - Employment	Lower Layer Super Output Area	2010	PART OF ESPON	LOCAL DATA - 754
CAEE	emp_fp_abs	Employed persons in the financial and professional services sector	2008	0702 Economy - Employment	Lower Layer Super Output Area	2010	PART OF ESPON	LOCAL DATA - 755
CAEE	gva_agr_abs	GVA at basic prices in agriculture, forestry and fishing	2000	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 756
CAEE	gva_agr_abs	GVA at basic prices in agriculture, forestry and fishing	2001	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 757
CAEE	gva_agr_abs	GVA at basic prices in agriculture, forestry and fishing	2002	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 758
CAEE	gva_agr_abs	GVA at basic prices in agriculture, forestry and fishing	2003	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 759
CAEE	gva_agr_abs	GVA at basic prices in agriculture, forestry and fishing	2004	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 760
CAEE	gva_agr_abs	GVA at basic prices in agriculture, forestry and fishing	2005	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 761
CAEE	gva_agr_abs	GVA at basic prices in agriculture, forestry and fishing	2006	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 762
CAEE	gva_agr_abs	GVA at basic prices in agriculture, forestry and fishing	2007	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 763
CAEE	gva_inds_abs	GVA at basic prices in manufacturing, building and construction	2000	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 764
CAEE	gva_inds_abs	GVA at basic prices in manufacturing, building and construction	2001	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 765
CAEE	gva_inds_abs	GVA at basic prices in manufacturing, building and construction	2002	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 766
CAEE	gva_inds_abs	GVA at basic prices in manufacturing, building and construction	2003	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 767
CAEE	gva_inds_abs	GVA at basic prices in manufacturing, building and construction	2004	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 768
CAEE	gva_inds_abs	GVA at basic prices in manufacturing, building and construction	2005	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 769
CAEE	gva_inds_abs	GVA at basic prices in manufacturing, building and construction	2006	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 770
CAEE	gva_inds_abs	GVA at basic prices in manufacturing, building and construction	2007	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 771
CAEE	gva_serv_abs	GVA at basic prices in market and non-market services	2000	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 772
CAEE	gva_serv_abs	GVA at basic prices in market and non-market services	2001	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 773
CAEE	gva_serv_abs	GVA at basic prices in market and non-market services	2002	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 774
CAEE	gva_serv_abs	GVA at basic prices in market and non-market services	2003	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 775
CAEE	gva_serv_abs	GVA at basic prices in market and non-market services	2004	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 776
CAEE	gva_serv_abs	GVA at basic prices in market and non-market services	2005	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 777
CAEE	gva_serv_abs	GVA at basic prices in market and non-market services	2006	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 778
CAEE	gva_t_abs	Total GVA	2007	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 779
CAEE	gva_t_abs	Total GVA	2000	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 780
CAEE	gva_t_abs	Total GVA	2001	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 781
CAEE	gva_t_abs	Total GVA	2002	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 782
CAEE	gva_t_abs	Total GVA	2003	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 783

- Local data (16)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
CAEE	gva_t_abs	Total GVA	2004	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 784
CAEE	gva_t_abs	Total GVA	2005	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 785
CAEE	gva_t_abs	Total GVA	2006	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 786
CAEE	gva_t_abs	Total GVA	2007	0703 Economy - Production and costs pe	NUTS3	2006	PART OF ESPON	LOCAL DATA - 787
TeDi	A_NS_1	Number of farm holdings	1991	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 788
TeDi	A_NS_1	Number of farm holdings	1995	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 789
TeDi	A_NS_1	Number of farm holdings	2000	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 790
TeDi	A_NS_1	Number of farm holdings	2001	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 791
TeDi	A_NS_1	Number of farm holdings	2003	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 792
TeDi	A_NS_1	Number of farm holdings	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 793
TeDi	A-NS_3a	Number of persons working in the agricultural sector	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 794
TeDi	A-NS_3b	Number of persons working in forestry and logging	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 795
TeDi	A-NS_3c	Number of persons working in fishing and aquaculture sector	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 796
TeDi	A-NS_5_upto24	Number of farm holders aged 24 and less	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 797
TeDi	A-NS_5_upto35	Number of farm holders aged 35 and less	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 798
TeDi	A-NS_5_upto39	Number of farm holders aged 39 and less	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 799
TeDi	A-NS_5_25_29	Number of farm holders aged 25-29	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 800
TeDi	A-NS_5_25_34	Number of farm holders aged 25-34	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 801
TeDi	A-NS_5_30_34	Number of farm holders aged 30-34	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 802
TeDi	A-NS_5_35_39	Number of farm holders aged 35-39	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 803
TeDi	A-NS_5_35_44	Number of farm holders aged 35-44	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 804
TeDi	A-NS_5_40_44	Number of farm holders aged 40-44	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 805
TeDi	A-NS_5_40_49	Number of farm holders aged 40-49	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 806
TeDi	A-NS_5_45_49	Number of farm holders aged 45-49	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 807
TeDi	A-NS_5_45_54	Number of farm holders aged 45-54	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 808
TeDi	A-NS_5_50_54	Number of farm holders aged 50-54	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 809
TeDi	A-NS_5_50_59	Number of farm holders aged 50-59	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 810
TeDi	A-NS_5_55_59	Number of farm holders aged 55-59	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 811
TeDi	A-NS_5_55_64	Number of farm holders aged 55-64	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 812
TeDi	A-NS_5_60_64	Number of farm holders aged 60-64	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 813
TeDi	A-NS_5_60_69	Number of farm holders aged 60-69	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 814
TeDi	A-NS_5_35_65	Number of farm holders aged 35-65	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 815
TeDi	A-NS_5_65_69	Number of farm holders aged 65-69	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 816
TeDi	A-NS_5_64_74	Number of farm holders aged 64-74	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 817
TeDi	A-NS_5_65more	Number of farm holders aged 65 or more	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 818
TeDi	A-NS_5_70more	Number of farm holders aged 70 or more	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 819
TeDi	A-NS_5_75more	Number of farm holders aged 75 or more	2007	0101 Agriculture & Fisheries - Farm struct	LAU2	2007	PART OF ESPON	LOCAL DATA - 820
TeDi	A-NS_7_CH_a	Arboriculture fruitière, viticulture, horticulture	1979-1985	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 821
TeDi	A-NS_7_CH_a	Arboriculture fruitière, viticulture, horticulture	1992-1997	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 822
TeDi	A-NS_7_CH_b	Prés et terres arables, pâturages locaux	1979-1985	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 823
TeDi	A-NS_7_CH_b	Prés et terres arables, pâturages locaux	1992-1997	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 824
TeDi	A-NS_7_CH_c	Alpages	1979-1985	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 825
TeDi	A-NS_7_CH_c	Alpages	1992-1997	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 826
TeDi	A-NS_7_CH_date	Année de relevé	1979-1985	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 827
TeDi	A-NS_7_CH_date	Année de relevé	1992-1997	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 828
TeDi	A-NS_7_CY_a	Temporary Crops	2003	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 829
TeDi	A-NS_7_CY_b	Tree Crops	2003	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 830
TeDi	A-NS_7_CY_c	Vines	2003	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 831
TeDi	A-NS_7_CY_d	Permanent Grasslands and Pastures	2003	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 832
TeDi	A-NS_7_CY_e	Fallow Land	2003	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 833
TeDi	A-NS_7_CY_f	Kitchen Gardens	2003	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 834
TeDi	A-NS_7_CY_g	Nurseries for fruit trees and other permanent crops	2003	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 835

- Local data (17)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
TeDi	A-NS_7_FL_a	Agricultural and horticultural land	2000	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 836
TeDi	A-NS_7_FL_a	Agricultural and horticultural land	2008	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 837
TeDi	A-NS_7_FL_b	Rough grazing and pasturage	2000	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 838
TeDi	A-NS_7_FL_b	Rough grazing and pasturage	2008	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 839
TeDi	A-NS_7_FL_c	Other land	2000	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 840
TeDi	A-NS_7_FL_c	Other land	2008	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 841
TeDi	A-NS_7_MT	Total arable land Malta	2001	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 842
TeDi	A-NS_7_MT	Total arable land Malta	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 843
TeDi	A-NS_7_NO	Agricultural area in use, total	2000	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 844
TeDi	A-NS_7_NO	Agricultural area in use, total	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 845
TeDi	A-NS_7_NO_a	Open fields and gardens	2000	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 846
TeDi	A-NS_7_NO_a	Open fields and gardens	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 847
TeDi	A-NS_7_NO_aa	Grain and oil seeds	2000	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 848
TeDi	A-NS_7_NO_aa	Grain and oil seeds	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 849
TeDi	A-NS_7_NO_aaa	Wheat	2000	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 850
TeDi	A-NS_7_NO_aaa	Wheat	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 851
TeDi	A-NS_7_NO_aab	Barley	2000	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 852
TeDi	A-NS_7_NO_aab	Barley	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 853
TeDi	A-NS_7_NO_aac	Oats	2000	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 854
TeDi	A-NS_7_NO_aac	Oats	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 855
TeDi	A-NS_7_NO_ab	Potato	2000	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 856
TeDi	A-NS_7_NO_ab	Potato	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 857
TeDi	A-NS_7_NO_ac	Crops for green fodder and silage	2000	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 858
TeDi	A-NS_7_NO_ac	Crops for green fodder and silage	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 859
TeDi	A-NS_7_NO_ad	Vegetables, field grown	2000	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 860
TeDi	A-NS_7_NO_ad	Vegetables, field grown	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 861
TeDi	A-NS_7_NO_b	Meadows for mowing and pastures	2000	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 862
TeDi	A-NS_7_NO_b	Meadows for mowing and pastures	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 863
TeDi	A-NS_7_NO_ba	Cultivated meadows	2000	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 864
TeDi	A-NS_7_NO_ba	Cultivated meadows	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 865
TeDi	A-NS_7_NO_bb	Permanent grassland and surface cultivated land	2000	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 866
TeDi	A-NS_7_NO_bb	Permanent grassland and surface cultivated land	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 867
TeDi	A-NS_7_RO	Total agricultural surface	1991	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 868
TeDi	A-NS_7_RO	Total agricultural surface	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 869
TeDi	A-NS_7_RO_a	Arable	1991	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 870
TeDi	A-NS_7_RO_a	Arable	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 871
TeDi	A-NS_7_RO_b	Pastures	1991	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 872
TeDi	A-NS_7_RO_b	Pastures	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 873
TeDi	A-NS_7_RO_c	hayfields	1991	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 874
TeDi	A-NS_7_RO_c	hayfields	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 875
TeDi	A-NS_7_RO_d	vineyards	1991	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 876
TeDi	A-NS_7_RO_d	vineyards	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 877
TeDi	A-NS_7_RO_e	orchards	1991	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 878
TeDi	A-NS_7_RO_e	orchards	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 879
TeDi	A-NS_7_SE	total arable land	1991	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 880
TeDi	A-NS_7_SE	total arable land	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 881
TeDi	A-NS_7_SE_a	spring wheat	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 882
TeDi	A-NS_7_SE_b	barley	1991	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 883
TeDi	A-NS_7_SE_b2	spring barley	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 884
TeDi	A-NS_7_SE_c	oats	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 885
TeDi	A-NS_7_SE_d	mixed grain and triticale	1991	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 886
TeDi	A-NS_7_SE_d2	mixed grain	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 887

- Local data (18)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
TeDi	A-NS_7_SE_e	green fodder	1991	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 888
TeDi	A-NS_7_SE_e	green fodder	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 889
TeDi	A-NS_7_SE_f	utilized ley for hay and pasture	1991	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 890
TeDi	A-NS_7_SE_f	utilized ley for hay and pasture	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 891
TeDi	A-NS_7_SE_g	not utilized ley for hay and pasture	1991	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 892
TeDi	A-NS_7_SE_h	ley for seeds	1991	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 893
TeDi	A-NS_7_SE_i	table potatoes	1991	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 894
TeDi	A-NS_7_SE_i	table potatoes	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 895
TeDi	A-NS_7_SE_j	potatoes for processing of starch	1991	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 896
TeDi	A-NS_7_SE_k	horticulture plants	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 897
TeDi	A-NS_7_SE_l	energy forest	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 898
TeDi	A-NS_7_SE_m	fallow land	1991	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 899
TeDi	A-NS_7_SE_m	fallow land	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 900
TeDi	A-NS_7_SE_n	other crops	1991	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 901
TeDi	A-NS_7_SE_o	other untilled arable land	1991	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 902
TeDi	A-NS_7_SE_o	other untilled arable land	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 903
TeDi	A-NS_7_SE_p	unspecified arable land	2007	0503 Land use - Rural land use attributes	LAU2	2007	PART OF ESPON	LOCAL DATA - 904
TeDi	D-NS_1a	Total population	1981	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 905
TeDi	D-NS_1a	Total population	1982	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 906
TeDi	D-NS_1a	Total population	1985	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 907
TeDi	D-NS_1a	Total population	1986	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 908
TeDi	D-NS_1a	Total population	1991	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 909
TeDi	D-NS_1a	Total population	1992	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 910
TeDi	D-NS_1a	Total population	2001	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 911
TeDi	D-NS_1b	Female population	2007	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 912
TeDi	D-NS_1b	Female population	1981	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 913
TeDi	D-NS_1b	Female population	1982	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 914
TeDi	D-NS_1b	Female population	1985	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 915
TeDi	D-NS_1b	Female population	1986	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 916
TeDi	D-NS_1b	Female population	1991	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 917
TeDi	D-NS_1b	Female population	1992	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 918
TeDi	D-NS_1b	Female population	2001	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 919
TeDi	D-NS_1c	Male population	2007	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 920
TeDi	D-NS_1c	Male population	1981	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 921
TeDi	D-NS_1c	Male population	1982	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 922
TeDi	D-NS_1c	Male population	1985	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 923
TeDi	D-NS_1c	Male population	1986	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 924
TeDi	D-NS_1c	Male population	1991	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 925
TeDi	D-NS_1c	Male population	1992	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 926
TeDi	D-NS_1c	Male population	2001	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 927
TeDi	D-NS_2_upto20	Population aged 19 and less	1990	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 928
TeDi	D-NS_2_upto20	Population aged 19 and less	1991	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 929
TeDi	D-NS_2_upto20	Population aged 19 and less	1992	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 930
TeDi	D-NS_2_upto20	Population aged 19 and less	2000	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 931
TeDi	D-NS_2_upto20	Population aged 19 and less	2001	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 932
TeDi	D-NS_2_upto20	Population aged 19 and less	2007	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 933
TeDi	D-NS_2_20_24	Population aged 20-24	1990	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 934
TeDi	D-NS_2_20_24	Population aged 20-24	1991	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 935
TeDi	D-NS_2_20_24	Population aged 20-24	1992	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 936
TeDi	D-NS_2_20_24	Population aged 20-24	2000	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 937
TeDi	D-NS_2_20_24	Population aged 20-24	2001	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 938
TeDi	D-NS_2_20_24	Population aged 20-24	2007	0201 Demography - Population structure	LAU2	2007	PART OF ESPON	LOCAL DATA - 939





- Local data (20)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
TeDi	D-NS_4a	Number of out migrants	1981	0204 Demography - Migrations	LAU2	2007	PART OF ESPON	LOCAL DATA - 992
TeDi	D-NS_4a	Number of out migrants	1991	0204 Demography - Migrations	LAU2	2007	PART OF ESPON	LOCAL DATA - 993
TeDi	D-NS_4a	Number of out migrants	2001	0204 Demography - Migrations	LAU2	2007	PART OF ESPON	LOCAL DATA - 994
TeDi	D-NS_4a	Number of out migrants	2003	0204 Demography - Migrations	LAU2	2007	PART OF ESPON	LOCAL DATA - 995
TeDi	D-NS_4a	Number of out migrants	2004	0204 Demography - Migrations	LAU2	2007	PART OF ESPON	LOCAL DATA - 996
TeDi	D-NS_4a	Number of out migrants	2005	0204 Demography - Migrations	LAU2	2007	PART OF ESPON	LOCAL DATA - 997
TeDi	D-NS_4a	Number of out migrants	2006	0204 Demography - Migrations	LAU2	2007	PART OF ESPON	LOCAL DATA - 998
TeDi	D-NS_4a	Number of out migrants	2007	0204 Demography - Migrations	LAU2	2007	PART OF ESPON	LOCAL DATA - 999
TeDi	D-NS_4a	Number of out migrants	2008	0204 Demography - Migrations	LAU2	2007	PART OF ESPON	LOCAL DATA - 1000
TeDi	D-NS_4b	Number of in migrants	1981	0204 Demography - Migrations	LAU2	2007	PART OF ESPON	LOCAL DATA - 1001
TeDi	D-NS_4b	Number of in migrants	1991	0204 Demography - Migrations	LAU2	2007	PART OF ESPON	LOCAL DATA - 1002
TeDi	D-NS_4b	Number of in migrants	2001	0204 Demography - Migrations	LAU2	2007	PART OF ESPON	LOCAL DATA - 1003
TeDi	D-NS_4b	Number of in migrants	2003	0204 Demography - Migrations	LAU2	2007	PART OF ESPON	LOCAL DATA - 1004
TeDi	D-NS_4b	Number of in migrants	2004	0204 Demography - Migrations	LAU2	2007	PART OF ESPON	LOCAL DATA - 1005
TeDi	D-NS_4b	Number of in migrants	2005	0204 Demography - Migrations	LAU2	2007	PART OF ESPON	LOCAL DATA - 1006
TeDi	D-NS_4b	Number of in migrants	2006	0204 Demography - Migrations	LAU2	2007	PART OF ESPON	LOCAL DATA - 1007
TeDi	D-NS_4b	Number of in migrants	2007	0204 Demography - Migrations	LAU2	2007	PART OF ESPON	LOCAL DATA - 1008
TeDi	D-NS_4b	Number of in migrants	2008	0204 Demography - Migrations	LAU2	2007	PART OF ESPON	LOCAL DATA - 1009
TeDi	D-NS_4c	Net migration	2002	0204 Demography - Migrations	LAU2	2007	PART OF ESPON	LOCAL DATA - 1010
TeDi	D-NS_4c	Net migration	2003	0204 Demography - Migrations	LAU2	2007	PART OF ESPON	LOCAL DATA - 1011
TeDi	D-NS_4c	Net migration	2004	0204 Demography - Migrations	LAU2	2007	PART OF ESPON	LOCAL DATA - 1012
TeDi	D-NS_4c	Net migration	2005	0204 Demography - Migrations	LAU2	2007	PART OF ESPON	LOCAL DATA - 1013
TeDi	D-NS_4c	Net migration	2006	0204 Demography - Migrations	LAU2	2007	PART OF ESPON	LOCAL DATA - 1014
TeDi	D-NS_5	Number of persons born abroad	2007	0204 Demography - Migrations	LAU2	2007	PART OF ESPON	LOCAL DATA - 1015
TeDi	E-NS_1a	Total active population	2007	0602 Social affairs - Labour market	LAU2	2007	PART OF ESPON	LOCAL DATA - 1016
TeDi	E-NS_1b	Active population female	2007	0602 Social affairs - Labour market	LAU2	2007	PART OF ESPON	LOCAL DATA - 1017
TeDi	E-NS_1c	Active population male	2007	0602 Social affairs - Labour market	LAU2	2007	PART OF ESPON	LOCAL DATA - 1018
TeDi	E-NS_4a	Total number of unemployed persons	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1019
TeDi	E-NS_4b	Number of unemployed persons, female	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1020
TeDi	E-NS_4c	Number of unemployed persons, male	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1021
TeDi	E-NS_2a	Total number of persons working in agriculture, hunting and forestry	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1022
TeDi	E-NS_2	Total number of persons working in fishing	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1023
TeDi	E-NS_2c	Total number of persons working in mining and quarrying	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1024
TeDi	E-NS_2d	Total number of persons working in manufacturing	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1025
TeDi	E-NS_2e	Total number of persons working in electricity, gas and water supply	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1026
TeDi	E-NS_2f	Total number of persons working in construction	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1027
TeDi	E-NS_2g	Total number of persons working in wholesale and retail trade; repair o	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1028
TeDi	E-NS_2h	Total number of persons working in hotels and restaurants	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1029
TeDi	E-NS_2i	Total number of persons working in transport, storage and communic	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1030
TeDi	E-NS_2j	Total number of persons working in financial intermediation	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1031
TeDi	E-NS_2k	Total number of persons working in real estate, renting and business a	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1032
TeDi	E-NS_2l	Total number of persons working in public administration and defence	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1033
TeDi	E-NS_2m	Total number of persons working in education	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1034
TeDi	E-NS_2n	Total number of persons working in health and social work	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1035
TeDi	E-NS_2o	Total number of persons working in other community, social and pers	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1036
TeDi	E-NS_2p	Total number of persons working in activities of households	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1037
TeDi	E-NS_2q	Total number of persons working in extra-territorial organizations and	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1038
TeDi	E-NS_2x	Unknown activity	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1039
TeDi	E-NS_5_15-19	Number of unemployed aged 15-19	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1040
TeDi	E-NS_5_20-24	Number of unemployed aged 20-24	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1041
TeDi	E-NS_5_25-29	Number of unemployed aged 25-29	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1042
TeDi	E-NS_5_30-34	Number of unemployed aged 30-34	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1043

- Local data (21)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
TeDi	E-NS_5_30-34	Number of unemployed aged 30-34	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1043
TeDi	E-NS_5_35-39	Number of unemployed aged 35-39	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1044
TeDi	E-NS_5_40-44	Number of unemployed aged 40-44	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1045
TeDi	E-NS_5_45-49	Number of unemployed aged 45-49	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1046
TeDi	E-NS_5_50-54	Number of unemployed aged 50-54	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1047
TeDi	E-NS_5_55-59	Number of unemployed aged 55-59	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1048
TeDi	E-NS_5_60-64	Number of unemployed aged 60-64	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1049
TeDi	E-NS_5_65more	Number of unemployed aged 65 and more	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1050
TeDi	E-NS_5_15-24	Number of unemployed aged 15-24	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1051
TeDi	E-NS_5_25-39	Number of unemployed aged 25-39	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1052
TeDi	E-NS_5_40-54	Number of unemployed aged 40-54	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1053
TeDi	E-NS_5_55more	Number of unemployed aged 55 and more	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1054
TeDi	E-NS_5_25-74	Number of unemployed aged 25-74	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1055
TeDi	E-NS_6	Number of people unemployed for more than 1 year	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1056
TeDi	E-NS_7	Total number of part-time employees	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1057
TeDi	E-NS_12	Number of companies created	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1058
TeDi	E-NS_13	Number of companies closed	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1059
TeDi	E-NS_8_upto9	Number of employees in companies with 9 employees or less	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1060
TeDi	E-NS_8_10-49	Number of employees in companies with 10 to 49 employees	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1061
TeDi	E-NS_8_50-100	Number of employees in companies with 50 to 100 employees	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1062
TeDi	E-NS_8_100-249	Number of employees in companies with 100 to 249 employees	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1063
TeDi	E-NS_8_100more	Number of employees in companies with 100 employees or more	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1064
TeDi	E-NS_8_250more	Number of employees in companies with 250 employees or more	2007	0702 Economy - Employment	LAU2	2007	PART OF ESPON	LOCAL DATA - 1065
TeDi	E-NS_10a	Number of persons with secondary education degree	2007	0601 Social affairs - Education	LAU2	2007	PART OF ESPON	LOCAL DATA - 1066
TeDi	E-NS_10b	Number of persons with tertiary education degree	2007	0601 Social affairs - Education	LAU2	2007	PART OF ESPON	LOCAL DATA - 1067
TeDi	E-NS_11	Number of students of higher education institutions	2007	0601 Social affairs - Education	LAU2	2007	PART OF ESPON	LOCAL DATA - 1068
TeDi	I-NS_2	Number of passengers at airports	2007	0302 Transport - flows	LAU2	2007	PART OF ESPON	LOCAL DATA - 1069
TeDi	I-NS_3	Freights handled by airports	2007	0302 Transport - flows	LAU2	2007	PART OF ESPON	LOCAL DATA - 1070
TeDi	I-NS_4	Number of passengers at maritime ports	2007	0302 Transport - flows	LAU2	2007	PART OF ESPON	LOCAL DATA - 1071
TeDi	I-NS_5	Freights handled by maritime ports	2007	0302 Transport - flows	LAU2	2007	PART OF ESPON	LOCAL DATA - 1072
ESPO DB	POP_T	Total population	2006	0201 Demography - Population structure	LAU2	2006	PART OF ESPON	LOCAL DATA - 1073
ESPO DB	M_RO_D	Maximum road distance	2006	0301 Transport - Accessibility	LAU2	2006	PART OF ESPON	LOCAL DATA - 1074
ESPO DB	A_RO_D	Average road distance	2006	0301 Transport - Accessibility	LAU2	2006	PART OF ESPON	LOCAL DATA - 1075
ESPO DB	S_RO_D	Sum road distance	2006	0301 Transport - Accessibility	LAU2	2006	PART OF ESPON	LOCAL DATA - 1076
ESPO DB	M_EC_D	Maximum Euclidean distance	2006	0301 Transport - Accessibility	LAU2	2006	PART OF ESPON	LOCAL DATA - 1077
ESPO DB	A_EC_D	Average Euclidean distance	2006	0301 Transport - Accessibility	LAU2	2006	PART OF ESPON	LOCAL DATA - 1078
ESPO DB	S_EC_D	Sum Euclidean distance	2006	0301 Transport - Accessibility	LAU2	2006	PART OF ESPON	LOCAL DATA - 1079
ESPO DB	M_RNEI	Maximum Road Network Efficiency Indicator	2006	0301 Transport - Accessibility	LAU2	2006	PART OF ESPON	LOCAL DATA - 1080
ESPO DB	A_RNEI	Road Network Efficiency Indicator	2006	0301 Transport - Accessibility	LAU2	2006	PART OF ESPON	LOCAL DATA - 1081
ESPO DB	S_RNEI	The sum of Road Network Efficiency Indicator	2006	0301 Transport - Accessibility	LAU2	2006	PART OF ESPON	LOCAL DATA - 1082
ESPO DB	LONG	Longitude	2006	9902 Non/Cross-thematic data - Geograp	LAU2	2006	PART OF ESPON	LOCAL DATA - 1083
ESPO DB	LAT	Latitude	2006	9902 Non/Cross-thematic data - Geograp	LAU2	2006	PART OF ESPON	LOCAL DATA - 1084
ESPO DB	POP_T	Total population	2001	0201 Demography - Population structure	LAU2	2006	PART OF ESPON	LOCAL DATA - 1085
ESPO DB	POP_T	Total population	2006	0201 Demography - Population structure	LAU2	2006	PART OF ESPON	LOCAL DATA - 1086
ESPO DB	REL_POP_GRO	Relative population growth	2001-2006	0201 Demography - Population structure	LAU2	2006	PART OF ESPON	LOCAL DATA - 1087
ESPO DB	AB_POP_GRO	Absolute population growth	2001-2006	0201 Demography - Population structure	LAU2	2006	PART OF ESPON	LOCAL DATA - 1088
ESPO DB	POP_T	Total population	1990	0201 Demography - Population structure	LAU2	2006	PART OF ESPON	LOCAL DATA - 1089
ESPO DB	POP_T	Total population	1991	0201 Demography - Population structure	LAU2	2006	PART OF ESPON	LOCAL DATA - 1090
ESPO DB	POP_T	Total population	1992	0201 Demography - Population structure	LAU2	2006	PART OF ESPON	LOCAL DATA - 1091
ESPO DB	POP_T	Total population	1993	0201 Demography - Population structure	LAU2	2006	PART OF ESPON	LOCAL DATA - 1092
ESPO DB	POP_T	Total population	1994	0201 Demography - Population structure	LAU2	2006	PART OF ESPON	LOCAL DATA - 1093
ESPO DB	POP_T	Total population	1995	0201 Demography - Population structure	LAU2	2006	PART OF ESPON	LOCAL DATA - 1094



- Local data (23)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
ESPON DB	Live_Birth_pm	Per thousand live Births	1998	0202 Demography - Natural changes	LAU2	2006	PART OF ESPON	LOCAL DATA - 1147
ESPON DB	Live_Birth_pm	Per thousand live Births	1999	0202 Demography - Natural changes	LAU2	2006	PART OF ESPON	LOCAL DATA - 1148
ESPON DB	Live_Birth_pm	Per thousand live Births	2000	0202 Demography - Natural changes	LAU2	2006	PART OF ESPON	LOCAL DATA - 1149
ESPON DB	Live_Birth_pm	Per thousand live Births	2001	0202 Demography - Natural changes	LAU2	2006	PART OF ESPON	LOCAL DATA - 1150
ESPON DB	Live_Birth_pm	Per thousand live Births	2002	0202 Demography - Natural changes	LAU2	2006	PART OF ESPON	LOCAL DATA - 1151
ESPON DB	Live_Birth_pm	Per thousand live Births	2003	0202 Demography - Natural changes	LAU2	2006	PART OF ESPON	LOCAL DATA - 1152
ESPON DB	Live_Birth_pm	Per thousand live Births	2004	0202 Demography - Natural changes	LAU2	2006	PART OF ESPON	LOCAL DATA - 1153
ESPON DB	Live_Birth_pm	Per thousand live Births	2005	0202 Demography - Natural changes	LAU2	2006	PART OF ESPON	LOCAL DATA - 1154
ESPON DB	Live_Birth_pm	Per thousand live Births	2006	0202 Demography - Natural changes	LAU2	2006	PART OF ESPON	LOCAL DATA - 1155
ESPON DB	AREA	LAU2 area	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1156
ESPON DB	C111	Continuous urban fabricIII	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1157
ESPON DB	C112	Discontinuous urban fabric	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1158
ESPON DB	C121	Industrial or commercial units	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1159
ESPON DB	C122	Road and rail networks and associated land Road and rail networks ar	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1160
ESPON DB	C123	Port areas	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1161
ESPON DB	C124	Airports	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1162
ESPON DB	C131	Mineral extraction sitesIII	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1163
ESPON DB	C132	Dump sites	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1164
ESPON DB	C133	Construction sites	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1165
ESPON DB	C141	Green urban areasIII	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1166
ESPON DB	C142	Sport and leisure facilities	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1167
ESPON DB	C211	Non-irrigated arable land	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1168
ESPON DB	C213	Rice fields	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1169
ESPON DB	C221	Non-irrigated arable land	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1170
ESPON DB	C222	Fruit trees and berry plantations	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1171
ESPON DB	C231	Pastures	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1172
ESPON DB	C242	Complex cultivation patterns	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1173
ESPON DB	C243	Land principally occupied by agriculture, with significant areas of natura	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1174
ESPON DB	C311	Broad-leaved forest	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1175
ESPON DB	C312	Coniferous forest	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1176
ESPON DB	C313	Mixed forest	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1177
ESPON DB	C321	Natural grasslands	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1178
ESPON DB	C322	Moors and heathland	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1179
ESPON DB	C324	Transitional woodland-shrub	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1180
ESPON DB	C331	Beaches, dunes, sands Beaches, dunes, sands	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1181
ESPON DB	C332	Bare rocks	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1182
ESPON DB	C333	Sparsely vegetated areas	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1183
ESPON DB	C411	Inland marshes	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1184
ESPON DB	C412	Peat bogs	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1185
ESPON DB	C421	Salt marshes	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1186
ESPON DB	C422	Salines	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1187
ESPON DB	C511	Water courses	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1188
ESPON DB	C512	Water bodies	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1189
ESPON DB	C521	Coastal lagoons	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1190
ESPON DB	C523	Sea and ocean	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1191
ESPON DB	CREL111	Continuous urban fabricIII (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1192
ESPON DB	CREL112	Discontinuous urban fabric (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1193
ESPON DB	CREL121	Industrial or commercial units (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1194
ESPON DB	CREL122	Road and rail networks and associated land Road and rail networks ar	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1195
ESPON DB	CREL123	Port areas (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1196
ESPON DB	CREL124	Airports (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1197
ESPON DB	CREL131	Mineral extraction sitesIII (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1198

- Local data (24)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
ESPON DB	CREL132	Dump sites (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1199
ESPON DB	CREL133	Construction sites (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1200
ESPON DB	CREL141	Green urban areasIII (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1201
ESPON DB	CREL142	Sport and leisure facilities (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1202
ESPON DB	CREL211	Non-irrigated arable land (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1203
ESPON DB	CREL213	Rice fields (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1204
ESPON DB	CREL221	Non-irrigated arable land (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1205
ESPON DB	CREL222	Fruit trees and berry plantations (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1206
ESPON DB	CREL231	Pastures (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1207
ESPON DB	CREL242	Complex cultivation patterns (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1208
ESPON DB	CREL243	Land principally occupied by agriculture, with significant areas of nature	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1209
ESPON DB	CREL311	Broad-leaved forest (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1210
ESPON DB	CREL312	Coniferous forest (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1211
ESPON DB	CREL313	Mixed forest (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1212
ESPON DB	CREL321	Natural grasslands (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1213
ESPON DB	CREL322	Moors and heathland (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1214
ESPON DB	CREL324	Transitional woodland-shrub (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1215
ESPON DB	CREL331	Beaches, dunes, sandsBeaches, dunes, sandsI (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1216
ESPON DB	CREL332	Bare rocks (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1217
ESPON DB	CREL333	Sparsely vegetated areas (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1218
ESPON DB	CREL411	Inland marshes (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1219
ESPON DB	CREL412	Peat bogs (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1220
ESPON DB	CREL421	Salt marshes (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1221
ESPON DB	CREL422	Salines (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1222
ESPON DB	CREL511	Water courses (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1223
ESPON DB	CREL512	Water bodies (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1224
ESPON DB	CREL521	Coastal lagoons (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1225
ESPON DB	CREL523	Sea and ocean (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1226
ESPON DB	AREA	LAU2 area	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1227
ESPON DB	C11	Urban fabric	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1228
ESPON DB	C12	Industrial, commercial and transport units	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1229
ESPON DB	C13	Mine, dump and construction sites	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1230
ESPON DB	C14	Artificial, non-agricultural vegetated areas	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1231
ESPON DB	C21	Arable land	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1232
ESPON DB	C22	Permanent crops	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1233
ESPON DB	C231	Pastures	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1234
ESPON DB	C24	Heterogeneous agricultural areas	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1235
ESPON DB	C31	Forests	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1236
ESPON DB	C32	Scrub and/or herbaceous vegetation associations	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1237
ESPON DB	C33	Open spaces with little or no vegetation	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1238
ESPON DB	C33	Open spaces with little or no vegetation	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1239
ESPON DB	C41	Open spaces with little or no vegetation	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1240
ESPON DB	C42	Open spaces with little or no vegetation	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1241
ESPON DB	C51	Inland waters	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1242
ESPON DB	C52	Marine waters	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1243
ESPON DB	C1	Artificial surfaces	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1244
ESPON DB	C2	Agricultural areas	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1245
ESPON DB	C3	Forest and semi natural areas	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1246
ESPON DB	C4	Wetlands	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1247
ESPON DB	C5	Water bodies	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1248
ESPON DB	CREL11	Urban fabric (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1249
ESPON DB	CREL12	Industrial, commercial and transport units (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1250

- Local data (25)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
ESPON DB	CREL13	Mine, dump and construction sites (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1251
ESPON DB	CREL14	Artificial, non-agricultural vegetated areas (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1252
ESPON DB	CREL21	Arable land (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1253
ESPON DB	CREL22	Permanent crops (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1254
ESPON DB	CREL231	Pastures (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1255
ESPON DB	CREL24	Heterogeneous agricultural areas (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1256
ESPON DB	CREL31	Forests (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1257
ESPON DB	CREL32	Scrub and/or herbaceous vegetation associations (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1258
ESPON DB	CREL33	Open spaces with little or no vegetation (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1259
ESPON DB	CREL33	Open spaces with little or no vegetation (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1260
ESPON DB	CREL41	Open spaces with little or no vegetation (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1261
ESPON DB	CREL42	Open spaces with little or no vegetation (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1262
ESPON DB	CREL51	Inland waters (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1263
ESPON DB	CREL52	Marine waters (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1264
ESPON DB	CREL1	Artificial surfaces (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1265
ESPON DB	CREL2	Agricultural areas (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1266
ESPON DB	CREL3	Forest and semi natural areas (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1267
ESPON DB	CREL4	Wetlands (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1268
ESPON DB	CREL5	Water bodies (%)	2006	0501 Land use - Land Use and land cover t	LAU2	2006	PART OF ESPON	LOCAL DATA - 1269
ESPON DB	GDP06LAU2	Local economic performance	2006	0701 Economy - Aggregated accounts	LAU2	2006	ESPON AREA	LOCAL DATA - 1270
ESPON DB	DENS_EURO	Density of estimated GDP06LAU2	2006	0701 Economy - Aggregated accounts	LAU2	2006	ESPON AREA	LOCAL DATA - 1271







- Neighborhood data (3)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
ESPON DB	unemp	Unemployment	2003	0702 Economy - Employment	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 105
ESPON DB	unemp	Unemployment	2004	0702 Economy - Employment	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 106
ESPON DB	unemp	Unemployment	2005	0702 Economy - Employment	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 107
ESPON DB	unemp	Unemployment	2006	0702 Economy - Employment	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 108
ESPON DB	unemp	Unemployment	2007	0702 Economy - Employment	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 109
ESPON DB	Net_migration_rate	Net migration rate	2000	0204 Demography - Migrations	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 110
ESPON DB	Net_migration_rate	Net migration rate	2001	0204 Demography - Migrations	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 111
ESPON DB	Net_migration_rate	Net migration rate	2002	0204 Demography - Migrations	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 112
ESPON DB	Net_migration_rate	Net migration rate	2003	0204 Demography - Migrations	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 113
ESPON DB	Net_migration_rate	Net migration rate	2004	0204 Demography - Migrations	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 114
ESPON DB	Net_migration_rate	Net migration rate	2005	0204 Demography - Migrations	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 115
ESPON DB	Net_migration_rate	Net migration rate	2006	0204 Demography - Migrations	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 116
ESPON DB	Net_migration_rate	Net migration rate	2007	0204 Demography - Migrations	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 117
ESPON DB	Pop_m	population males	2000	0201 Demography - Population structure	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 118
ESPON DB	Pop_m	population males	2001	0201 Demography - Population structure	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 119
ESPON DB	Pop_m	population males	2002	0201 Demography - Population structure	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 120
ESPON DB	Pop_m	population males	2003	0201 Demography - Population structure	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 121
ESPON DB	Pop_m	population males	2004	0201 Demography - Population structure	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 122
ESPON DB	Pop_m	population males	2005	0201 Demography - Population structure	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 123
ESPON DB	Pop_m	population males	2006	0201 Demography - Population structure	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 124
ESPON DB	Pop_m	population males	2007	0201 Demography - Population structure	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 125
ESPON DB	Pop_f	population females	2000	0201 Demography - Population structure	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 126
ESPON DB	Pop_f	population females	2001	0201 Demography - Population structure	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 127
ESPON DB	Pop_f	population females	2002	0201 Demography - Population structure	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 128
ESPON DB	Pop_f	population females	2003	0201 Demography - Population structure	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 129
ESPON DB	Pop_f	population females	2004	0201 Demography - Population structure	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 130
ESPON DB	Pop_f	population females	2005	0201 Demography - Population structure	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 131
ESPON DB	Pop_f	population females	2006	0201 Demography - Population structure	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 132
ESPON DB	Pop_f	population females	2007	0201 Demography - Population structure	SNUTS	2010	Western Balkans & Turkey	NEIGHB. - 133

- Urban data (1)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
FOCI	pop_t	Total population	2000	0201 Demography - Population structure	FUA	2010	ESPON	URBAN DATA - 1
FOCI	pop_t	population of cities	1986-2006	0201 Demography - Population structure	FUA	2010	ESPON	URBAN DATA - 2
FOCI	gdp_cap	Total population	2001	0201 Demography - Population structure	FUA	2010	ESPON	URBAN DATA - 3
FOCI	Dem_Sit_MetroReg_Pop_Ch	Demographic situation in metropolis-region context - population chan	2000-2005	0201 Demography - Population structure	MEGA	2010	ESPON	URBAN DATA - 4
FOCI	pop_t	Total population	1995	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 5
FOCI	pop_t	Total population	1996	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 6
FOCI	pop_t	Total population	1997	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 7
FOCI	pop_t	Total population	1998	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 8
FOCI	pop_t	Total population	1999	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 9
FOCI	pop_t	Total population	2000	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 10
FOCI	pop_t	Total population	2001	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 11
FOCI	pop_t	Total population	2002	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 12
FOCI	pop_t	Total population	2003	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 13
FOCI	pop_t	Total population	2004	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 14
FOCI	pop_t	Total population	2005	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 15
FOCI	pop_t	Total population	2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 16
FOCI	pop_t	Total population	2007	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 17
FOCI	pop_t	Total population	2005	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 18
FOCI	pop_luz_rt	ratio between LUZ population and NUTS2 population	2005	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 19
FOCI	pop_luz_t	Population of the LUZ	2004-2006	0201 Demography - Population structure	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 20
FOCI	pop_luz_t	Population of the LUZ	1999-2001	0201 Demography - Population structure	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 21
FOCI	pop_core_t	population of the core city	2004-2006	0201 Demography - Population structure	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 22
FOCI	pop_core_t	population of the core city	1999-2001	0201 Demography - Population structure	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 23
FOCI	pop_per_t	Population of the suburbs (LUZ minus core)	2004	0201 Demography - Population structure	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 24
FOCI	pop_per_t	Population of the suburbs (LUZ minus core)	2000	0201 Demography - Population structure	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 25
FOCI	pop_grow_core	Annual growth of the population of the core city	2000-2005	0201 Demography - Population structure	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 26
FOCI	pop_grow_per	Annual growth of the population of the suburbs	2000-2005	0201 Demography - Population structure	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 27
FOCI	pop_grow_diff	Difference between annual growth of the suburbs and the core	2000-2005	0201 Demography - Population structure	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 28
FOCI	pop_core_t	population of the core city	2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 29
FOCI	pop_core_nat	Nationals in the core city	2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 30
FOCI	pop_core_Eunat	EU national in the core city	2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 31
FOCI	pop_core_nonEunat	Non EU national in the core city	2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 32
FOCI	pop_luz_t	Population of the LUZ	2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 33
FOCI	pop_luz_nat	Nationals in the LUZ	2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 34
FOCI	pop_luz_Eunat	EU nationals in the LUZ	2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 35
FOCI	pop_luz_nonEunat	Non EU national in the LUZ	2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 36
FOCI	pop_luz_nat_sh	Share of nationals in the LUZ	2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 37
FOCI	pop_luz_Eunat_sh	Share of EU nationals in the LUZ	2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 38
FOCI	luznoneup	Share of non EU nationals in the LUZ	2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 39
FOCI	pop_core_nat_sh	Share of national in the core city	2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 40
FOCI	pop_core_Eunat_sh	Share of EU nationals in the core city	2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 41
FOCI	pop_luz_nonEunat_sh	Share of Non EU national in the core city	2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 42
FOCI	pop_luzcore_nonEunat_sh	Ratio between non EU national in the core city and in the LUZ	2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 43
FOCI	pop_luz_t	Population of the LUZ	1990	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 44
FOCI	pop_luz_t	Population of the LUZ	1991	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 45
FOCI	pop_luz_t	Population of the LUZ	1992	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 46
FOCI	pop_luz_t	Population of the LUZ	1993	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 47
FOCI	pop_luz_t	Population of the LUZ	1994	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 48
FOCI	pop_luz_t	Population of the LUZ	1995	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 49
FOCI	pop_luz t	Population of the LUZ	1996	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 50

- Urban data (2)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
FOCI	pop_luz_t	Population of the LUZ	1997	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 51
FOCI	pop_luz_2_t	Population of the LUZ	1998	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 52
FOCI	pop_luz_3_t	Population of the LUZ	1999	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 53
FOCI	pop_luz_4_t	Population of the LUZ	2000	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 54
FOCI	pop_luz_5_t	Population of the LUZ	2001	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 55
FOCI	pop_luz_6_t	Population of the LUZ	2002	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 56
FOCI	pop_luz_7_t	Population of the LUZ	2003	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 57
FOCI	pop_luz_8_t	Population of the LUZ	2004	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 58
FOCI	pop_luz_9_t	Population of the LUZ	2005	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 59
FOCI	pop_luz_10_t	Population of the LUZ	2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 60
FOCI	pop_luz_11_t	Population of the LUZ	2007	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 61
FOCI	pop_grow_luz	Population growth	1990-2007	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 62
FOCI	pop_grow_luz	Population growth	1995-2007	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 63
FOCI	pop_grow_luz	Population growth	2001-2007	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 64
FOCI	pop_luz_2_t	Population of the LUZ	2004-2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 65
FOCI	pop_luz_3_t	Population of the LUZ	1999-2003	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 66
FOCI	pop_luz_4_t	Population of the LUZ	1994-1999	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 67
FOCI	pop_luz_5_t	Population of the LUZ	1989-1994	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 68
FOCI	pop_grow_luz	Population growth	1991-2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 69
FOCI	pop_grow_luz	Population growth	2000-2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 70
FOCI	pop_res_t	Total Resident Population	2004-2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 71
FOCI	pop_res_t	Total Resident Population	1999-2002	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 72
FOCI	pop_luz_2_t	Population of the LUZ	2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 73
FOCI	Dem_Sit_MetroReg_Nat_Incr	Demographic situation in metropolis-region context - natural increase	2000-2005	0202 Demography - Natural changes	MEGA	2010	ESPON	URBAN DATA - 74
FOCI	household_overcrow_rt	Share of households with more than 1 person per room	2004-2006	0203 Demography - Households	LUZ	2008	ESPON	URBAN DATA - 75
FOCI	household_overcrow_rt	Share of households with more than 1 person per room	1999-2002	0203 Demography - Households	LUZ	2008	ESPON	URBAN DATA - 76
FOCI	Dem_Sit_MetroReg_Mig_Bal	Demographic situation in metropolis-region context - migration balance	2000-2005	0204 Demography - Migrations	MEGA	2010	ESPON	URBAN DATA - 77
FOCI	typo_luz_demo	Typology of the internal demographic dynamics core versus suburbs	2000-2005	0205 Demography - Territorial impact	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 78
FOCI	Nb_MEGAs_reachable_by_return	Number of MEGAs reachable by rail return trips	2009	0301 Transport - Accessibility	MEGA	2010	ESPON	URBAN DATA - 79
FOCI	Nb_MEGAs_reachable_by_return	Number of MEGAs reachable by air return trips	2009	0301 Transport - Accessibility	MEGA	2010	ESPON	URBAN DATA - 80
FOCI	Nb_MEGAs_reachable_by_return	Number of MEGAs reachable by intermodal rail plus air return trips	2009	0301 Transport - Accessibility	MEGA	2010	ESPON	URBAN DATA - 81
FOCI	total_nb_MEGAs_reachable_by_return	Number of MEGAs reachable by rail, air and intermodal return trips	2009	0301 Transport - Accessibility	MEGA	2010	ESPON	URBAN DATA - 82
FOCI	tt1010v	Percentage of journeys to work by public transport (rail, metro, bus, tram)	1999-2002	0302 Transport - Flows	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 83
FOCI	FUAs_Commuting_out	Commuting out from FUAs	2001	0302 Transport - flows	FUA	2010	ESPON	URBAN DATA - 84
FOCI	FUAs_Commuting_in	Commuting in FUAs	2001	0302 Transport - flows	FUA	2010	ESPON	URBAN DATA - 85
FOCI	arable	Arable land and permanent crops lost	1990-2000	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 86
FOCI	pastures	Pastures and mosaic farmland lost	1990-2000	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 87
FOCI	forest	Forests and transitional woodland shrub lost	1990-2000	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 88
FOCI	grassland	Natural grassland, heathland, sclerophyllous vegetation lost	1990-2000	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 89
FOCI	open	Open space with little or no vegetation lost	1990-2000	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 90
FOCI	sealing	Sealed area per inhabitant in European cities	2006	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 91
FOCI	sealing_cc	Degree of soil sealing in core city	2006	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 92
FOCI	sealing_luz	Degree of soil sealing in LUZ	2006	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 93
FOCI	built_cc	Percentage of built-up area in the core city	1990	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 94
FOCI	built_cc	Percentage of built-up area in the core city	2000	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 95
FOCI	built_luz	Percentage of built-up area in Large Urban Zone	1990	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 96
FOCI	built_luz	Percentage of built-up area in Large Urban Zone	2000	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 97
FOCI	built_ch_cc	Percentage of new of built-up area over total built-up area at the beginning	1990-2000	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 98
FOCI	built_ch_cc	Percentage of new of built-up area over total built-up area at the beginning	2000-2006	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 99
FOCI	built_ch_luz	Percentage of new of built-up area over total built-up area at the beginning	1990-2000	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 100
FOCI	built_ch_luz	Percentage of new of built-up area over total built-up area at the beginning	2000-2006	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 101
FOCI	lt_cc	Land take per capita in core city	1990-2000	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 102

- Urban data (3)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
FOCI	lt_cc	Land take per capita in core city	2000-2006	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 103
FOCI	lt_luz	Land take per capita in LUZ	1990-2000	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 104
FOCI	lt_luz	Land take per capita in LUZ	2000-2006	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 105
FOCI	redevelop_cc	Degree of redevelopment in the core city.	1990-2000	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 106
FOCI	redevelop_cc	Degree of redevelopment in the core city.	2000-2006	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 107
FOCI	redevelop_luz	Degree of redevelopment in the large urban zone.	1990-2000	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 108
FOCI	redevelop_luz	Degree of redevelopment in the large urban zone.	2000-2006	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 109
FOCI	residential_cc	Growth rate of residential areas in the core city.	1990-2000	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 110
FOCI	residential_cc	Growth rate of residential areas in the core city.	2000-2006	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 111
FOCI	residential_luz	Growth rate of residential areas in the large urban zone.	1990-2000	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 112
FOCI	residential_luz	Growth rate of residential areas in the large urban zone.	2000-2006	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 113
FOCI	ind_cc	Growth rate of industrial, commercial and transport areas in the core city.	1990-2000	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 114
FOCI	ind_cc	Growth rate of industrial, commercial and transport areas in the core city.	2000-2006	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 115
FOCI	ind_luz	Growth rate of industrial, commercial and transport areas in the large urban zone.	1990-2000	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 116
FOCI	ind_luz	Growth rate of industrial, commercial and transport areas in the large urban zone.	2000-2006	0502 Land Use - Urban land use attributes	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 117
FOCI	edu_stud1_rt	Students not completing their compulsory education	2004-2006	0601 Social affairs - Education	LUZ	2008	ESPON	URBAN DATA - 118
FOCI	act_edu1-2_rt	Working age population qualified at level 1 or 2 ISCED	2004-2006	0601 Social affairs - Education	LUZ	2008	ESPON	URBAN DATA - 119
FOCI	act_edu3-4_rt	Working age population qualified at level 3 or 4 ISCED	2004-2006	0601 Social affairs - Education	LUZ	2008	ESPON	URBAN DATA - 120
FOCI	act_edu5-6_rt	Working age population qualified at level 5 or 6 ISCED	2004-2006	0601 Social affairs - Education	LUZ	2008	ESPON	URBAN DATA - 121
FOCI	edu_stud1_rt	Students not completing their compulsory education	1999-2002	0601 Social affairs - Education	LUZ	2008	ESPON	URBAN DATA - 122
FOCI	act_edu1-2_rt	Working age population qualified at level 1 or 2 ISCED	1999-2002	0601 Social affairs - Education	LUZ	2008	ESPON	URBAN DATA - 123
FOCI	act_edu3-4_rt	Working age population qualified at level 3 or 4 ISCED	1999-2002	0601 Social affairs - Education	LUZ	2008	ESPON	URBAN DATA - 124
FOCI	act_edu5-6_rt	Working age population qualified at level 5 or 6 ISCED	1999-2002	0601 Social affairs - Education	LUZ	2008	ESPON	URBAN DATA - 125
FOCI	act_edu1_rt	Share of active population with less or equal to low secondary school diploma	2000	0601 Social affairs - Education	LUZ	2008	ESPON	URBAN DATA - 126
FOCI	act_edu2_rt	Share of active population with high secondary school diploma	2000	0601 Social affairs - Education	LUZ	2008	ESPON	URBAN DATA - 127
FOCI	act_edu3_rt	Share of active population with tertiary diploma	2000	0601 Social affairs - Education	LUZ	2008	ESPON	URBAN DATA - 128
FOCI	act_edu1_rt	Share of active population with less or equal to low secondary school diploma	2005	0601 Social affairs - Education	LUZ	2008	ESPON	URBAN DATA - 129
FOCI	act_edu2_rt	Share of active population with high secondary school diploma	2005	0601 Social affairs - Education	LUZ	2008	ESPON	URBAN DATA - 130
FOCI	act_edu3_rt	Share of active population with tertiary diploma	2005	0601 Social affairs - Education	LUZ	2008	ESPON	URBAN DATA - 131
FOCI	earn_1-4_rt	Ratio of first to fourth quintile earnings	2004-2006	0602 Social affairs - Labour market	LUZ	2008	ESPON	URBAN DATA - 132
FOCI	inc_60lessmed_rt	Households with less than 60% of the national median annual disposable income	2004-2006	0602 Social affairs - Labour market	LUZ	2008	ESPON	URBAN DATA - 133
FOCI	inc_50lessav_rt	Households receiving less than half of the national average household income	2004-2006	0602 Social affairs - Labour market	LUZ	2008	ESPON	URBAN DATA - 134
FOCI	earn_1-4_rt	Ratio of first to fourth quintile earnings	1999-2002	0602 Social affairs - Labour market	LUZ	2008	ESPON	URBAN DATA - 135
FOCI	inc_60lessmed_rt	Households with less than 60% of the national median annual disposable income	1999-2002	0602 Social affairs - Labour market	LUZ	2008	ESPON	URBAN DATA - 136
FOCI	inc_50lessav_rt	Households receiving less than half of the national average household income	1999-2002	0602 Social affairs - Labour market	LUZ	2008	ESPON	URBAN DATA - 137
FOCI	socsec_household_rt	Households reliant upon social security	2004-2006	0603 Social affairs - Living conditions	LUZ	2008	ESPON	URBAN DATA - 138
FOCI	socsec_indiv_rt	Individuals reliant on social security	2004-2006	0603 Social affairs - Living conditions	LUZ	2008	ESPON	URBAN DATA - 139
FOCI	dwel_bad_rt	Dwellings lacking basic amenities	2004-2006	0603 Social affairs - Living conditions	LUZ	2008	ESPON	URBAN DATA - 140
FOCI	inf_mort_rt	Infant Mortality rate	2004-2006	0603 Social affairs - Living conditions	LUZ	2008	ESPON	URBAN DATA - 141
FOCI	violent_death_rt	Murders and violent deaths / average population	2004-2006	0603 Social affairs - Living conditions	LUZ	2008	ESPON	URBAN DATA - 142
FOCI	car_thefts_rt	Number of car thefts / average population	2004-2006	0603 Social affairs - Living conditions	LUZ	2008	ESPON	URBAN DATA - 143
FOCI	victim_rt	Incidence rate of victimisation	2004-2006	0603 Social affairs - Living conditions	LUZ	2008	ESPON	URBAN DATA - 144
FOCI	socsec_household_rt	Households reliant upon social security	1999-2002	0603 Social affairs - Living conditions	LUZ	2008	ESPON	URBAN DATA - 145
FOCI	socsec_indiv_rt	Individuals reliant on social security	1999-2002	0603 Social affairs - Living conditions	LUZ	2008	ESPON	URBAN DATA - 146
FOCI	dwel_bad_rt	Dwellings lacking basic amenities	1999-2002	0603 Social affairs - Living conditions	LUZ	2008	ESPON	URBAN DATA - 147
FOCI	inf_mort_rt	Infant Mortality rate	1999-2002	0603 Social affairs - Living conditions	LUZ	2008	ESPON	URBAN DATA - 148
FOCI	violent_death_rt	Murders and violent deaths / average population	1999-2002	0603 Social affairs - Living conditions	LUZ	2008	ESPON	URBAN DATA - 149
FOCI	car_thefts_rt	Number of car thefts / average population	1999-2002	0603 Social affairs - Living conditions	LUZ	2008	ESPON	URBAN DATA - 150
FOCI	victim_rt	Incidence rate of victimisation	1999-2002	0603 Social affairs - Living conditions	LUZ	2008	ESPON	URBAN DATA - 151
FOCI	Synt_indic	synthetic indicator of satisfaction	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 152
FOCI	PS1010	Satisfied with public transport (synthetic index 0-100)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 153
FOCI	PS1012	public transport: very satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 154

- Urban data (4)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
FDCI	PS1013	public transport: rather satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 155
FDCI	PS1014	public transport: rather unsatisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 156
FDCI	PS1015	public transport: not at all satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 157
FDCI	PS1016	public transport: no reply	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 158
FDCI	PS1017	public transport: satisfied (rather+strong)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 159
FDCI	PS1018	public transport: unsatisfied (rather+strong)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 160
FDCI	PS1020	Satisfied with schools (synthetic index 0-100)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 161
FDCI	PS1022	schools: very satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 162
FDCI	PS1023	schools: rather satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 163
FDCI	PS1024	schools: rather unsatisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 164
FDCI	PS1025	schools: not at all satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 165
FDCI	PS1026	schools: no reply	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 166
FDCI	PS1027	schools: satisfied (rather+strong)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 167
FDCI	PS1028	schools: unsatisfied (rather+strong)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 168
FDCI	PS1030	Satisfied with hospitals (synthetic index 0-100)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 169
FDCI	PS1032	hospitals: very satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 170
FDCI	PS1033	hospitals: rather satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 171
FDCI	PS1034	hospitals: rather unsatisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 172
FDCI	PS1035	hospitals: not at all satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 173
FDCI	PS1036	hospitals: no reply	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 174
FDCI	PS1037	hospitals: satisfied (rather+strong)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 175
FDCI	PS1038	hospitals: unsatisfied (rather+strong)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 176
FDCI	PS1040	Satisfied with doctors (synthetic index 0-100)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 177
FDCI	PS1042	doctors: very satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 178
FDCI	PS1043	doctors: rather satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 179
FDCI	PS1044	doctors: rather unsatisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 180
FDCI	PS1045	doctors: not at all satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 181
FDCI	PS1046	doctors: no reply	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 182
FDCI	PS1047	doctors: satisfied (rather+strong)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 183
FDCI	PS1048	doctors: unsatisfied (rather+strong)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 184
FDCI	PS1050	Satisfied with green space (synthetic index 0-100)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 185
FDCI	PS1052	greenspace: very satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 186
FDCI	PS1053	greenspace: rather satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 187
FDCI	PS1054	greenspace: rather unsatisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 188
FDCI	PS1055	greenspace: not at all satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 189
FDCI	PS1056	greenspace: no reply	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 190
FDCI	PS1057	greenspace: satisfied (rather+strong)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 191
FDCI	PS1058	greenspace: unsatisfied (rather+strong)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 192
FDCI	PS1060	Satisfied with sport facilities (synthetic index 0-100)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 193
FDCI	PS1062	sportfacilities: very satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 194
FDCI	PS1063	sportfacilities: rather satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 195
FDCI	PS1064	sportfacilities: rather unsatisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 196
FDCI	PS1065	sportfacilities: not at all satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 197
FDCI	PS1066	sportfacilities: no reply	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 198
FDCI	PS1067	sportfacilities: satisfied (rather+strong)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 199
FDCI	PS1068	sportfacilities: unsatisfied (rather+strong)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 200
FDCI	PS1070	Satisfied with cinemas (synthetic index 0-100)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 201
FDCI	PS1072	cinemas: very satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 202
FDCI	PS1073	cinemas: rather satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 203
FDCI	PS1074	cinemas: rather unsatisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 204
FDCI	PS1075	cinemas: not at all satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 205
FDCI	PS1076	cinemas: no reply	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 206

- Urban data (5)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
FOCI	PS1077	cinemas: satisfied (rather+strong)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 207
FOCI	PS1078	cinemas: unsatisfied (rather+strong)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 208
FOCI	PS1080	Satisfied with cultural facilities (synthetic index 0-100)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 209
FOCI	PS1082	cultural facilities: very satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 210
FOCI	PS1083	cultural facilities: rather satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 211
FOCI	PS1084	cultural facilities: rather unsatisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 212
FOCI	PS1085	cultural facilities: not at all satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 213
FOCI	PS1086	cultural facilities: no reply	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 214
FOCI	PS1087	cultural facilities: satisfied (rather+strong)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 215
FOCI	PS1088	cultural facilities: unsatisfied (rather+strong)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 216
FOCI	PS1090	Satisfied with public internet access (synthetic index 0-100)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 217
FOCI	PS1092	public-internet: very satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 218
FOCI	PS1093	public-internet: rather satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 219
FOCI	PS1094	public-internet: rather unsatisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 220
FOCI	PS1095	public-internet: not at all satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 221
FOCI	PS1096	public-internet: no reply	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 222
FOCI	PS1097	public-internet: satisfied (rather+strong)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 223
FOCI	PS1098	public-internet: unsatisfied (rather+strong)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 224
FOCI	PS1100	Satisfied with internet access at home (synthetic index 0-100)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 225
FOCI	PS1102	internet access at home: very satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 226
FOCI	PS1103	internet access at home: rather satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 227
FOCI	PS1104	internet access at home: rather unsatisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 228
FOCI	PS1105	internet access at home: not at all satisfied	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 229
FOCI	PS1106	internet access at home: no reply	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 230
FOCI	PS1107	internet access at home: satisfied (rather+strong)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 231
FOCI	PS1108	internet access at home: unsatisfied (rather+strong)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 232
FOCI	PS2010	It is easy to find a good job here (synthetic index 0-100)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 233
FOCI	PS2012	easy-to-find-a-job: strongly agree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 234
FOCI	PS2013	easy-to-find-a-job: somewhat agree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 235
FOCI	PS2014	easy-to-find-a-job: somewhat disagree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 236
FOCI	PS2015	easy-to-find-a-job: strongly disagree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 237
FOCI	PS2016	easy-to-find-a-job: no reply	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 238
FOCI	PS2017	easy-to-find-a-job: agree (strongly+somewhat)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 239
FOCI	PS2018	easy-to-find-a-job: disagree (strongly+somewhat)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 240
FOCI	PS2020	Foreigner here are well integrated (synthetic index 0-100)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 241
FOCI	PS2022	integration of foreigners: strongly agree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 242
FOCI	PS2023	integration of foreigners: somewhat agree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 243
FOCI	PS2024	integration of foreigners: somewhat disagree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 244
FOCI	PS2025	integration of foreigners: strongly disagree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 245
FOCI	PS2026	integration of foreigners: no reply	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 246
FOCI	PS2027	integration of foreigners: agree (strongly+somewhat)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 247
FOCI	PS2028	integration of foreigners: disagree (strongly+somewhat)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 248
FOCI	PS2030	Easy to find good housing at reasonable price (synthetic index 0-100)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 249
FOCI	PS2032	easy-to-find-good-housing: strongly agree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 250
FOCI	PS2033	easy-to-find-good-housing: somewhat agree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 251
FOCI	PS2034	easy-to-find-good-housing: somewhat disagree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 252
FOCI	PS2035	easy-to-find-good-housing: strongly disagree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 253
FOCI	PS2036	easy-to-find-good-housing: no reply	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 254
FOCI	PS2037	easy-to-find-good-housing: agree (strongly+somewhat)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 255
FOCI	PS2038	easy-to-find-good-housing: disagree (strongly+somewhat)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 256
FOCI	PS2040	Administrative services help efficiently (synthetic index 0-100)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 257
FOCI	PS2042	administration-helpful: strongly agree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 258

- Urban data (6)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
FOCI	PS2043	administration-helpful: somewhat agree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 259
FOCI	PS2044	administration-helpful: somewhat disagree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 260
FOCI	PS2045	administration-helpful: strongly disagree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 261
FOCI	PS2046	administration-helpful: no reply	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 262
FOCI	PS2047	administration-helpful: agree (strongly+somewhat)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 263
FOCI	PS2048	administration-helpful: disagree (strongly+somewhat)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 264
FOCI	PS2050	Air pollution is a big problem here (synthetic index 0-100)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 265
FOCI	PS2052	Air pollution is a big problem here (synthetic index 0-100)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 266
FOCI	PS2053	pollution-is-a-problem: somewhat agree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 267
FOCI	PS2054	pollution-is-a-problem: somewhat disagree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 268
FOCI	PS2055	pollution-is-a-problem: strongly disagree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 269
FOCI	PS2056	pollution-is-a-problem: no reply	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 270
FOCI	PS2057	pollution-is-a-problem: agree (strongly+somewhat)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 271
FOCI	PS2058	pollution-is-a-problem: disagree (strongly+somewhat)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 272
FOCI	PS2060	Noise is a big problem here (synthetic index 0-100)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 273
FOCI	PS2062	noise-is-a-problem: strongly agree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 274
FOCI	PS2063	noise-is-a-problem: somewhat agree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 275
FOCI	PS2064	noise-is-a-problem: somewhat disagree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 276
FOCI	PS2065	noise-is-a-problem: strongly disagree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 277
FOCI	PS2066	noise-is-a-problem: no reply	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 278
FOCI	PS2067	noise-is-a-problem: agree (strongly+somewhat)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 279
FOCI	PS2068	noise-is-a-problem: disagree (strongly+somewhat)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 280
FOCI	PS2070	This is a clean city (synthetic index 0-100)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 281
FOCI	PS2072	clean-city: strongly agree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 282
FOCI	PS2073	clean-city: somewhat agree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 283
FOCI	PS2074	clean-city: somewhat disagree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 284
FOCI	PS2075	clean-city: strongly disagree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 285
FOCI	PS2076	clean-city: no reply	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 286
FOCI	PS2077	clean-city: agree (strongly+somewhat)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 287
FOCI	PS2078	clean-city: disagree (strongly+somewhat)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 288
FOCI	PS2080	Resources are spent in a responsible way (synthetic index 0-100)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 289
FOCI	PS2082	resources: strongly agree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 290
FOCI	PS2083	resources: somewhat agree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 291
FOCI	PS2084	resources: somewhat disagree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 292
FOCI	PS2085	resources: strongly disagree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 293
FOCI	PS2086	resources: no reply	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 294
FOCI	PS2087	resources: agree (strongly+somewhat)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 295
FOCI	PS2088	resources: disagree (strongly+somewhat)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 296
FOCI	PS2090	Satisfied to live in this city (synthetic index 0-100)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 297
FOCI	PS2092	overall-satisfied: strongly agree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 298
FOCI	PS2093	overall-satisfied: somewhat agree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 299
FOCI	PS2094	overall-satisfied: somewhat disagree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 300
FOCI	PS2095	overall-satisfied: strongly disagree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 301
FOCI	PS2096	overall-satisfied: no reply	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 302
FOCI	PS2097	overall-satisfied: agree (strongly+somewhat)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 303
FOCI	PS2098	overall-satisfied: disagree (strongly+somewhat)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 304
FOCI	PS2100	In 5 years, it will be more pleasant to live here (synthetic index 0-100)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 305
FOCI	PS2102	in five years it will be better: strongly agree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 306
FOCI	PS2103	in five years it will be better: somewhat agree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 307
FOCI	PS2104	in five years it will be better: somewhat disagree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 308
FOCI	PS2105	in five years it will be better: strongly disagree	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 309
FOCI	PS2106	in five years it will be better: no reply	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 310



- Urban data (7)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
FOCI	PS2107	in five years it will be better: agree (strongly+somewhat)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 311
FOCI	PS2108	in five years it will be better: disagree (strongly+somewhat)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 312
FOCI	PS3010	Difficulty paying the bills at the end of the month (synthetic index 0-100)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 313
FOCI	PS3012	it is difficult to pay my bills: always	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 314
FOCI	PS3013	it is difficult to pay my bills: sometimes	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 315
FOCI	PS3014	it is difficult to pay my bills: rarely or never	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 316
FOCI	PS3015	it is difficult to pay my bills: no reply	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 317
FOCI	PS3020	Feel safe in this neighbourhood (synthetic index 0-100)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 318
FOCI	PS3022	the neighborhood is safe: always	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 319
FOCI	PS3023	the neighborhood is safe: sometimes	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 320
FOCI	PS3024	the neighborhood is safe: rarely or never	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 321
FOCI	PS3025	the neighborhood is safe: no reply	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 322
FOCI	PS3030	Feel safe in this city (synthetic index 0-100)	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 323
FOCI	PS3032	the city is safe: always	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 324
FOCI	PS3033	the city is safe: sometimes	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 325
FOCI	PS3034	the city is safe: rarely or never	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 326
FOCI	PS3035	the city is safe: no reply	2009	0603 Social affairs - Living conditions	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 327
FOCI	homicide_rt	Homicides/average population	1999-2001	0603 Social affairs - Living conditions	LUZ	2008	ESPON	URBAN DATA - 328
FOCI	suicide_rt	Suicide/average population	1999-2001	0603 Social affairs - Living conditions	LUZ	2008	ESPON	URBAN DATA - 329
FOCI	gdp_t	Total GDP	2007	0701 Economy - aggregated accounts	FUA	2010	ESPON	URBAN DATA - 330
FOCI	gdp_cap	Average GDP per capita	2007	0701 Economy - aggregated accounts	FUA	2010	ESPON	URBAN DATA - 331
FOCI	gdp_cap	GDP PPS per capita	2007	0701 Economy - aggregated accounts	FUA	2010	ESPON	URBAN DATA - 332
FOCI	GDP_cap_ch	GDP PPS per capita change	2001	0701 Economy - aggregated accounts	FUA	2010	ESPON	URBAN DATA - 333
FOCI	Dis_Dev_Metro_Reg	Disparities in the development level between the metropolis and its reg	1995	0701 Economy - Aggregated accounts	MEGA	2010	ESPON	URBAN DATA - 334
FOCI	Dis_Dev_Metro_Reg	Disparities in the development level between the metropolis and its reg	2004	0701 Economy - Aggregated accounts	MEGA	2010	ESPON	URBAN DATA - 335
FOCI	GROW_Dis_Dev_Metro_Reg	Change of disparities in the development level between the metropolis	1995-2004	0701 Economy - Aggregated accounts	MEGA	2010	ESPON	URBAN DATA - 336
FOCI	TYP_Macro_Reg_GDP_per_CP	Typology of macroregions based on GDP per capita ratio level and ch	1995-2004	0701 Economy - Aggregated accounts	MEGA	2010	ESPON	URBAN DATA - 337
FOCI	TYP_Macro_Reg_GDP_GROW	Typology of macroregions based on GDP growth 1995-2004 - real gro	1995-2004	0701 Economy - Aggregated accounts	MEGA	2010	ESPON	URBAN DATA - 338
FOCI	TYP_Macro_Reg_GDP_GROW	Typology of macroregions based on GDP growth 1995-2004 - each co	1995-2004	0701 Economy - Aggregated accounts	MEGA	2010	ESPON	URBAN DATA - 339
FOCI	gdp_pps_t	GDP in PPS	1995	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 340
FOCI	gdp_pps_t	GDP in PPS	1996	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 341
FOCI	gdp_pps_t	GDP in PPS	1997	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 342
FOCI	gdp_pps_t	GDP in PPS	1998	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 343
FOCI	gdp_pps_t	GDP in PPS	1999	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 344
FOCI	gdp_pps_t	GDP in PPS	2000	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 345
FOCI	gdp_pps_t	GDP in PPS	2001	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 346
FOCI	gdp_pps_t	GDP in PPS	2002	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 347
FOCI	gdp_pps_t	GDP in PPS	2003	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 348
FOCI	gdp_pps_t	GDP in PPS	2004	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 349
FOCI	gdp_pps_t	GDP in PPS	2005	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 350
FOCI	gdp_pps_t	GDP in PPS	2006	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 351
FOCI	va_t	Total Added value	2006	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 352
FOCI	gdp_cap_pps_ix	GDP per inhabitant in PPS according to EU average	1995	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 353
FOCI	gdp_cap_pps_ix	GDP per inhabitant in PPS according to EU average	1996	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 354
FOCI	gdp_cap_pps_ix	GDP per inhabitant in PPS according to EU average	1997	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 355
FOCI	gdp_cap_pps_ix	GDP per inhabitant in PPS according to EU average	1998	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 356
FOCI	gdp_cap_pps_ix	GDP per inhabitant in PPS according to EU average	1999	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 357
FOCI	gdp_cap_pps_ix	GDP per inhabitant in PPS according to EU average	2000	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 358
FOCI	gdp_cap_pps_ix	GDP per inhabitant in PPS according to EU average	2001	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 359
FOCI	gdp_cap_pps_ix	GDP per inhabitant in PPS according to EU average	2002	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 360
FOCI	gdp_cap_pps_ix	GDP per inhabitant in PPS according to EU average	2003	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 361
FOCI	gdp_cap_pps_ix	GDP per inhabitant in PPS according to EU average	2004	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 362

- Urban data (8)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
FOCI	gdp_cap_pps_ix	GDP per inhabitant in PPS according to EU average	2005	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 363
FOCI	gdp_cap_pps_ix	GDP per inhabitant in PPS according to EU average	2006	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 364
FOCI	gdp_cap_pps_evn	Difference between GDP per inhabitant at 2 dates.	1995-2006	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 365
FOCI	gdp_cap_pps_evn	Difference between GDP per inhabitant at 2 dates.	1995-2001	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 366
FOCI	gdp_cap_pps_evn	Difference between GDP per inhabitant at 2 dates.	2001-2006	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 367
FOCI	gdp_cap_pps_nat_ix	GDP per inhabitant in PPS according to national average	1995	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 368
FOCI	gdp_cap_pps_nat_ix	GDP per inhabitant in PPS according to national average	1996	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 369
FOCI	gdp_cap_pps_nat_ix	GDP per inhabitant in PPS according to national average	1997	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 370
FOCI	gdp_cap_pps_nat_ix	GDP per inhabitant in PPS according to national average	1998	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 371
FOCI	gdp_cap_pps_nat_ix	GDP per inhabitant in PPS according to national average	1999	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 372
FOCI	gdp_cap_pps_nat_ix	GDP per inhabitant in PPS according to national average	2000	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 373
FOCI	gdp_cap_pps_nat_ix	GDP per inhabitant in PPS according to national average	2001	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 374
FOCI	gdp_cap_pps_nat_ix	GDP per inhabitant in PPS according to national average	2002	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 375
FOCI	gdp_cap_pps_nat_ix	GDP per inhabitant in PPS according to national average	2003	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 376
FOCI	gdp_cap_pps_nat_ix	GDP per inhabitant in PPS according to national average	2004	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 377
FOCI	gdp_cap_pps_nat_ix	GDP per inhabitant in PPS according to national average	2005	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 378
FOCI	gdp_cap_pps_nat_ix	GDP per inhabitant in PPS according to national average	2006	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 379
FOCI	gdp_cap_pps_nat_evn	Difference between GDP per inhabitant at 2 dates.	1995-2006	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 380
FOCI	gdp_cap_pps_nat_evn	Difference between GDP per inhabitant at 2 dates.	1995-2001	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 381
FOCI	gdp_cap_pps_nat_evn	Difference between GDP per inhabitant at 2 dates.	2001-2006	0701 Economy - Aggregated accounts	LUZ	2008	ESPON	URBAN DATA - 382
FOCI	Ch_Empl-per100_Pop	Change in employees per 100 population in percentage points	2002-2005	0702 Economy - Employment	MEGA	2010	ESPON	URBAN DATA - 383
FOCI	Ch_Act_Rate	Change of activity rate in percentage points	2002-2005	0702 Economy - Employment	MEGA	2010	ESPON	URBAN DATA - 384
FOCI	Ch_Unemp_Rate	unemployment rate - change in percentage points	2002-2005	0702 Economy - Employment	MEGA	2010	ESPON	URBAN DATA - 385
FOCI	emp_indep_rt	Share of non agricultural self employed in the total employment	2000	0702 Economy - Employment	LUZ	2008	ESPON	URBAN DATA - 386
FOCI	emp_indep_rt	Share of non agricultural self employed in the total employment	2005	0702 Economy - Employment	LUZ	2008	ESPON	URBAN DATA - 387
FOCI	unemp_rt	Unemployment rate	2004-2006	0702 Economy - Employment	LUZ	2008	ESPON	URBAN DATA - 388
FOCI	unemp_rt	Unemployment rate	1999-2002	0702 Economy - Employment	LUZ	2008	ESPON	URBAN DATA - 389
FOCI	unemp_t	Unemployment rate	1999	0702 Economy - Employment	LUZ	2008	ESPON	URBAN DATA - 390
FOCI	unemp_t	Unemployment rate	2000	0702 Economy - Employment	LUZ	2008	ESPON	URBAN DATA - 391
FOCI	unemp_t	Unemployment rate	2001	0702 Economy - Employment	LUZ	2008	ESPON	URBAN DATA - 392
FOCI	unemp_t	Unemployment rate	2002	0702 Economy - Employment	LUZ	2008	ESPON	URBAN DATA - 393
FOCI	unemp_t	Unemployment rate	2003	0702 Economy - Employment	LUZ	2008	ESPON	URBAN DATA - 394
FOCI	unemp_t	Unemployment rate	2004	0702 Economy - Employment	LUZ	2008	ESPON	URBAN DATA - 395
FOCI	unemp_t	Unemployment rate	2005	0702 Economy - Employment	LUZ	2008	ESPON	URBAN DATA - 396
FOCI	unemp_t	Unemployment rate	2006	0702 Economy - Employment	LUZ	2008	ESPON	URBAN DATA - 397
FOCI	unemp_t	Unemployment rate	2007	0702 Economy - Employment	LUZ	2008	ESPON	URBAN DATA - 398
FOCI	unemp_nat	Unemployment rate regarding national average	1999	0702 Economy - Employment	LUZ	2008	ESPON	URBAN DATA - 399
FOCI	unemp_nat	Unemployment rate regarding national average	2000	0702 Economy - Employment	LUZ	2008	ESPON	URBAN DATA - 400
FOCI	unemp_nat	Unemployment rate regarding national average	2001	0702 Economy - Employment	LUZ	2008	ESPON	URBAN DATA - 401
FOCI	unemp_nat	Unemployment rate regarding national average	2002	0702 Economy - Employment	LUZ	2008	ESPON	URBAN DATA - 402
FOCI	unemp_nat	Unemployment rate regarding national average	2003	0702 Economy - Employment	LUZ	2008	ESPON	URBAN DATA - 403
FOCI	unemp_nat	Unemployment rate regarding national average	2004	0702 Economy - Employment	LUZ	2008	ESPON	URBAN DATA - 404
FOCI	unemp_nat	Unemployment rate regarding national average	2005	0702 Economy - Employment	LUZ	2008	ESPON	URBAN DATA - 405
FOCI	unemp_nat	Unemployment rate regarding national average	2006	0702 Economy - Employment	LUZ	2008	ESPON	URBAN DATA - 406
FOCI	unemp_nat	Unemployment rate regarding national average	2007	0702 Economy - Employment	LUZ	2008	ESPON	URBAN DATA - 407
FOCI	active_t	Total active population	2001	0702 Economy - Employment	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 408
FOCI	unemp_var	Unemployment gaps within the city, weighted variance	2001	0702 Economy - Employment	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 409
FOCI	unemp_var2	Unemployment gaps within the city, standard deviation	2001	0702 Economy - Employment	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 410
FOCI	Ch_Econ_Str_MetroReg_Ind	Change in the economic structure in metropolis-region context - indus	1989-2005	0703 Economy - Production and costs pe	MEGA	2010	ESPON	URBAN DATA - 411
FOCI	Ch_Econ_Str_MetroReg_Serv	Change in the economic structure in metropolis-region context - servi	1989-2005	0703 Economy - Production and costs pe	MEGA	2010	ESPON	URBAN DATA - 412
FOCI	Str_Sim_MetroReg_Context	Structural similarity in metropolis-region context	1998	0703 Economy - Production and costs pe	MEGA	2010	ESPON	URBAN DATA - 413
FOCI	Str_Sim_MetroReg_Context	Structural similarity in metropolis-region context	2005	0703 Economy - Production and costs pe	MEGA	2010	ESPON	URBAN DATA - 414

- Urban data (9)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
FOCI	Str_Sim_MetroReg_Context_Ch	Structural similarity changes in metropolis-region context	1998-2005	0703 Economy - Production and costs pe	MEGA	2010	ESPON	URBAN DATA - 415
FOCI	va_ab_t	Added value of NACE AB	2006	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 416
FOCI	va_ce_t	Added value of NACE C to E	2006	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 417
FOCI	va_f_t	Added value of construction Nace F	2006	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 418
FOCI	va_gh_t	Added value of NACE G to I	2006	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 419
FOCI	va_jk_t	Added value of NACE J and K	2006	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 420
FOCI	va_l_p_t	Added value of NACE L to P	2006	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 421
FOCI	gdp_cap_pps	GDP per inhabitant in PPS	1995	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 422
FOCI	gdp_cap_pps	GDP per inhabitant in PPS	1996	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 423
FOCI	gdp_cap_pps	GDP per inhabitant in PPS	1997	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 424
FOCI	gdp_cap_pps	GDP per inhabitant in PPS	1998	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 425
FOCI	gdp_cap_pps	GDP per inhabitant in PPS	1999	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 426
FOCI	gdp_cap_pps	GDP per inhabitant in PPS	2000	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 427
FOCI	gdp_cap_pps	GDP per inhabitant in PPS	2001	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 428
FOCI	gdp_cap_pps	GDP per inhabitant in PPS	2002	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 429
FOCI	gdp_cap_pps	GDP per inhabitant in PPS	2003	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 430
FOCI	gdp_cap_pps	GDP per inhabitant in PPS	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 431
FOCI	gdp_cap_pps	GDP per inhabitant in PPS	2005	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 432
FOCI	gdp_cap_pps	GDP per inhabitant in PPS	2006	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 433
FOCI	va_ab_rt	Share of the value added in NACE AB	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 434
FOCI	va_c_rt	Share of the value added in NACE C	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 435
FOCI	va_d_rt	Share of the value added in NACE D	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 436
FOCI	va_e_rt	Share of the value added in NACE E	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 437
FOCI	va_f_rt	Share of the value added in NACE F	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 438
FOCI	va_g_rt	Share of the value added in NACE G	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 439
FOCI	va_h_rt	Share of the value added in NACE H	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 440
FOCI	va_i_rt	Share of the value added in NACE I	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 441
FOCI	va_l_rt	Share of the value added in NACE J	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 442
FOCI	va_k_rt	Share of the value added in NACE K	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 443
FOCI	va_l_rt	Share of the value added in NACE L	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 444
FOCI	va_m_rt	Share of the value added in NACE M	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 445
FOCI	va_n_rt	Share of the value added in NACE N	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 446
FOCI	va_op_rt	Share of the value added in NACE OP	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 447
FOCI	va_da_rt	Share of the value added in NACE DA	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 448
FOCI	va_dbod_rt	Share of the value added in NACE DB and DC	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 449
FOCI	va_dd_rt	Share of the value added in NACE DD	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 450
FOCI	va_de_rt	Share of the value added in NACE DE	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 451
FOCI	va_dfdg_rt	Share of the value added in NACE DF and DG	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 452
FOCI	va_dh_rt	Share of the value added in NACE DH	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 453
FOCI	va_di_rt	Share of the value added in NACE DI	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 454
FOCI	va_dj_rt	Share of the value added in NACE DJ	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 455
FOCI	va_dk_rt	Share of the value added in NACE DK	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 456
FOCI	va_dl_rt	Share of the value added in NACE DL	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 457
FOCI	va_dm_rt	Share of the value added in NACE DM	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 458
FOCI	va_dn_rt	Share of the value added in NACE DN	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 459
FOCI	va_sect1_PC1	First component of the PCA analysis on sectoral structure	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 460
FOCI	va_sect1_PC2	Second component of the PCA analysis on sectoral structure	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 461
FOCI	va_sect1_PC3	Third component of the PCA analysis on sectoral structure	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 462
FOCI	va_sect1_PC4	Fourth component of the PCA analysis on sectoral structure	2004	0703 Economy - Production and costs pe	LUZ	2008	ESPON	URBAN DATA - 463
FOCI	specialisation	research specialisation in NBIC	1986-2006	0704 Economy - Research and innovatioi	FUA	2010	ESPON	URBAN DATA - 464
FOCI	transn_t	Number of Headquarters of Transnational firms in the 2000 biggest wc	2005	0704 Economy - Research and innovatioi	LUZ	2008	ESPON	URBAN DATA - 465
FOCI	transn_sales	sales of Transnational firms in the 2000 biggest world firms whose he	2005	0704 Economy - Research and innovatioi	LUZ	2008	ESPON	URBAN DATA - 466

- Urban data (10)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
FOCI	transn_emp	Employees of Transnational firms in the 2000 biggest world firms who:	2005	0704 Economy - Research and innovati	LUZ	2008	ESPON	URBAN DATA - 467
FOCI	patent_rt	patents for 1000000 inhabitants, average	2001-2003	0704 Economy - Research and innovati	LUZ	2008	ESPON	URBAN DATA - 468
FOCI	rd_gdp_rt	Share of total GDP in R&D	2003	0704 Economy - Research and innovati	LUZ	2008	ESPON	URBAN DATA - 469
FOCI	emp_so_rt	Share of scientific and technical employment on total employment	2005	0704 Economy - Research and innovati	LUZ	2008	ESPON	URBAN DATA - 470
FOCI	pond_luz	Weighter used for statistical analyses	2010	0902 Non/cross thematic data - Geograp	LUZ	2008	ESPON	URBAN DATA - 471
FOCI	pop_luz_rt	Weighter used for statistical analyses	2005	0902 Non/cross thematic data - Geograp	LUZ	2008	ESPON	URBAN DATA - 472
FOCI	pop_0-15_rt	Share of people between 0 and 14 years	1999-2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 473
FOCI	pop_15-24_rt	Share of people between 15 and 24 years	1999-2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 474
FOCI	pop_25-54_rt	Share of people between 25 and 54 years	1999-2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 475
FOCI	pop_55-64_rt	Share of people between 55 and 64 years	1999-2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 476
FOCI	pop_65+_rt	Share of people more or equal to 65 years	1999-2006	0201 Demography - Population structure	LUZ	2008	ESPON	URBAN DATA - 477
FOCI	Total_inertia	Total inertia	2001	9901 Non/cross thematic data - Integrativ	FUA	2010	ESPON	URBAN DATA - 478
FOCI	TYP_MacroReg_Trends_Compat	Types of macroregions based on trends compatibility	1998-2005	9901 Non/cross thematic data - Integrativ	MEGA	2010	ESPON	URBAN DATA - 479
FOCI	Modern_Economy	Spatial distribution of principal component values in metropolitan area	2005	9901 Non/cross thematic data - Integrativ	MEGA	2010	ESPON	URBAN DATA - 480
FOCI	Population_Potential	Spatial distribution of principal component values in metropolitan area	2005	9901 Non/cross thematic data - Integrativ	MEGA	2010	ESPON	URBAN DATA - 481
FOCI	National-Growth_Poles	Spatial distribution of principal component values in metropolitan area	2005	9901 Non/cross thematic data - Integrativ	MEGA	2010	ESPON	URBAN DATA - 482
FOCI	Weak_Suburban_Zone	Spatial distribution of principal component values in metropolitan area	2005	9901 Non/cross thematic data - Integrativ	MEGA	2010	ESPON	URBAN DATA - 483
FOCI	Classification_metropolitan_area	Classification of metropolitan areas	2005	9901 Non/cross thematic data - Integrativ	MEGA	2010	ESPON	URBAN DATA - 484
FOCI	Modern_Economy_Hinterlands	Spatial distribution of principal component values in regional hinterland	2005	9901 Non/cross thematic data - Integrativ	MEGA	2010	ESPON	URBAN DATA - 485
FOCI	Well_Developed_Labour_Market	Spatial distribution of principal component values in regional hinterland	2005	9901 Non/cross thematic data - Integrativ	MEGA	2010	ESPON	URBAN DATA - 486
FOCI	Population_Potential_Hinterlands	Spatial distribution of principal component values in regional hinterland	2005	9901 Non/cross thematic data - Integrativ	MEGA	2010	ESPON	URBAN DATA - 487
FOCI	Duality_economy	Spatial distribution of principal component values in regional hinterland	2005	9901 Non/cross thematic data - Integrativ	MEGA	2010	ESPON	URBAN DATA - 488
FOCI	TYP_Hinterlands	Classification of regional hinterlands	2005	9901 Non/cross thematic data - Integrativ	MEGA	2010	ESPON	URBAN DATA - 489
FOCI	IntraReg_Dev_Disparity	Regression residual - intraregional development disparities	2005	9901 Non/cross thematic data - Integrativ	MEGA	2010	ESPON	URBAN DATA - 490
FOCI	IntraReg_Dev_Disparity_Grow	Regression residuals - change of intraregional development disparities	1998-2005	9901 Non/cross thematic data - Integrativ	MEGA	2010	ESPON	URBAN DATA - 491
FOCI	areacc	Area of core city	1989-1993	9902 Non/cross thematic data - Geograp	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 492
FOCI	oc_luz	Core city LUZ ratio	1989-1993	9902 Non/cross thematic data - Geograp	Core city (Urban Audit)	2008	ESPON	URBAN DATA - 493
FOCI	beetweenness	beetweenness centrality index	1986-2006	9902 Non/cross thematic data - Geograp	FUA	2010	ESPON	URBAN DATA - 494
FOCI	degreesNValues	degrees centrality index	1986-2006	9902 Non/cross thematic data - Geograp	FUA	2010	ESPON	URBAN DATA - 495
FOCI	degreesValues	degrees centrality index	1986-2006	9902 Non/cross thematic data - Geograp	FUA	2010	ESPON	URBAN DATA - 496
FOCI	burt	burt index	1986-2006	9902 Non/cross thematic data - Geograp	FUA	2010	ESPON	URBAN DATA - 497
FOCI	NUTS3	NUTS3 unit corresponding to FUA	2006	9902 Non/cross-thematic data - Geograp	FUA	2010	ESPON	URBAN DATA - 498
FOCI	MEGA_TRANS_REGI	Classification of FUAs to MEGAs, Trans-national FUAs, Regional FUAs	2010	9902 Non/cross-thematic data - Geograp	FUA	2010	ESPON	URBAN DATA - 499
FOCI	N_S_w_E	Location of FUAs in Northern, Southern, Western and Eastern Europe	2010	9902 Non/cross-thematic data - Geograp	FUA	2010	ESPON	URBAN DATA - 500
FOCI	EU15_NEW12_OTHER	Location of FUAs in EU15, 12 *New member countries, countries outside EU	2010	9902 Non/cross-thematic data - Geograp	FUA	2010	ESPON	URBAN DATA - 501
FOCI	luz_nuts2_rt	ratio between LUZ population and NUTS2 area	2003-2006	9902 Non/cross-thematic data - Geograp	LUZ	2008	ESPON	URBAN DATA - 502
FOCI	id	contactability by air, all and rail-air on a cities pair, possible return trips	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 503
FOCI	tpsDispo	time available at destination city	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 504
FOCI	dNomOri	Name of Origin transport node	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 505
FOCI	dNomDes	Name of Destination transport node	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 506
FOCI	dNomConurbOri	Name of Origin conurbation	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 507
FOCI	dNomConurbDes	Name of Destination conurbation	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 508
FOCI	dHDep	time constrain for take-off or train departure	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 509
FOCI	dHArr	time of arrival at destination transport node of first trip (local time)	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 510
FOCI	dTemps	difference between time of arrival (Harr) and departure constrain (Hdep)	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 511
FOCI	dTVeh	difference between time of arrival (Harr) and effective time of departure	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 512
FOCI	dTMAP	walking time	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 513
FOCI	dTAtt	waiting time while on trip	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 514
FOCI	dTAtt1	waiting time before leaving initial transport node	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 515
FOCI	dTexte	detail of trip	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 516
FOCI	dType	modal structure of morning trip: 101 rail, 202 air, 102 rail then air, 201 air then rail	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 517
FOCI	tpsDispo	time available for meeting at destination city	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 518

- Urban data (11)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
FOCI	aType	modal structure of the return trip: 101 rail, 202 air, 102 rail then air, 201 air	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 519
FOCI	aHDep	time of leaving the transport nod for evening return trip	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 520
FOCI	aHArr	time constraint for return at the transport node of initial city	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 521
FOCI	aTEmps	difference between arrival constraint (aHarr) and departure from trans	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 522
FOCI	aTVEH	difference between affective time of arrival and time of departure from	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 523
FOCI	aTMAP	walking time for evening return trip	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 524
FOCI	aTAtt	waiting time while on evening return trip	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 525
FOCI	aTAtt1	remaining time after reaching transport node (due to timetables) when	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 526
FOCI	aTexte	detail of evening return trip	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 527
FOCI	dOri	id of Origin transport node of morning trip	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 528
FOCI	dDes	id of Destination transport node of morning trip	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 529
FOCI	dConurbOri	id of Origin MEGA	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 530
FOCI	dConurbDes	id of Destination MEGA	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 531
FOCI	aNomOri	label of departure transport node for evening return trip	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 532
FOCI	aNomDes	label of arrival transport node for evening return trip	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 533
FOCI	orix	longitude of departure transport node	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 534
FOCI	orix	longitude of departure transport node	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 535
FOCI	desx	longitude of destination transport node	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 536
FOCI	desy	latitude of destination transport node	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 537
FOCI	UAIdOri	id code of Urban Audit city of origin	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 538
FOCI	UAIdDes	id code of Urban Audit city of destination	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 539
FOCI	UANameOri	label (name) of Urban Audit city of origin	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 540
FOCI	UANameDes	label (name) of Urban Audit city of destination	2009	0301 Transport - Accessibility	City-pair	/	ESPON	URBAN DATA - 541
ESPON DB	TYPE	Type of UMZ	2010	9902 Non/Cross-thematic data - Geograp	UMZ	2010	ESPON SPACE	URBAN DATA - 542
ESPON DB	COUNTRY	Identifier of the country	2010	9902 Non/Cross-thematic data - Geograp	UMZ	2010	ESPON SPACE	URBAN DATA - 543
ESPON DB	POPUMZV41	Total population	2001	0201 Demography - Population structure	UMZ	2010	ESPON SPACE	URBAN DATA - 544
ESPON DB	IDLAU	Identifier of the LAU	2010	9902 Non/Cross-thematic data - Geograp	UMZ	2010	ESPON SPACE	URBAN DATA - 545
ESPON DB	NAME	Name of the UMZ	2010	9902 Non/Cross-thematic data - Geograp	UMZ	2010	ESPON SPACE	URBAN DATA - 546
ESPON DB	PERIMETER	UMZ Perimeter	2010	9902 Non/Cross-thematic data - Geograp	UMZ	2010	ESPON SPACE	URBAN DATA - 547
ESPON DB	AREA	UMZ areas	2010	9902 Non/Cross-thematic data - Geograp	UMZ	2010	ESPON SPACE	URBAN DATA - 548
ESPON DB	WGS84_X	X coordinate	2010	9902 Non/Cross-thematic data - Geograp	UMZ	2010	ESPON SPACE	URBAN DATA - 549
ESPON DB	WGS84_Y	Y coordinate	2010	9902 Non/Cross-thematic data - Geograp	UMZ	2010	ESPON SPACE	URBAN DATA - 550
ESPON DB	pop_mua	MUA Population	2001	0201 Demography - Population structure	FUA	2008	ESPON SPACE	URBAN DATA - 551
ESPON DB	pop_fua01	FUA population in 2001	2001	0201 Demography - Population structure	FUA	2008	ESPON SPACE	URBAN DATA - 552
ESPON DB	pop_fua06	FUA population in 2006	2006	0201 Demography - Population structure	FUA	2008	ESPON SPACE	URBAN DATA - 553
ESPON DB	pop_growth_01_06	population growth between 2001 and 2006	2001-2006	0201 Demography - Population structure	FUA	2008	ESPON SPACE	URBAN DATA - 554
ESPON DB	gdp_Meur_fua	GDP of the FUA	2006	0701 Economy - Aggregated accounts	FUA	2008	ESPON SPACE	URBAN DATA - 555
ESPON DB	gdp_inh_fua	GDP of the FUA by inhabitants	2006	0701 Economy - Aggregated accounts	FUA	2008	ESPON SPACE	URBAN DATA - 556
ESPON DB	pe_ab	share of ab sector added value	2006	0703 Economy - Production and costs pe	FUA	2008	ESPON SPACE	URBAN DATA - 557
ESPON DB	pe_ce	share of ce sector added value	2006	0703 Economy - Production and costs pe	FUA	2008	ESPON SPACE	URBAN DATA - 558
ESPON DB	pe_f	share of f sector added value	2006	0703 Economy - Production and costs pe	FUA	2008	ESPON SPACE	URBAN DATA - 559
ESPON DB	pe_g_h_i	share of g_h_i sector added value	2006	0703 Economy - Production and costs pe	FUA	2008	ESPON SPACE	URBAN DATA - 560
ESPON DB	pe_l_k	share of l_k sector added value	2006	0703 Economy - Production and costs pe	FUA	2008	ESPON SPACE	URBAN DATA - 561
ESPON DB	pe_l_to_p	share of l_to_p sector added value	2006	0703 Economy - Production and costs pe	FUA	2008	ESPON SPACE	URBAN DATA - 562
ESPON DB	chom2006	Unemployment in the FUA	2006	0702 Economy - Employment	FUA	2008	ESPON SPACE	URBAN DATA - 563
ESPON DB	fua_area	FUA area	2008	9902 Non/Cross-thematic data - Geograp	FUA	2008	ESPON SPACE	URBAN DATA - 564
ESPON DB	mua_area	MUA area	2008	9902 Non/Cross-thematic data - Geograp	FUA	2008	ESPON SPACE	URBAN DATA - 565
ESPON DB	compactness	Compactness of the FUA	2008	9902 Non/Cross-thematic data - Geograp	FUA	2008	ESPON SPACE	URBAN DATA - 566





- World data (3)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
ESPON DB	pop_t	Total population	2012	0201 Demography - Population structure	Countries	2008	WORLD	WORLD DATA - 105
ESPON DB	pop_t	Total population	2013	0201 Demography - Population structure	Countries	2008	WORLD	WORLD DATA - 106
ESPON DB	pop_t	Total population	2014	0201 Demography - Population structure	Countries	2008	WORLD	WORLD DATA - 107
ESPON DB	pop_t	Total population	2015	0201 Demography - Population structure	Countries	2008	WORLD	WORLD DATA - 108
ESPON DB	area_t	Total population	2005	9302 Non/cross thematic data - Geograp	Countries	2008	WORLD	WORLD DATA - 109
ESPON DB	pop_t	Total surface area	2005	0201 Demography - Population structure	Countries	2008	WORLD	WORLD DATA - 110
ESPON DB	pop_urb_t	Urban Population - Total	2005	0502 Land Use - Urban land use attributes	Countries	2008	WORLD	WORLD DATA - 111
ESPON DB	area_agri_t	Agricultural area	2005	0501 Land Use - Land use and land cover t	Countries	2008	WORLD	WORLD DATA - 112
ESPON DB	gdp_t	Gross Domestic Product	2005	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 113
ESPON DB	gni_t	Gross National Income	2005	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 114
ESPON DB	co2_t	Emissions of CO2	2005	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 115
ESPON DB	co2_cement_cdiac	Emissions of CO2 - from Cement Production (CDIAC)	2000	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 116
ESPON DB	co2_cement_cdiac	Emissions of CO2 - from Cement Production (CDIAC)	2001	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 117
ESPON DB	co2_cement_cdiac	Emissions of CO2 - from Cement Production (CDIAC)	2002	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 118
ESPON DB	co2_cement_cdiac	Emissions of CO2 - from Cement Production (CDIAC)	2003	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 119
ESPON DB	co2_cement_cdiac	Emissions of CO2 - from Cement Production (CDIAC)	2004	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 120
ESPON DB	co2_cement_cdiac	Emissions of CO2 - from Cement Production (CDIAC)	2005	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 121
ESPON DB	co2_cement_cdiac	Emissions of CO2 - from Cement Production (CDIAC)	2006	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 122
ESPON DB	co2_cement_cdiac	Emissions of CO2 - from Cement Production (CDIAC)	2007	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 123
ESPON DB	co2_cement_cdiac	Emissions of CO2 - from Cement Production (CDIAC)	2008	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 124
ESPON DB	co2_flaring_cdiac	Emissions of CO2 - from Gas Flaring (CDIAC)	2000	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 125
ESPON DB	co2_flaring_cdiac	Emissions of CO2 - from Gas Flaring (CDIAC)	2001	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 126
ESPON DB	co2_flaring_cdiac	Emissions of CO2 - from Gas Flaring (CDIAC)	2002	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 127
ESPON DB	co2_flaring_cdiac	Emissions of CO2 - from Gas Flaring (CDIAC)	2003	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 128
ESPON DB	co2_flaring_cdiac	Emissions of CO2 - from Gas Flaring (CDIAC)	2004	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 129
ESPON DB	co2_flaring_cdiac	Emissions of CO2 - from Gas Flaring (CDIAC)	2005	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 130
ESPON DB	co2_flaring_cdiac	Emissions of CO2 - from Gas Flaring (CDIAC)	2006	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 131
ESPON DB	co2_flaring_cdiac	Emissions of CO2 - from Gas Flaring (CDIAC)	2007	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 132
ESPON DB	co2_flaring_cdiac	Emissions of CO2 - from Gas Flaring (CDIAC)	2008	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 133
ESPON DB	co2_gas_cdiac	Emissions of CO2 - from Gas Fuels Consumption (CDIAC)	2000	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 134
ESPON DB	co2_gas_cdiac	Emissions of CO2 - from Gas Fuels Consumption (CDIAC)	2001	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 135
ESPON DB	co2_gas_cdiac	Emissions of CO2 - from Gas Fuels Consumption (CDIAC)	2002	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 136
ESPON DB	co2_gas_cdiac	Emissions of CO2 - from Gas Fuels Consumption (CDIAC)	2003	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 137
ESPON DB	co2_gas_cdiac	Emissions of CO2 - from Gas Fuels Consumption (CDIAC)	2004	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 138
ESPON DB	co2_gas_cdiac	Emissions of CO2 - from Gas Fuels Consumption (CDIAC)	2005	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 139
ESPON DB	co2_gas_cdiac	Emissions of CO2 - from Gas Fuels Consumption (CDIAC)	2006	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 140
ESPON DB	co2_gas_cdiac	Emissions of CO2 - from Gas Fuels Consumption (CDIAC)	2007	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 141
ESPON DB	co2_gas_cdiac	Emissions of CO2 - from Gas Fuels Consumption (CDIAC)	2008	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 142
ESPON DB	co2_liquid_cdiac	Emissions of CO2 - from Liquid Fuels Consumption (CDIAC)	2000	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 143
ESPON DB	co2_liquid_cdiac	Emissions of CO2 - from Liquid Fuels Consumption (CDIAC)	2001	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 144
ESPON DB	co2_liquid_cdiac	Emissions of CO2 - from Liquid Fuels Consumption (CDIAC)	2002	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 145
ESPON DB	co2_liquid_cdiac	Emissions of CO2 - from Liquid Fuels Consumption (CDIAC)	2003	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 146
ESPON DB	co2_liquid_cdiac	Emissions of CO2 - from Liquid Fuels Consumption (CDIAC)	2004	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 147
ESPON DB	co2_liquid_cdiac	Emissions of CO2 - from Liquid Fuels Consumption (CDIAC)	2005	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 148
ESPON DB	co2_liquid_cdiac	Emissions of CO2 - from Liquid Fuels Consumption (CDIAC)	2006	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 149
ESPON DB	co2_liquid_cdiac	Emissions of CO2 - from Liquid Fuels Consumption (CDIAC)	2007	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 150
ESPON DB	co2_liquid_cdiac	Emissions of CO2 - from Liquid Fuels Consumption (CDIAC)	2008	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 151
ESPON DB	co2_solid_cdiac	Emissions of CO2 - from Solid Fuels Consumption (CDIAC)	2000	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 152
ESPON DB	co2_total_cdiac	Emissions of CO2 - from Fossil Fuels - Total (CDIAC)	2001	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 153
ESPON DB	co2_total_cdiac	Emissions of CO2 - from Fossil Fuels - Total (CDIAC)	2002	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 154
ESPON DB	co2_total_cdiac	Emissions of CO2 - from Fossil Fuels - Total (CDIAC)	2003	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 155
ESPON DB	co2_total_cdiac	Emissions of CO2 - from Fossil Fuels - Total (CDIAC)	2004	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 156



- World data (4)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
ESPON DB	co2_total_cdiac	Emissions of CO2 - from Fossil Fuels - Total (CDIAC)	2005	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 157
ESPON DB	co2_total_cdiac	Emissions of CO2 - from Fossil Fuels - Total (CDIAC)	2006	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 158
ESPON DB	co2_total_cdiac	Emissions of CO2 - from Fossil Fuels - Total (CDIAC)	2007	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 159
ESPON DB	co2_total_cdiac	Emissions of CO2 - from Fossil Fuels - Total (CDIAC)	2008	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 160
ESPON DB	gdp	GDP (constant 2000 US\$)	2000	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 161
ESPON DB	gdp	GDP (constant 2000 US\$)	2001	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 162
ESPON DB	gdp	GDP (constant 2000 US\$)	2002	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 163
ESPON DB	gdp	GDP (constant 2000 US\$)	2003	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 164
ESPON DB	gdp	GDP (constant 2000 US\$)	2004	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 165
ESPON DB	gdp	GDP (constant 2000 US\$)	2005	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 166
ESPON DB	gdp	GDP (constant 2000 US\$)	2006	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 167
ESPON DB	gdp	GDP (constant 2000 US\$)	2007	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 168
ESPON DB	gdp	GDP (constant 2000 US\$)	2008	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 169
ESPON DB	gdp_growth_rate	GDP growth (annual %)	2000	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 170
ESPON DB	gdp_growth_rate	GDP growth (annual %)	2001	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 171
ESPON DB	gdp_growth_rate	GDP growth (annual %)	2002	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 172
ESPON DB	gdp_growth_rate	GDP growth (annual %)	2003	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 173
ESPON DB	gdp_growth_rate	GDP growth (annual %)	2004	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 174
ESPON DB	gdp_growth_rate	GDP growth (annual %)	2005	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 175
ESPON DB	gdp_growth_rate	GDP growth (annual %)	2006	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 176
ESPON DB	gdp_growth_rate	GDP growth (annual %)	2007	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 177
ESPON DB	gdp_growth_rate	GDP growth (annual %)	2008	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 178
ESPON DB	gdp_ppp	GDP, PPP (constant 2005 international \$)	2000	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 179
ESPON DB	gdp_ppp	GDP, PPP (constant 2005 international \$)	2001	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 180
ESPON DB	gdp_ppp	GDP, PPP (constant 2005 international \$)	2002	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 181
ESPON DB	gdp_ppp	GDP, PPP (constant 2005 international \$)	2003	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 182
ESPON DB	gdp_ppp	GDP, PPP (constant 2005 international \$)	2004	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 183
ESPON DB	gdp_ppp	GDP, PPP (constant 2005 international \$)	2005	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 184
ESPON DB	gdp_ppp	GDP, PPP (constant 2005 international \$)	2006	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 185
ESPON DB	gdp_ppp	GDP, PPP (constant 2005 international \$)	2007	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 186
ESPON DB	gdp_ppp	GDP, PPP (constant 2005 international \$)	2008	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 187
ESPON DB	unemployment	Unemployment, total (% of total labor force)	2000	0702 Economy - Employment	Countries	2008	WORLD	WORLD DATA - 188
ESPON DB	unemployment	Unemployment, total (% of total labor force)	2001	0702 Economy - Employment	Countries	2008	WORLD	WORLD DATA - 189
ESPON DB	unemployment	Unemployment, total (% of total labor force)	2002	0702 Economy - Employment	Countries	2008	WORLD	WORLD DATA - 190
ESPON DB	unemployment	Unemployment, total (% of total labor force)	2003	0702 Economy - Employment	Countries	2008	WORLD	WORLD DATA - 191
ESPON DB	unemployment	Unemployment, total (% of total labor force)	2004	0702 Economy - Employment	Countries	2008	WORLD	WORLD DATA - 192
ESPON DB	unemployment	Unemployment, total (% of total labor force)	2005	0702 Economy - Employment	Countries	2008	WORLD	WORLD DATA - 193
ESPON DB	unemployment	Unemployment, total (% of total labor force)	2006	0702 Economy - Employment	Countries	2008	WORLD	WORLD DATA - 194
ESPON DB	unemployment	Unemployment, total (% of total labor force)	2007	0702 Economy - Employment	Countries	2008	WORLD	WORLD DATA - 195
ESPON DB	unemployment	Unemployment, total (% of total labor force)	2008	0702 Economy - Employment	Countries	2008	WORLD	WORLD DATA - 196
ESPON DB	pop_total	Total population (split in 157 indicators -> age structure)	1950	0201 Demography - Population structure	Countries	2008	WORLD	WORLD DATA - 197
ESPON DB	pop_total	Total population (split in 157 indicators -> age structure)	1951	0201 Demography - Population structure	Countries	2008	WORLD	WORLD DATA - 198
ESPON DB	pop_total	Total population (split in 157 indicators -> age structure)	1952	0201 Demography - Population structure	Countries	2008	WORLD	WORLD DATA - 199
ESPON DB	pop_total	Total population (split in 157 indicators -> age structure)	1953	0201 Demography - Population structure	Countries	2008	WORLD	WORLD DATA - 200
ESPON DB	pop_total	Total population (split in 157 indicators -> age structure)	1954	0201 Demography - Population structure	Countries	2008	WORLD	WORLD DATA - 201
ESPON DB	pop_total	Total population (split in 157 indicators -> age structure)	1955	0201 Demography - Population structure	Countries	2008	WORLD	WORLD DATA - 202
ESPON DB	pop_total	Total population (split in 157 indicators -> age structure)	1956	0201 Demography - Population structure	Countries	2008	WORLD	WORLD DATA - 203
ESPON DB	pop_total	Total population (split in 157 indicators -> age structure)	1957	0201 Demography - Population structure	Countries	2008	WORLD	WORLD DATA - 204
ESPON DB	pop_total	Total population (split in 157 indicators -> age structure)	1958	0201 Demography - Population structure	Countries	2008	WORLD	WORLD DATA - 205
ESPON DB	pop_total	Total population (split in 157 indicators -> age structure)	1959	0201 Demography - Population structure	Countries	2008	WORLD	WORLD DATA - 206
ESPON DB	pop_total	Total population (split in 157 indicators -> age structure)	1960	0201 Demography - Population structure	Countries	2008	WORLD	WORLD DATA - 207
ESPON DB	pop_total	Total population (split in 157 indicators -> age structure)	1961	0201 Demography - Population structure	Countries	2008	WORLD	WORLD DATA - 208



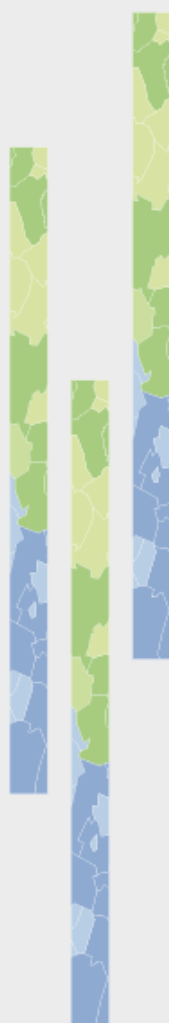
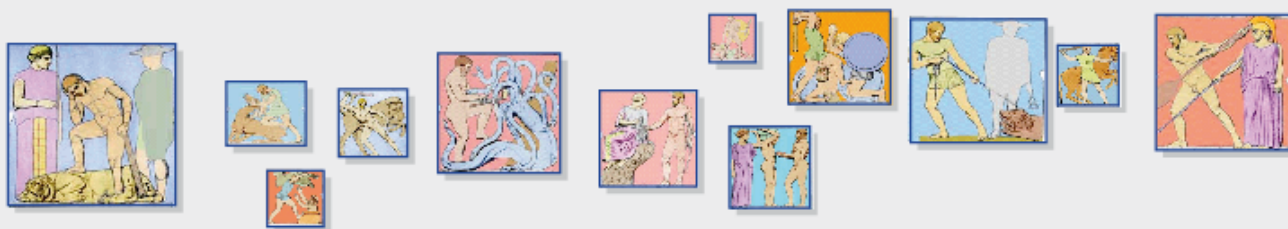


- World data (7)

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
ESPON DB	co2_unfccc_inc	Emissions of CO2 - (National Reports, UNFCCC), Including Land Use	2006	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 313
ESPON DB	co2_unfccc_inc	Emissions of CO2 - (National Reports, UNFCCC), Including Land Use	2007	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 314
ESPON DB	co2_unfccc_inc	Emissions of CO2 - (National Reports, UNFCCC), Including Land Use	2008	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 315
ESPON DB	co2_unfccc_exc	Emissions of CO2 - (National Reports, UNFCCC), Excluding Land Use	2000	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 316
ESPON DB	co2_unfccc_exc	Emissions of CO2 - (National Reports, UNFCCC), Excluding Land Use	2001	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 317
ESPON DB	co2_unfccc_exc	Emissions of CO2 - (National Reports, UNFCCC), Excluding Land Use	2002	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 318
ESPON DB	co2_unfccc_exc	Emissions of CO2 - (National Reports, UNFCCC), Excluding Land Use	2003	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 319
ESPON DB	co2_unfccc_exc	Emissions of CO2 - (National Reports, UNFCCC), Excluding Land Use	2004	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 320
ESPON DB	co2_unfccc_exc	Emissions of CO2 - (National Reports, UNFCCC), Excluding Land Use	2005	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 321
ESPON DB	co2_unfccc_exc	Emissions of CO2 - (National Reports, UNFCCC), Excluding Land Use	2006	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 322
ESPON DB	co2_unfccc_exc	Emissions of CO2 - (National Reports, UNFCCC), Excluding Land Use	2007	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 323
ESPON DB	co2_unfccc_exc	Emissions of CO2 - (National Reports, UNFCCC), Excluding Land Use	2008	0402 Energy & environment - Climate cha	Countries	2008	WORLD	WORLD DATA - 324
ESPON DB	gdp	Present value of external debt (current US\$)	2000	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 325
ESPON DB	gdp	Present value of external debt (current US\$)	2001	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 326
ESPON DB	gdp	Present value of external debt (current US\$)	2002	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 327
ESPON DB	gdp	Present value of external debt (current US\$)	2003	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 328
ESPON DB	gdp	Present value of external debt (current US\$)	2004	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 329
ESPON DB	gdp	Present value of external debt (current US\$)	2005	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 330
ESPON DB	gdp	Present value of external debt (current US\$)	2006	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 331
ESPON DB	gdp	Present value of external debt (current US\$)	2007	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 332
ESPON DB	gdp	Present value of external debt (current US\$)	2008	0701 Economy - Aggregated accounts	Countries	2008	WORLD	WORLD DATA - 333

- **Grid data**

Provider	Indicator - Code	Indicator - Name	Year(s)	Theme-Sub theme	Geo object	Geo version	SPACE	CATEGORY
ESPON DB	gdp	Gross Domestic Product	2003	0701 Economy - Aggregated accounts	1 km grid	2008	ESPON SPACE	GRID DATA
ESPON DB	gdp	Gross Domestic Product	2006	0701 Economy - Aggregated accounts	1 km grid	2008	ESPON SPACE	GRID DATA
ESPON DB	active	Active population	2003	0602 Social affairs - Labour market	1 km grid	2008	ESPON SPACE	GRID DATA
ESPON DB	active	Active population	2006	0602 Social affairs - Labour market	1 km grid	2008	ESPON SPACE	GRID DATA
ESPON DB	unemp	Unemployed persons	2003	0702 Economy - Employment	1 km grid	2008	ESPON SPACE	GRID DATA
ESPON DB	unemp	Unemployed persons	2006	0702 Economy - Employment	1 km grid	2008	ESPON SPACE	GRID DATA
ESPON DB	No code (OLAP Cube)	Geographical dimension: Biogeographic regions (9 categories)	2008	9902 Non/cross thematic data - Geograp	OLAP	V3.0	ESPON SPACE	GRID DATA
ESPON DB	No code (OLAP Cube)	Geographical dimension: Elevation breakdown (6 categories)	2006	9902 Non/cross thematic data - Geograp	OLAP	V3.0	ESPON SPACE	GRID DATA
ESPON DB	No code (OLAP Cube)	Geographical dimension: LUZ (311 units)	2006	9902 Non/cross thematic data - Geograp	OLAP	V3.0	ESPON SPACE	GRID DATA
ESPON DB	No code (OLAP Cube)	Geographical dimension: Massifs (17 massifs)	2006	9902 Non/cross thematic data - Geograp	OLAP	V3.0	ESPON SPACE	GRID DATA
ESPON DB	No code (OLAP Cube)	Geographical dimension: NUTS 2003 nomenclature (NUTS0-1-2-3)	2006	9902 Non/cross thematic data - Geograp	OLAP	V3.0	ESPON SPACE	GRID DATA
ESPON DB	No code (OLAP Cube)	Geographical dimension: NUTS 2006 nomenclature (NUTS0-1-2-3)	2006	9902 Non/cross thematic data - Geograp	OLAP	V3.0	ESPON SPACE	GRID DATA
ESPON DB	No code (OLAP Cube)	Geographical dimension: River basin districts (131 basins)	2006	9902 Non/cross thematic data - Geograp	OLAP	V3.0	ESPON SPACE	GRID DATA
ESPON DB	No code (OLAP Cube)	Geographical dimension: Regional sea basins (5 basins)	2006	9902 Non/cross thematic data - Geograp	OLAP	V3.0	ESPON SPACE	GRID DATA
ESPON DB	No code (OLAP Cube)	Socio-economic variables: GDP	2003	9902 Non/cross thematic data - Geograp	OLAP	V3.0	ESPON SPACE	GRID DATA
ESPON DB	No code (OLAP Cube)	Socio-economic variables: GDP	2006	9902 Non/cross thematic data - Geograp	OLAP	V3.0	ESPON SPACE	GRID DATA
ESPON DB	No code (OLAP Cube)	Socio-economic variables: active population	2003	9902 Non/cross thematic data - Geograp	OLAP	V3.0	ESPON SPACE	GRID DATA
ESPON DB	No code (OLAP Cube)	Socio-economic variables: active population	2006	9902 Non/cross thematic data - Geograp	OLAP	V3.0	ESPON SPACE	GRID DATA
ESPON DB	No code (OLAP Cube)	Socio-economic variables: unemployment	2003	9902 Non/cross thematic data - Geograp	OLAP	V3.0	ESPON SPACE	GRID DATA
ESPON DB	No code (OLAP Cube)	Socio-economic variables: unemployment	2006	9902 Non/cross thematic data - Geograp	OLAP	V3.0	ESPON SPACE	GRID DATA
ESPON DB	No code (OLAP Cube)	Land Cover: Corine Land Cover	1990	9902 Non/cross thematic data - Geograp	OLAP	V3.0	ESPON SPACE	GRID DATA
ESPON DB	No code (OLAP Cube)	Land Cover: Corine Land Cover	2000	9902 Non/cross thematic data - Geograp	OLAP	V3.0	ESPON SPACE	GRID DATA
ESPON DB	No code (OLAP Cube)	Land Cover: Corine Land Cover	2006	9902 Non/cross thematic data - Geograp	OLAP	V3.0	ESPON SPACE	GRID DATA
ESPON DB	No code (OLAP Cube)	Land Cover: Corine Land Cover change	1990-2006	9902 Non/cross thematic data - Geograp	OLAP	V3.0	ESPON SPACE	GRID DATA
ESPON DB	No code (OLAP Cube)	Land Cover: Corine Land Cover change	1990-2000	9902 Non/cross thematic data - Geograp	OLAP	V3.0	ESPON SPACE	GRID DATA
ESPON DB	No code (OLAP Cube)	Land Cover: Corine Land Cover change	2000-2006	9902 Non/cross thematic data - Geograp	OLAP	V3.0	ESPON SPACE	GRID DATA
ESPON DB	No code (OLAP Cube)	Measures: population density	2001	9902 Non/cross thematic data - Geograp	OLAP	V3.0	ESPON SPACE	GRID DATA
ESPON DB	No code (OLAP Cube)	Measures: land area	2006	9902 Non/cross thematic data - Geograp	OLAP	V3.0	ESPON SPACE	GRID DATA



## Modeling territorial changes and time series database building process: empirical approach and applications

### CONTENT

The problem of time series data is the lack of data a territorial unit either because the territorial unit has changed in the course of time or because data are simply missing.

Three complementary methods have been investigated: data collection in historical regional databases, NUTS changes modeling and missing values estimation.

The building of a dictionary of changes is a fundamental step to build harmonized spatiotemporal database. It consists to describing and modeling of territorial changes.

This theoretical framework was tested through data sets.

**ESPON 2013 DATABASE**  
**MARCH 2011**



# LIST OF AUTHORS

Ben Rebah Maher, UMS 2414 RIATE

Plumejeaud Christine, LIG

Ysebaert Ronan, UMS 2414 RIATE

Peeters Didier, IGEAT, Université Libre de Bruxelles

## Contact

[maher.benrebah@ums-riate.fr](mailto:maher.benrebah@ums-riate.fr)

[christine.plumejeaud@imag.fr](mailto:christine.plumejeaud@imag.fr)

[ronan.ysebaert@ums-riate.fr](mailto:ronan.ysebaert@ums-riate.fr)

[dpeeter1@ulb.ac.be](mailto:dpeeter1@ulb.ac.be)

UMS 2414 RIATE

Tel. (+ 33) 1 57 27 65 35

LIG Grenoble

Tel. (+33) 4 76 82 72 11

Université Libre de Bruxelles

Tel. (+32) 2 650 50 77

# TABLE OF CONTENT

Introduction .....	3
<b>1 Territorial changing information sources .....</b>	<b>5</b>
1.1 Legal source: Official journal of the European Union .....	5
1.2 Eurostat.....	6
1.3 National Statistical Institutes.....	7
1.4 Conclusion of the first part .....	10
<b>2 Nuts changes knowledge: from elementary to systemic approach of territorial changes.....</b>	<b>11</b>
2.1 Elementary changes .....	11
2.2 Systemic approach of NUTS changes .....	14
<b>3 Building historical database of territorial changes: from conceptual approach to operational solutions.....</b>	<b>16</b>
3.1 Improved Snapshot model (presentation) .....	16
3.2 The space-time composite model: reconstructing genealogy of Nuts versions .....	18
3.3 Towards a NUTS changes mapping.....	20
3.3.1 A general view .....	20
3.3.2 Typology of changes.....	21
3.3.3 Maps by times intervals .....	21
3.3.4 Examples of lineage (genealogy) visualization .....	21
3.4 Tables structure of NUTS changes .....	22
3.4.1 Changes location tables .....	22
3.4.2 Changes typology tables .....	22
3.4.3 "Units life" tables .....	23
3.4.4 "genealogy units" tables .....	23
<b>4 Example of time series data building process: the European regional cohesion indicators .....</b>	<b>24</b>
4.1 Data collection and temporal series maps .....	24
4.2 Building a harmonized temporal data base for the European regional cohesion analysis .....	25
4.2.1 Cohesion reports datasets exploration: missing values and lack of compatibility between NUTS versions .....	26
4.2.2 NUTS harmonisation process .....	27
4.3 Spatio-temporal analysis of the European regional cohesion .....	28
<b>Conclusion .....</b>	<b>31</b>
<b>Bibliography.....</b>	<b>32</b>



# Introduction

ESPON DB 2013 project aims to improve the access to time series data. The issue of time series is a recurring necessity for ESPON projects and several European institutions primarily DG REGIO and ESUROSTAT. In spite of its importance, this process has not very well initiated by the previous ESPON DB project (2006)<sup>1</sup>.

This brief technical report is for the purpose of providing a background to time series challenge of ESPON DB project during the timeframe December 2008 to December 2009. In this first period of the project, activities done are mainly conceptual. Some applications and operational results have been developed since November 2009 (figure 25).

The issue of time series data could be fundamentally assimilated to the lack of data for a territorial unit either because the territorial unit has changed in the course of time or because data are simply missing. Difficulties to build time series data can be related firstly to the lack of achieved databases EUROSTAT, as the principal provider of European statistics, does not archive its database versions. It keeps just the last version of database. Secondly, information about historical changes of NUTS is often missing or uncertain.

Time series approach can be organized in two main steps. Firstly, collection and exploration of historical data bases (NewCronos from EUROSTAT, cohesion reports from DG-REGIO...) was undertaken. This work aims to provide a review of continuous time-data series could be built from this data bases. Additionally, we have explored NUTS changes between 1995 and 2006. The dictionary of NUTS changes is the main result of this exploration. It allows a review of territorial changes (codes, names and geometries). But the most contribution of the dictionary of changes is the identification of the genealogy (lineage) of NUTS which is very useful for the harmonization of time-series data.

The result of this first step will be used to build continuous time-series data. The conceptual model and the implementation of the computing (automation) of the process is in progress.

---

<sup>1</sup> [http://www.espon.eu/mmp/online/website/content/tools/832/index\\_EN.html](http://www.espon.eu/mmp/online/website/content/tools/832/index_EN.html)

1.7 In what sense is the MAUP a problem? (p. 15)

[...]

**The most important problem is about international and historical comparisons: do the elementary spatial units which are used for the analysis have the same meaning in two different countries? At two different time periods?** It is not easy to determine if a difference in the results is due to a difference in the processes which are underlying the observed phenomena, or simply to a difference in the meaning of the spatial entities that are used for the observation.

[http://www.espon.eu/mmp/online/website/content/projects/261/431/index\\_EN.html](http://www.espon.eu/mmp/online/website/content/projects/261/431/index_EN.html)

4.3.4.1 Temporal integration

(p. 110)

[...] **Identificators or the geometries of the NUTS change strongly during the period.**

These changes introduce very big difficulties in the survey of variables in the time. It doesn't exist any simple ties between two dates.

(p. 120)

**Changes of geometry and changes of units identification, don't permit to get directly evolutions of population basing on initial data,** as the shows following example:

We estimate an evolution for a middle time (1990-2000) and represent it for different geographical grids (NUTS 23 1988 and NUTS 23 1999) whereas data of population initial are the similar, calculations of evolution (1990-2000) defer very strongly from a geographical grid to the other.

[http://www.espon.eu/mmp/online/website/content/tools/127/index\\_EN.html](http://www.espon.eu/mmp/online/website/content/tools/127/index_EN.html)

6.5. Conclusion (vol. 1 p 242)

The synthesis of the regional insertion of the ESPON region into the world economy and the typology of gateway cities that we have elaborated in this final section of the report cannot be considered as definitive results as their elaboration was based on a limited number of criteria. **Better results could be obtained in the future if, for example, international trades statistics can be obtain for the regional level or if coherent time series could be analyzed concerning the evolution of air traffic linking European cities to the rest of the World.** The current set of results does however uncover some important findings in accordance with the objectives of the ESDP.

[http://www.espon.eu/mmp/online/website/content/projects/260/720/index\\_EN.html](http://www.espon.eu/mmp/online/website/content/projects/260/720/index_EN.html)

# 1 Territorial changing information sources

Data availability and data quality are crucial for the understanding and the formalization of Nuts genealogy. Our attempt to harmonize NUTS versions is the result of a meticulous combination of several sources provided by European and national institutions.

## 1.1 Legal source: Official journal of the European Union

The Official Journal is the legal source. It constitutes the legal framework of regulation of NUTS since 2003. The regulation EC n° 1059/2003 defines the NUTS and states the conditions of their modifications. This information is very useful to understand and formalize the changes of NUTS. This founder juridical text is amended and updated when new countries joined the European Union (figure1).

**Figure 1:** Example of regulations amendment following the accession of Czech Republic and Romania

REGULATIONS			
REGULATION (EC) No 176/2008 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 20 February 2008 amending Regulation (EC) No 1059/2003 on the establishment of a common classification of territorial units for statistics (NUTS) by reason of the accession of Bulgaria and Romania to the European Union			
Annex I to Regulation (EC) No 1059/2003 is amended as follows:			
1. the following table is inserted between BE — BELGIQUE/BELGIË and CZ — ČESKÁ REPUBLIKA:			
<b>БЪЛГАРИЯ</b>			
CODE	NUTS 1	NUTS 2	NUTS 3
BG			
BG3	СЕВЕРНА И ЮГОИЗТОЧНА БЪЛГАРИЯ		
BG31		Северозападен	
BG311			Видин
BG312			Монтана
BG313			Враца
BG314			Плевен
BG315			Ловеч

## 1.2 Eurostat

Eurostat is the most valuable source. Many kinds of documents are produced by Eurostat<sup>2</sup> allowing with NUTS changes, among which the most interesting is the description of changes occurring between each version. However, this description does not usually define types of changes. It is also, sometimes very imprecise, in the case of complex territorial modification, like the Danish territorial modification in 2006 which is described as follow: "Following an extensive regional reform in Denmark, where new administrative regions were created, Denmark will be divided into NUTS level 2 regions. The previous NUTS 3 regions do not generally correspond to the new NUTS level 2 regions. [...] "The previous 15 administrative regions have been abolished and in their place, 11 new non administrative regions have been created by combining municipalities. Only two NUTS 3 level 3 regions remain intact".

Concerning the update and the of EUROSTAT database, EUROSTAT does not archive its database versions. First of all, it keeps just the last version of database. Secondly, information about historical changes of NUTS is often missing or uncertain.

Besides the data available on the Eurostat internet portal, we obtained a CDROM with the Windows-only New Cronos application (figure 2), i. e. the Eurostat archives. This CDROM was unsuitable for the needs of the project because of its web interface designed exclusively for data consultation and not for data exportation. The data were also stored on the CDROM in a specific file format unknown from us which led us to spend time on finding technical workarounds to finally extract and store these in a format we could handle.

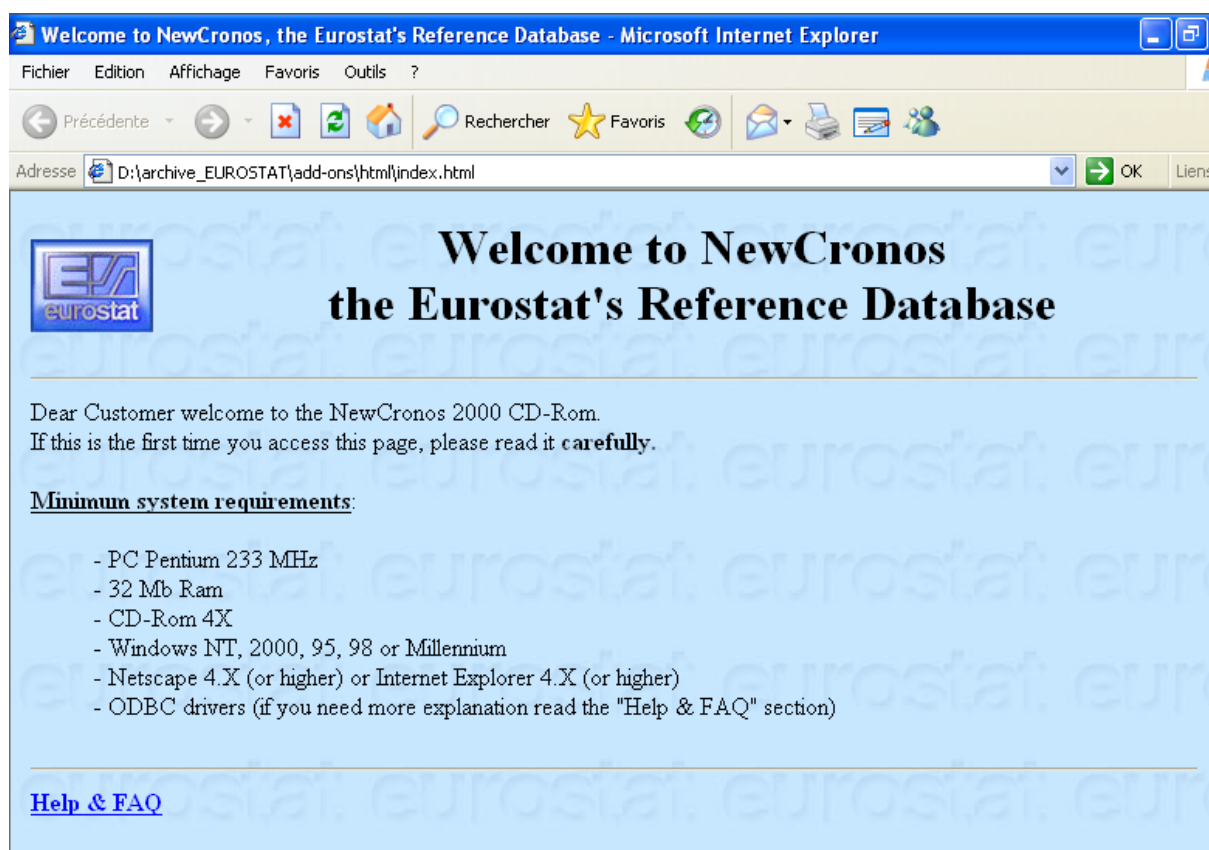
The data appeared to be organised in 271 tables and 16 categories. We made an inventory of their content in order to have an idea of their completeness, so to say the time span covered and the territories covered. The nuts are from 1999, and all European countries that are currently EU members are represented, but this of course depending on the type of data, the nuts level, the years considered, and logically the completeness of these archive decreases with older data.

To provide here an exhaustive list of the content, even in a synthesised way, would be a non-sense because of the number of variables and parameters. The data currently available on the Eurostat internet portal and the data included in these archives are partially the same, except that they do not use the same nuts reference system.

These data will be included in the Database system but will depend on the time series conversion tool to mix them with the current Eurostat data. Reversely, since they refer to an older nuts genealogy (1999) they might be useful in the next step of the time series harmonisation challenge, but probably as a validation mean, to be compared with the values our tool will compute for the 1999 nuts references.

---

<sup>2</sup> [http://ec.europa.eu/eurostat/ramon/nuts/splash\\_regions.html](http://ec.europa.eu/eurostat/ramon/nuts/splash_regions.html)



*Figure 2: NewCronos Web interface*

### 1.3 National Statistical Institutes

The European national statistical offices provide historical databases of national administrative boundaries. These sources are very useful for the understanding of local changes (national) which may affect geometry or structure of NUTS levels. National boundaries historical database is also very essential in the case of accessing new countries (EU15, EU25, and EU27) because EUROSTAT databases do not provide long term information about the historical administrative boundaries of these New Members.

Although a relatively high numbers of countries describe the changes of their administrative boundaries, the attempt does not construct a real temporal database. The most frequent method is to list changes as events (juridical rules) without relations between administrative units states. Historical "communes" database in France, done by the INSEE, describe changes of the local level from 1930 until today. However, this database is not easily workable because of describing textual change (Figure 3)

ID MUNICIPALITY	DATE OF THE CHANGE	DESCRIPTION
51 261	14/12/1930	Fresnes-lès-Reims become Fresne-lès-Reims.
67 161	21/02/1948	Gottenhausen become Gottenhouse.
12 307	03/12/1952	Curan is created thanks to some areas of Salles-Curan.
51 136	01/03/2006	Châtillon-sur-Marne is separated from Cuisles.
21 551	01/01/2009	Saint-Germain-Source-Seine is merged to Blessey, which become Source-Seine

*Figure 3: Extract of the "Historical French communes' database" done by INSEE*

Danish National statistical Institute has built population time series data based in the current Nuts version (figure3). Earlier 2003 territorial nuts changes data sets are also available.

NUTS VERSION	CODE	NAME	1979	1980	...	2004	2005	2006	...	2009
NUTS 2003	DK001	Copenhagen	505974	498850	...	501664	502362	501158	...	No data
NUTS 2003	DK001	Frederiksberg	88835	88287	...	91721	91886	91855	...	No data
NUTS 2003	DK002	Copenhagen County	629928	627245	...	618407	618237	618529	...	No data
NUTS 2003	DK003	Frederiksborg County	325855	329141	...	373688	375705	378686	...	No data
NUTS 2003	DK004	Roskilde County	199672	202017	...	237089	239049	241523	...	No data
NUTS 2003	DK005	West Zealand County	275985	277833	...	302479	304761	307207	...	No data
NUTS 2003	DK006	Storstrøm County	259445	260081	...	261884	262144	262781	...	No data
NUTS 2003	DK007	Bornholm (excl. Christiansø)	47605	47780	...	43673	43347	43245	...	No data
NUTS 2003	DK008	Funen County	451727	452965	...	475082	476580	478347	...	No data
NUTS 2003	DK009	South Jutland County	248985	249949	...	252936	252980	252433	...	No data
NUTS 2003	DK00A	Ribe County	211492	212624	...	224595	224454	224261	...	No data
NUTS 2003	DK00B	Vejle County	323418	325774	...	355691	358055	360921	...	No data
NUTS 2003	DK00C	Ringkøbing County	261028	262751	...	274830	274574	275065	...	No data
NUTS 2003	DK00D	Århus County	571702	573916	...	653472	657671	661370	...	No data
NUTS 2003	DK00E	Viborg County	230536	231517	...	234659	234434	234896	...	No data
NUTS 2003	DK00F	North Jutland County	479349	481335	...	495669	495068	495090	...	No data
NUTS 2006	DK011	Province København by	No data	No data	...	No data	646986	645875	...	667228
NUTS 2006	DK012	Province Københavns omegn	No data	No data	...	No data	504634	504317	...	508183
NUTS 2006	DK013	Province Nordsjælland	No data	No data	...	No data	436570	440036	...	444215
NUTS 2006	DK014	Province Bornholm	No data	No data	...	No data	43445	43337	...	42659
NUTS 2006	DK021	Province Østsjælland	No data	No data	...	No data	228712	231150	...	233605
NUTS 2006	DK022	Province Vest- og sydsjælland	No data	No data	...	No data	577242	580361	...	587647
NUTS 2006	DK031	Province Fyn	No data	No data	...	No data	476580	478347	...	484346
NUTS 2006	DK032	Province Sydjylland	No data	No data	...	No data	707171	707504	...	715321
NUTS 2006	DK042	Province Østjylland	No data	No data	...	No data	792934	798671	...	820558
NUTS 2006	DK041	Province Vestjylland	No data	No data	...	No data	419853	421054	...	427174
NUTS 2006	DK050	Province Nordjylland	No data	No data	...	No data	577278	576807	...	580515

**Figure 4:** Population 1979-2009 in the NUTS3 of Denmark (both NUTS 2003 and 2006 version, according to the Danish National Statistic Institute)

Italian National statistical Institute proposes another example of time series handling. It provides information related to national territorial changes and its correspondence with regional (European) level (figure 5).

However, the Italian example may be considered as the best attempt because it provides much information to describe the change of administrative units: type of change, juridical texts and relation between versions of unit (genealogy). It allows also to analyse the effect of national administrative boundaries change on the NUTS geometry and hierarchy.

## Cartografia: confini amministrativi e dei sistemi locali del lavoro

Censimento 2001, 31 dicembre 2008 e 1 gennaio 2010

L'Istat fornisce la **versione generalizzata dei confini amministrativi** (Regioni, Province e Comuni) e dei **sistemi locali del lavoro**. Gli strati informativi sono costituiti da tre livelli gerarchici a copertura nazionale per i limiti di regione, provincia e comune.

I dati sono in formato shapefile; tale formato dati è stato reso pubblico già da parecchi anni ed utilizzato per lo scambio di dati in ambito GIS (Geographic Information System). I dati cartografici forniti sono nel sistema di riferimento ED\_1950\_UTM zona 32; il dettaglio tecnico della proiezione è riportato nel file apposito, associato a ciascun file geografico.

La scala dei dati non è certificabile uniformemente dall'Istat, in quanto le basi di acquisizione utilizzate provengono da fonti e scale differenti, che variano dal 1:5.000 in ambito urbano fino 1:25.000 in ambito extraurbano. I dati sono stati inoltre generalizzati e semplificati nelle forme geometriche, per renderne disponibile una versione da utilizzare agevolmente, per la creazione di cartografia simbolica o di riferimento a livello nazionale.

I file geografici di regioni, province e comuni, già pubblicati alla data del Censimento del 2001, sono stati aggiornati comprendendo le variazioni (territoriali e di nome) intercorse tra la data del Censimento 2001 e il 31 dicembre 2008 e successivamente al 1 gennaio 2010. Sono stati quindi acquisiti i codici e le denominazioni delle tre nuove province (Monza e della Brianza, Fermo e Barletta-Andria-Trani) e le nuove codifiche dei comuni ad esse appartenenti. Inoltre sono state acquisite anche altre variazioni comunali (si veda la **Struttura dei dati**). Per una più approfondita descrizione delle variazioni amministrative e territoriali intervenute successivamente alla data del Censimento del 2001 si può consultare la pagina web con i **codici dei comuni, delle province e delle regioni**.

Sono inoltre forniti i confini dei **Sistemi locali del lavoro** e delle **NUTS2** (Nomenclature of territorial units for statistics), che rappresenta l'articolazione ufficiale europea del territorio di livello 2 (Regioni e le Province autonome per l'Italia) finalizzata alla produzione di statistiche.

### descrizione dati

#### Struttura dei dati

#### confini amministrativi

Censimento 2001

- Regioni
- Province
- Comuni

31 dicembre 2008

- Regioni
- Province
- Comuni

1 gennaio 2010

- Regioni
- Province
- Comuni

#### altra cartografia

#### Sistemi locali del lavoro

Censimento 2001

- NUTS2

2008

#### per informazioni

Informazioni territoriali e sistema informativo geografico  
tel. 06 4673.4861  
email int@istat.it

Figure 5: Web Interface of Italian National statistical Institute

Furthermore the National Statistical Institutes, other government departments such as the interior ministers could publish documents related to national territorial changes. The Danish ministry of the interior and social affairs has published guide paper to understand the local boundaries reforms, which have consequences on the definition of NUTS3 and NUTS2 units (figure 6). This kind of document is not usually available.

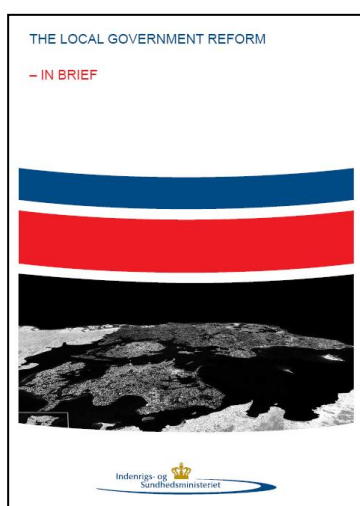


Figure 6: Explanation of the Danish Reform on Territorial units

## **1.4 Conclusion of the first part**

Building NUTS temporal database is very complex for the following main reasons:

- NUTS changes vary greatly from country to country due to their different structures;
- The available data are very heterogeneous;
- Data quality varies largely;
- Lack of good practice and experiences of handling territorial boundaries.

Based on a compilation of several sources and methods, we will propose formalization adapted to the specificity of NUTS.



## 2 Nuts changes knowledge: from elementary to systemic approach of territorial changes

The benchmarking of sources and experiences has showed the complexity of NUTS territorial changes. Following Swianczyny, (2000) who stated that: "In order to create a truly time integrative GIS, the focus has to change from spatial to temporal and from analyzing changes between events to the analysis of the change itself", we propose an appropriate approach to formalize the Nuts changes. This approach will be based on an explicit description of changes.

Based on the characteristics of NUTS, determined by regulations, we can distinguish the following changes: name, geometry, code and hierarchical level. These changes can occur at the same time because territorial changes are very complex. The changes analysis may be presented from two angles: elementary approach and systemic approach.

### 2.1 Elementary changes

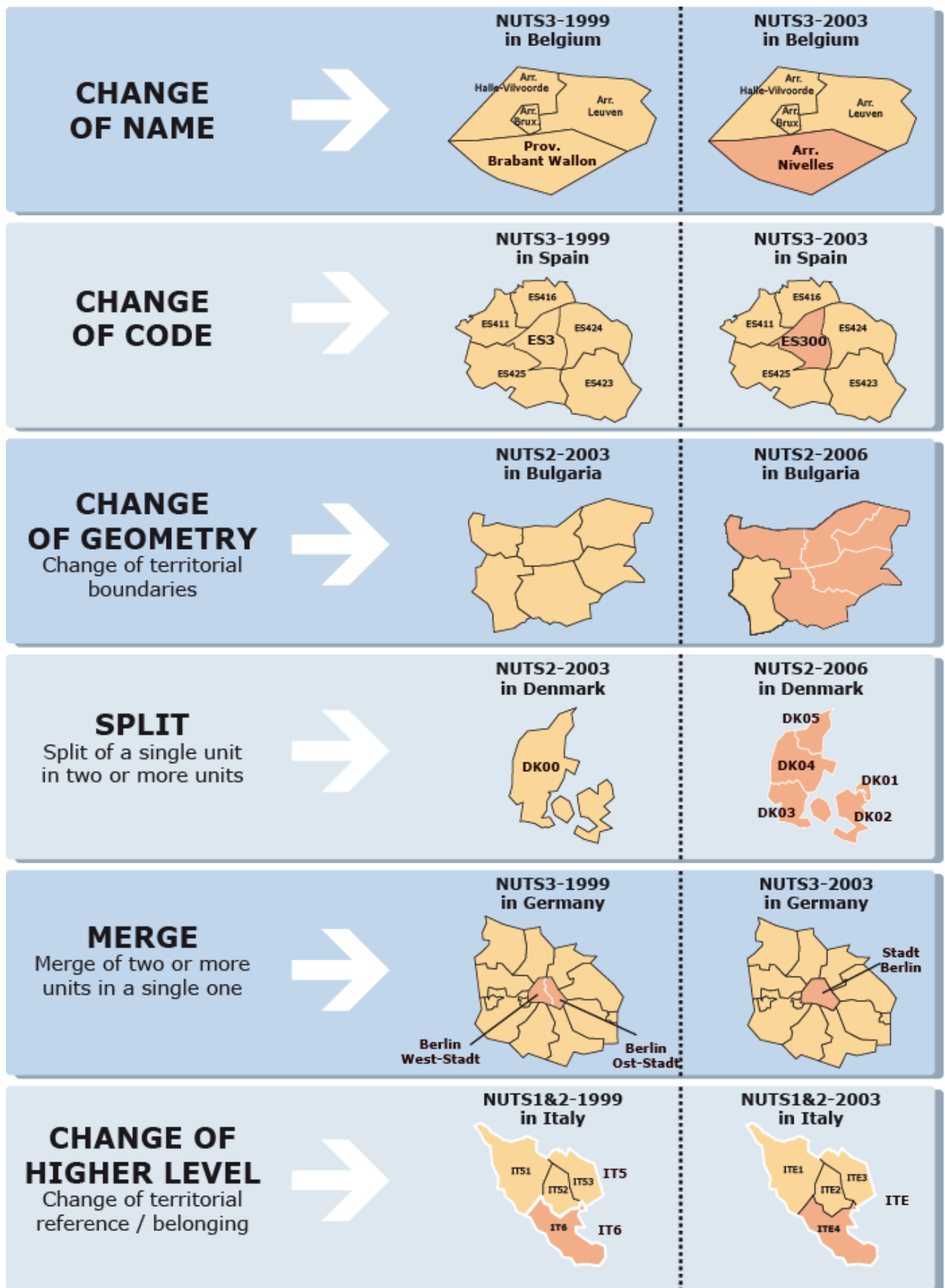
Elementary approach consists in describing the change of territorial units one undependably of the others (figure 7).

- Change of name: two cases can be distinguished. If the unit in question belongs to two levels (it is at the same time a NUTS 2 and a NUTS 3) the change of name can concern either one or the two levels.

1999: BE31 Brabant Wallon

2003: B310 Arrondissement Nivelles

- Change of code: it may result from different types of territorial modifications: political decision, territorial reorganisation. In the first case, we can list many changes of NUTS 2 level in 2003. In the second case, we point the code changing of Italian NUTS 2 and NUTS 3 units since 2003 due to regional reorganisation of NUTS 1.
- Change of geometry: It is the most complicated change type. Generally, the deformation of a spatial unit can be done in three different ways: the loss of area, the gain of area, or the redistribution of boundaries even while keeping the same area value. Most of the time, there is no relation between the different versions of the NUTS, like in Poland (figure 9). This kind of situation makes the harmonisation very difficult to implement.
- Change of hierarchy: it consists to the change of the territorial reference or of belonging. As shown by the figure 7, Italian Nuts 2 level have changed their Nuts 1 units of belonging because of territorial reforms of nuts 1 level in 2003.

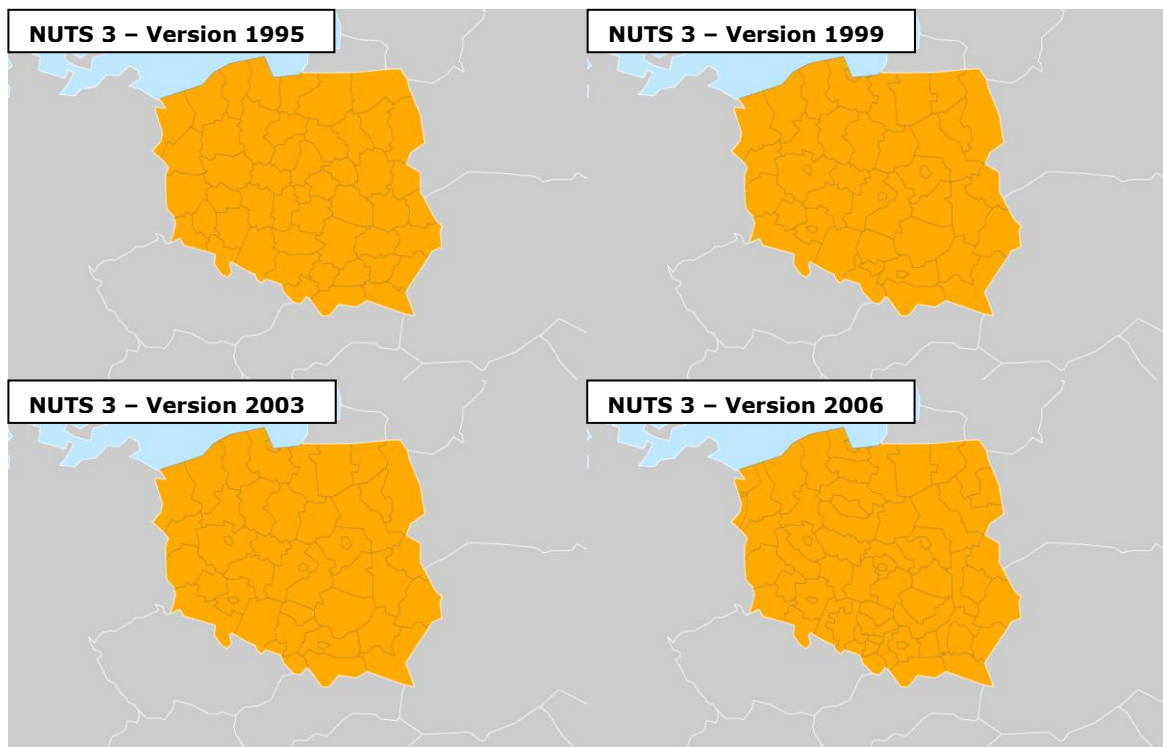


**Figure 7:** Examples of elementary Nuts changes

CODE 1999	CODE 2003	Name 1999	Name 2003	Status
ES3	ES3	Comunidad de Madrid	Comunidad de Madrid	No change
ES3	ES30	Comunidad de Madrid	Comunidad de Madrid	Changed
ES3	ES300	Comunidad de Madrid	Comunidad de Madrid	Changed
IT1	ITC	Nord Ovest	Nord Ovest	Changed
IT11	ITC1	Piemonte	Piemonte	Changed
IT111	ITC11	Torino	Torino	Changed
IT112	ITC12	Vercelli	Vercelli	Changed
IT113	ITC13	Biella	Biella	Changed
IT114	ITC14	Verbano-Cusio-Ossola	Verbano-Cusio-Ossola	Changed
IT115	ITC15	Novara	Novara	Changed
IT116	ITC16	Cuneo	Cuneo	Changed
IT117	ITC17	Asti	Asti	Changed
IT118	ITC18	Alessandria	Alessandria	Changed

**Figure 8:** Examples of changing of unit's code in Italy and Spain from 1999 to 2003

Considering the characteristics of Nuts units which are mentioned, limiting the investigation to basic (elementary) changes does not allow to reconstruct genealogy of Nuts. To achieve this, we have considered nuts structure as a system and we have focused on the relationship between changes.



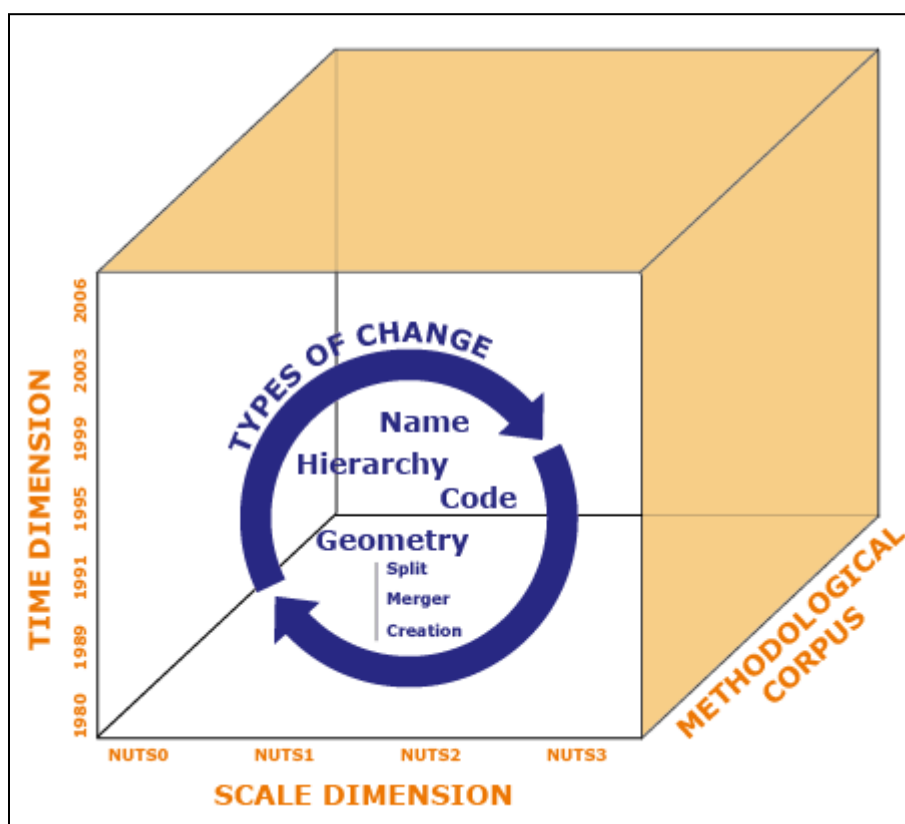
**Figure9:** Complex geometry change in Polish nuts 3 units

## 2.2 Systemic approach of NUTS changes

Because the formalization of Nuts changes is complex and has to take into account several parameters (type of changes, temporal period and scalar dimension), we propose a cubic model which emphasizes the relationships between these parameters. This means that the result of territorial modifications depends on the type of changes (name, code, and geometry), the period of time and the territorial level. Thus, we used the concept of systemic approach.

The systemic conception emphasizes the relationships between the changes:

- A change affecting a unit may have implications on the other units.
- A change happening on a given level may have implications on the other levels.



**Figure 10:** Cube structure of NUTS formalisation

We demonstrate our approach through the analysis of the example of Italian Nuts between 1995 and 2006 (figure 10):

Concerning the temporal dimension, two orders can be distinguished:

- The period of time determines the degree of discontinuity of the data sets. Indeed, the extension of the period increases the discontinuity because of the complexity of changes that may have occurred. In the case of Italian Nuts, if we consider the whole period (1995-2006), we can see a big discontinuity in the data sets. However, the data set will be complete between 2003 and 2006.
- The building of time series data could be considered in either a prospective or retrospective territorial approach. The prospective view consists in transposing old data sets onto a recent version of Nuts (data 1995 onto Nuts 2006 for example). However, the retrospective view consists in transposing recent data sets onto old Nuts versions (data 2006 to Nuts 1995). Each approach requires a

different methodology. For example, 2003 version data should be disaggregated to be integrated in Italian Nuts 1 level 1999 version. However, the 1995 version data should be aggregated.

- As for the Scalar dimension, it is linked to the hierarchical structure of Nuts (Nuts 1 level is subdivided into Nuts 2 level which is in turn subdivided into Nuts 3 level). In fact, the changes which occur in higher levels (1 and 2) have various consequences on lower territorial levels. As it was shown by the figure 7, the territorial reform of Italian Nuts 1 level in 2003, consisting in merging and changing codes of units, has caused a change of codes of Nuts 2 and Nuts 3 units. Moreover, reforms of higher Nuts levels (Nuts 1 and Nuts 2) could have more complex implications on lower levels. The creation of 5 new Nuts 2 units in Denmark in 2003, by splitting DK00, has caused very complex territorial reorganization on Nuts 3 level units (Figure 7).

Regarding Relationships between changes, the change of geometry is a determining factor in the time series data building process. On the whole, three types of unit spatial changes can be identified: the loss of area, the gain of area and deformation (which means territorial boundaries redistribution without loss of area). Based on these primary types of changes, we have developed a conceptual corpus to describe further types of changes (dictionary of changes). The dictionary of changes aims to answer the following questions: what happened? How did it happen? And what were the results?

For example, the Danish territorial reforms in 2003 could be described as follows:

Nuts 1 level: there are no changes

Nuts 2 level:

- The Split of DK00 (change of geometry)
- Official disappearance of DK00
- Creation of 5 new Nuts 2 units: DK01, DK02, DK03, DK04 and DK05

Nuts 3 level:

- Change of code which means change of belonging to a superior unit (hierarchy): Funy DK008 (2003) and DK031 (2006), Bornholm DK007 (2003) and DK014 (2006)
- Complex changes of geometry for the rest of units which have caused the disappearance of 12 units and the creation of 10 new units

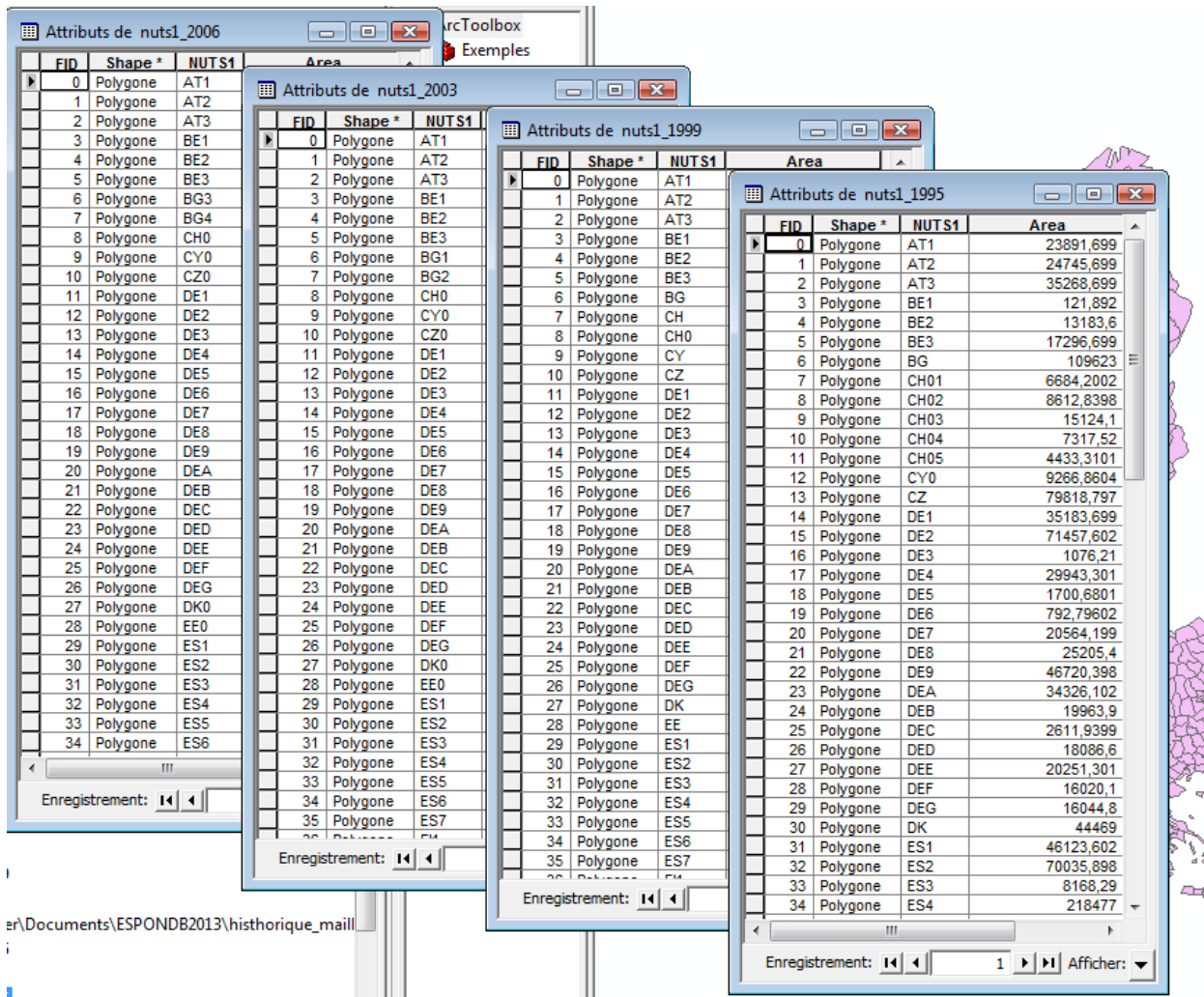
### **3 Building historical database of territorial changes: from conceptual approach to operational solutions**

This process of formalization of Nuts changes is not an end in itself. Aiming to build continuous time series data, its first objective is to understand how spatial units have been changed. Information describing Nuts changes represent a very useful metadata. In this final section of this technical report, we will examine how these results could be presented to the users. A first application made on cohesion reports will also be presented.

The results of this exploration may be presented in different ways depending on the users' needs. The examples that we present illustrate the progress of the complexity of the issue of nuts changes formalization. Location of change, identification of change and genealogy (lineage) of spatial units are the most important information that could be allowed to the users at this stage of work.

#### **3.1 Improved Snapshot model (presentation)**

One of the simplest spatio-temporal data models is the snapshot model (Armstrong 1988). Temporal information was incorporated into this spatial data model by time-stamping layers. Every layer shows the states of NUTS at different times without explicit relation between versions. Changes and genealogy cannot be depicted. This prevents harmonized database to be built.



**Figure 11:** Snapshot of NUTS versions between 1995 and 2006 (example of Nuts 1 level)

This method could be improved by location of changes occurred. A binary code was attributed to the state of Nuts: 0 for stability and 1 in case of change. This simple coding can identify potential discontinuities in the time series. The figure 11 shows an extract of tables covering all Nuts levels.

Code 2006	NUTS0	NUTS level 1	NUTS level 2	NUTS level 3	Change	Change since 2003
DK0		DANMARK			same	0
DK01			Hovedstaden		changed	1
DK011				Byen København	changed	1
DK012				Københavns omegn	changed	1
DK013				Nordsjælland	changed	1
DK014				Bornholm	changed	1
DK02			Sjælland		changed	1
DK021				Østsjælland	changed	1
DK022				Vest- og Sydsjælland	changed	1
DK03			Syddanmark		changed	1
DK031				Fyn	changed	1
DK032				Syddjylland	changed	1
DK04			Midtjylland		changed	1
DK041				Vestjylland	changed	1
DK042				Østjylland	changed	1
DK05			Nordjylland		changed	1
DK050				Nordjylland	changed	1

**Figure 12:** Extract of table of location of changes: Danish Nuts between 2006 and 2003

The figure 12 shows more developed stage of change describing process. Identifying kinds of change and the consequences, even in scalar dimension and relationship between changes, were added to the location of change.

- The column Check: change/no change
- The column Change: identify the initial change
- The column Life: indicates if units exists (E) or did not exist (D: deleted)
- The column Hierarchy : change of geometry (0/1)
- The column Geometry: specifies the change of geometry of the unit

Semantic describing changes (code) should be improved. A table of Metadata will be established when this semantic description will be consolidated.

Code 2003	Code 2006	Country	NUTS level 1	NUTS level 2	NUTS level 3	CHECK	CHANGE	LIFE	HIERARCHY	GEOMETRY
	DK01			Hovedstaden		changed	GEOM	N	0	GEOM+
	DK011				Byen København	changed	GEOM	N	0	GEOM+
	DK012				Københavns omegn	changed	GEOM	N	0	GEOM+
	DK013				Nordsjælland	changed	GEOM	N	0	GEOM+
DK007	DK014				Bornholm	changed	GEOM	N	0	GEOM+
	DK02			Sjælland		changed	GEOM	N	0	GEOM+
	DK021				Østsjælland	changed	GEOM	N	0	GEOM+
	DK022				Vest- og Sydsjælland	changed	GEOM	N	0	GEOM+
	DK03			Syddanmark		changed	GEOM	N	0	GEOM+
DK008	DK031				Fyn	changed	GEOM	N	0	GEOM+
	DK032				Sydjylland	changed	GEOM	N	0	GEOM+
	DK04			Midtjylland		changed	GEOM	N	0	GEOM+
	DK041				Vestjylland	changed	GEOM	N	0	GEOM+
	DK042				Østjylland	changed	GEOM	N	0	GEOM+
	DK05			Nordjylland		changed	GEOM	N	0	GEOM+
	DK050				Nordjylland	changed	GEOM	N	0	GEOM+
DK00				Danmark		changed	GEOM	D	0	0
DK001					København og Frederiksberg kommuner	changed	GEOM	D	0	0
DK002					Københavns amt	changed	GEOM	D	0	0
DK003					Frederiksborg amt	changed	GEOM	D	0	0
DK004					Roskilde amt	changed	GEOM	D	0	0
DK005					Vestsjællands amt	changed	GEOM	D	0	0
DK006					Storstrøms amt	changed	GEOM	D	0	0
DK009					Sønderjyllands amt	changed	GEOM	D	0	0
DK00A					Ribe amt	changed	GEOM	D	0	0
DK00B					Vejle amt	changed	GEOM	D	0	0
DK00C					Ringkøbing amt	changed	GEOM	D	0	0
DK00D					Århus amt	changed	GEOM	D	0	0
DK00E					Viborg amt	changed	GEOM	D	0	0
DK00F					Nordjyllands amt	changed	GEOM	D	0	0

Figure 13: Extract of table of identification of changes: Danish Nuts between 2006 and 2003

### 3.2 The space-time composite model: reconstructing genealogy of Nuts versions

This model was proposed by Langran (1992). It consists to decompose NUTS, geometries through time by intersecting different versions (1995-1999-2003-2006). Small spatio-temporal entities (polygons) are created as the results of this intersection. It represents the lowest common denominator.

As it is shown by the figure 13 the intersection of geometries of Danish NUTS 3 2003 and 2006 versions results 22 polygons. The belonging to NUTS is defined as a temporal attribute. For example, the polygon n°11 (table and selected in the maps) belong to DK00E unit in 2003 and to DK041 unit in 2006. In fact, genealogy of units can be deduced and may be represented by a graph.



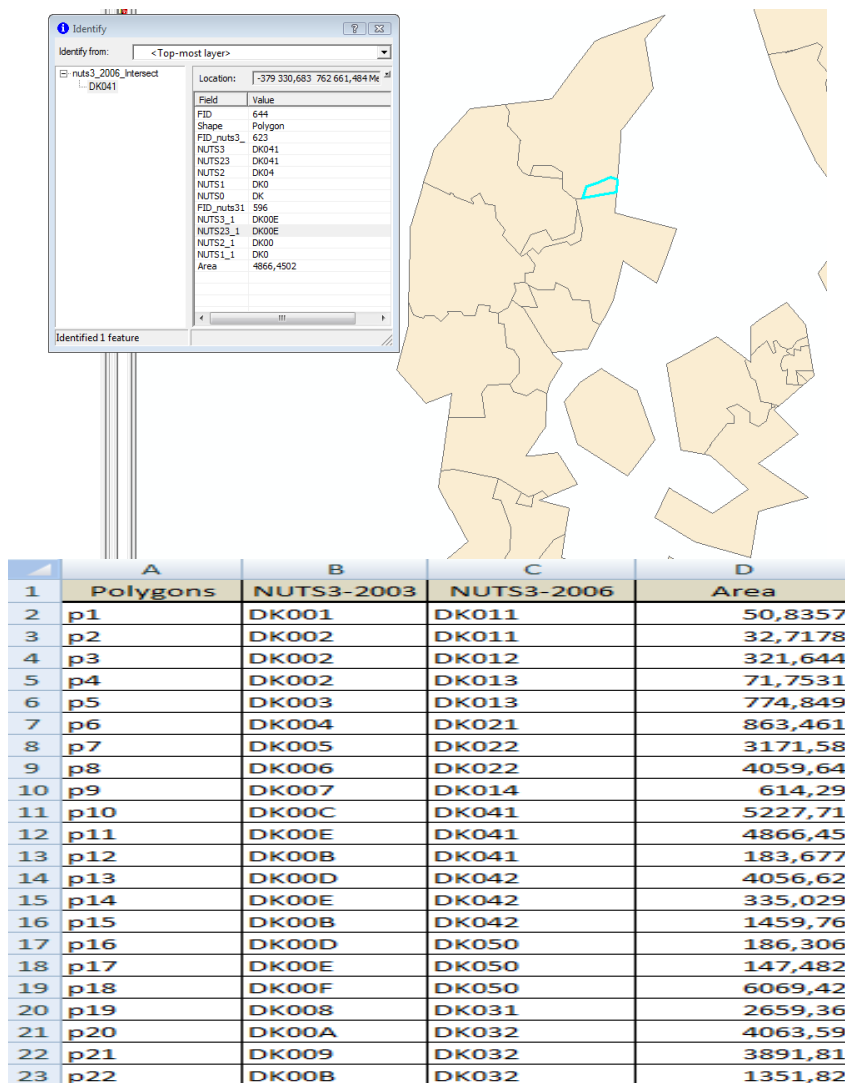


Figure 14: Space time composite model

We have improved this cartographic method by quantifying lineage of Nuts. Information collected through the exploration of sources has allowed us to establish a relationship between spatial units. We tried to quantify this genealogy by calculating the proportion of area transferred in case of change. The proportion of the population transferred will also be tested. This method is very useful for estimating missing data because of territorial changes

The figure 14 illustrates the table built. It does not yet cover all the ESPON area because of lack of accurate data especially for the new member countries.

NAME	code 2006	% Geom	code 2003	% Geom	code 1999	% Geom	code 1995	NAME
Denmark	0	0	DK00	100	DK00	100	DK00	Danmark
	DK01	4,2	DK00	4,2	DK00	4,2	DK00	Hovedstaden
	DK02	18,2	DK00	18,2	DK00	18,2	DK00	Sjælland
	DK03	26,9	DK00	26,9	DK00	26,9	DK00	Syddanmark
	DK04	36,3	DK00	36,3	DK00	36,3	DK00	Midtjylland
	DK05	14,4	DK00	14,4	DK00	14,4	DK00	Nordjylland

Figure 15: Extract of table of genealogy of Nuts: Danish Nuts2 level between 2006 and 1995

Another simple approach is to tag every object (NUTS) with a pair of timestamps, one for the time of creation and one for the time of cessation. Current objects have their cessation time given by a special value "NOW", "CURRENT", or "NULL".

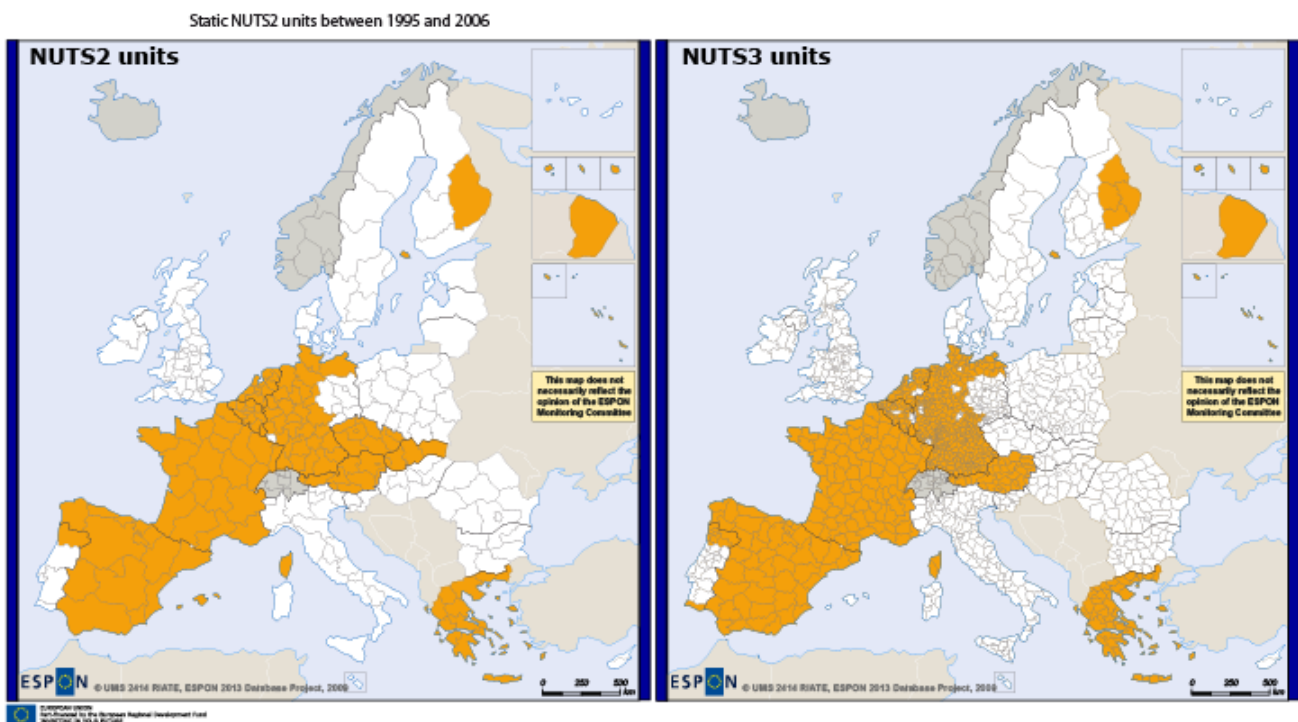
To conclude this section, we emphasize that these applications were developed in the following time periods: 2006-2003; 2006-1999; 2006-1995, 2003-1999; 2003-1995 and 1999-1995.

### 3.3 Towards a NUTS changes mapping

A cartographic display of the results of formalization would be very useful for the better understanding of Nuts territorial changes. However the visualization of changes is as complex as the formalization and it depends on all parameters presented in the previous section. We propose the following thoughts.

#### 3.3.1 A general view

The map of static units distinguishes static units (no change observed) from changed units as shown by the figure 15. Yellow units are static. However, white units have been changed since 1995.



**Figure 16:** Static NUTS 2 and Nuts 3 units between 1995 and 2006 (all criteria).

### 3.3.2 *Typology of changes*

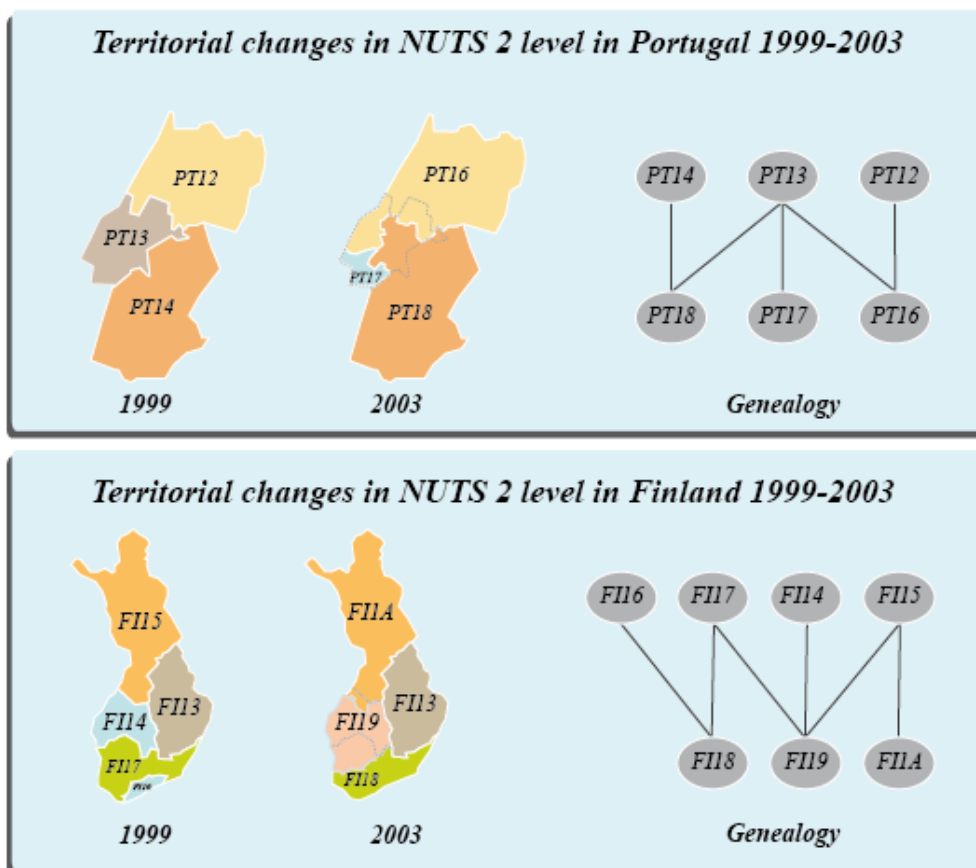
These maps give more information about changes. It aims to emphasize on the diversity of changes. The most important kinds of changes are: name, code, geometry and hierarchy.

### 3.3.3 *Maps by times intervals*

The interval of time (period) is a very important parameter for building time series database. Naturally, the enlargement of the covered period makes the building process more difficult. Mapping changes by period of time could help users in their building time series process.

### 3.3.4 *Examples of lineage (genealogy) visualization*

The attempt is to propose a graphic visualization of the units' genealogy. We selected the most demonstrative examples in order to emphasize in the complexity of Nuts changes. These patterns are in an early stage implementation and should be improved.



**Figure 17:** examples of NUTS genealogy visualization

### 3.4 Tables structure of NUTS changes

Table structuring of NUTS changes is a very challenging topic. The aim of this topic is to provide a very friendly using data about territorial changes. The complexity of NUTS changes requires simplifying the information to be useful and comprehensive by a large public dealing with time series data implementation.

We propose a several ways of structuring change tables. These tables are complementary and propose different visions of territorial changes that meet to needs of many users in terms of spatiotemporal databases.

#### 3.4.1 Changes location tables

It is a binary description of NUTS situation between each version: 1995-1999, 1999-2003 and 2003-2006 (more time intervals combinations are possible). The aim of this table is to locate changes regarding all analysis criteria (name, code, geometry and belonging to a higher level).

Code 2006	NUTS0	NUTS level 1	NUTS level 2	NUTS level 3	Change since 2003
DK	DANMARK				0
DK0		DANMARK			0
DK01			Hovedstaden		1
DK011				Byen København	1
DK012				Københavns omegn	1
DK013				Nordsjælland	1
DK014				Bornholm	1
DK02			Sjælland		1
DK021				Østsjælland	1

**Figure 18:** extract of change location table (Danish Nuts between 2003 and 2006)

#### 3.4.2 Changes typology tables

These tables give more detailed information about changes. The identification of nuts changes is very helpful for understanding territorial changes. It could contribute to the database harmonization. Regarding to nuts changes formalization presented in the previous section (2.2), we identified four kinds of change: name, code, geometry and belonging to a higher level.

Code 2003	Code 1999	NUTS level 1	NUTS level 2	NUTS level 3	OP	CHANGE	LIFE	HIERARCHY	GEOMETRY
DE3	DE3	BERLIN				0	0	E	0
DE30	DE3		Berlin		CODE	CODE	E		1
	0 DE300			Berlin	MERGER	NO	D		0 NO
DE301	0			Berlin-West, Stadt	MERGER	NO	N		0 GEOM+
DE302	0			Berlin-Ost, Stadt	MERGER	NO	N		0 GEOM+
DE4	DE4	BRANDENBURG			SPLIT	H	D		1 NO
	0 DE4		Brandenburg		SPLIT	NO	D		0 NO
DE41	0			Brandenburg - Nordost	SPLIT	NO	N		0 GEOM+
DE42	0			Brandenburg - Südwest	SPLIT	NO	N		0 GEOM+
DE411	DE403			Frankfurt (Oder), Kreisfreie Stadt	SPLIT	CODE	E		2
DE412	DE405			Barnim	SPLIT	CODE	E		2
DE413	DE409			Märkisch-Oderland	SPLIT	CODE	E		2
DE414	DE40A			Oberhavel	SPLIT	CODE	E		2
DE415	DE40C			Oder-Spree	SPLIT	CODE	E		2
DE416	DE40D			Ostprignitz-Ruppin	SPLIT	CODE	E		2
DE417	DE40F			Prignitz	SPLIT	CODE	E		2
DE418	DE40I			Uckermark	SPLIT	CODE	E		2

**Figure 19:** extract of change typology table (German Nuts between 2003 and 1999)

### 3.4.3 "Units life" tables

The aim of these tables is to provide a general view of the official existence of units in each nuts version. The use of legal definition of statistical units (Official journal of the European Union) avoided the several interpretations of the notion of "life". We regroup in one table all units from 1995 to 2006 and we indicate each version, by binary description, if it exists or it was deleted.

Code 2003	Code 1999	NUTS level 1	NUTS level 2	NUTS level 3	OP	LIFE
DE3	DE3	BERLIN			0	1
DE30	DE3		Berlin		CODE	1
	0 DE300			Berlin	MERGER	0
DE301	0			Berlin-West, Stadt	MERGER	1
DE302	0			Berlin-Ost, Stadt	MERGER	1
DE4	DE4	BRANDENBURG			SPLIT	0
	0 DE4		Brandenburg		SPLIT	0
DE41	0			Brandenburg - Nordost	SPLIT	1
DE42	0			Brandenburg - Südwest	SPLIT	1
DE411	DE403			Frankfurt (Oder), Kreisfreie Stadt	SPLIT	1
DE412	DE405			Barnim	SPLIT	1
DE413	DE409			Märkisch-Oderland	SPLIT	1
DE414	DE40A			Oberhavel	SPLIT	1
DE415	DE40C			Oder-Spree	SPLIT	1
DE416	DE40D			Ostprignitz-Ruppin	SPLIT	1
DE417	DE40F			Prignitz	SPLIT	1
DE418	DE40I			Uckermark	SPLIT	1

**Figure 20:** extract of "Units life" table (German Nuts between 2003 and 1999)

### 3.4.4 "genealogy units" tables

It is the most complicated table and the most innovative. It provides a conversion keys between nuts versions in order to harmonize temporal databases. To emphasize on the lineage of units, we have defined tow possible kinds of proportionalities: area and population. The proportionality of area transferred is very useful for the conversion of environmental indicators (figure 15). However the proportionality of population is useful for social economic indicators conversion.

## 4 Example of time series data building process: the European regional cohesion indicators

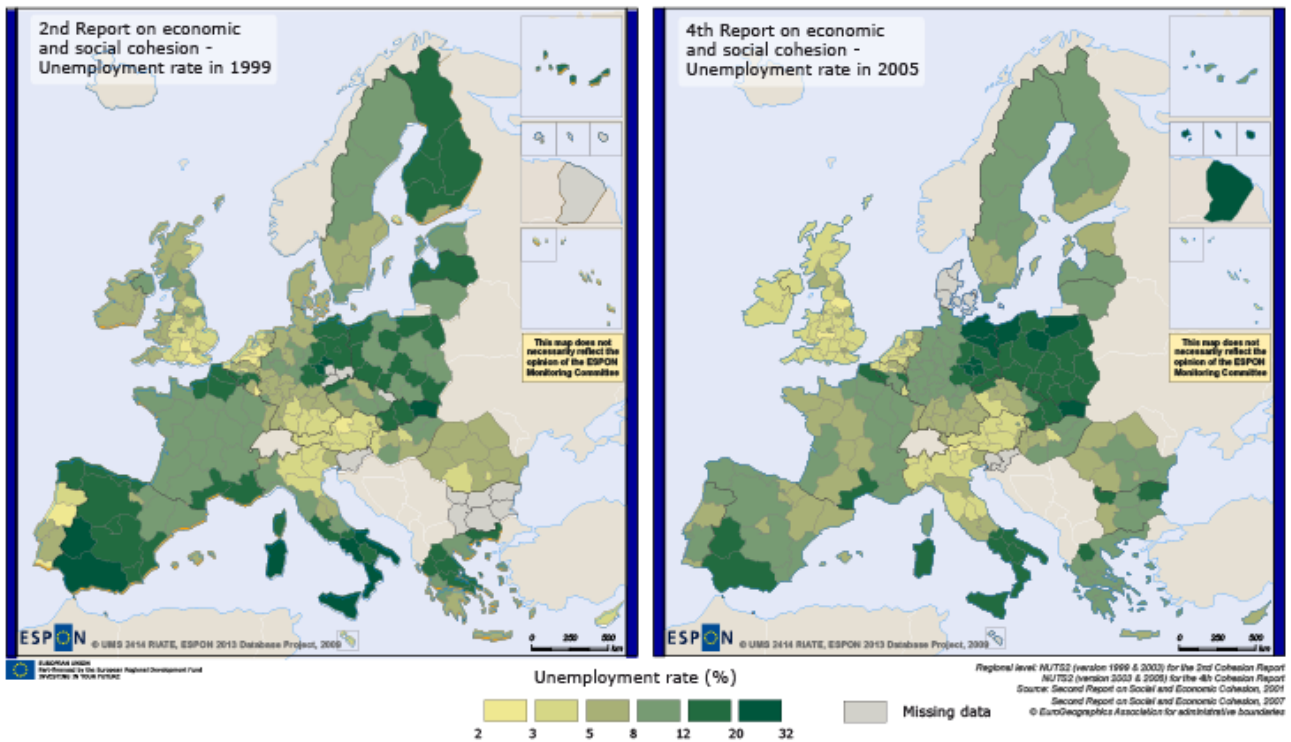
This section presents a harmonization use-case based on the formalization previously developed. It emphasizes on the contribution of time series data building and harmonization to the thematic analysis. The analysis of the European regional cohesion spatio-temporal evolution offers a very interesting example of methodological demonstration of the time-series building process. The analysis focused on data attached to the three latest reports on territorial cohesion, which include all methodological difficulties such as missing values and mixture of NUTS versions.

### 4.1 Data collection and temporal series maps

Based on the results of the exploration of nuts changes, the first step of the time series data building process is to collect data and to identify the latest version of nuts. The aim of this work is to produce thematic maps based on the 2003 and 2006 Nuts version (figure 16 and figure 17). These data sets were delivered at ESPON Coordination Unit in November (ESPOON DB Update) and were presented at ESPON seminar in Malmö (2-4 December 2009).

id	level	name	NUTS_VER source	
			1999	2006
	<b>TEMPORAL_START</b>			
	<b>TEMPORAL_END</b>			
BE	NUTS0	België/Belgique	2006	1
BE10	NUTS2	Région de Bruxelles-Capitale / Brussels Hoofdstedelijk Gewest	2006	1
BE2	NUTS1	Vlaams Gewest	2006	1
BE21	NUTS2	Prov. Antwerpen	2006	1
BE22	NUTS2	Prov. Limburg (BE)	2006	1
BE23	NUTS2	Prov. Oost-Vlaanderen	2006	1
BE24	NUTS2	Prov. Vlaams-Brabant	2006	1
BE25	NUTS2	Prov. West-Vlaanderen	2006	1
BE3	NUTS1	Région Wallonne	2006	1
BE31	NUTS2	Prov. Brabant Wallon	2006	1
BE32	NUTS2	Prov. Hainaut	2006	1
BE33	NUTS2	Prov. Liège	2006	1
BE34	NUTS2	Prov. Luxembourg (BE)	2006	1
BE35	NUTS2	Prov. Namur	2006	1
BG	NUTS0	Bulgaria	2006	1

**Figure 21:** Nuts version identification of spatial units used by the 4<sup>th</sup> Cohesion report



**Figure 22:** Mapping the Unemployment rate in 1999 and 2005

This task aims to provide data users and policy makers with maps allowing the comparison between two situations of unemployment in the European regions. Mapping is an important added value.

## 4.2 Building a harmonized temporal data base for the European regional cohesion analysis

The harmonization of time series data is a fundamental step for a more advanced temporal analysis and mapping. The built of time-series data can be done by two complementary approaches: The ESTI framework for estimating missing values is one of them<sup>1</sup> and the modelling of NUTS changes.

The methodology consisted on these main steps:

- Based on the dictionary of changes of territorial units, we defined the versions of NUTS. The 2003 NUTS version was considered as a reference. The data were transposed into this version.
- Missing data were estimated using the ESTI framework methods
- Selected indicators cover the following sectors: economy, demography, unemployment and GDP per capita
- We proceed first to a hierarchical cluster on the second report data and we allocate the results to third and fourth reports data

These steps will be developed in the next sections.

#### 4.2.1 Cohesion reports datasets exploration: missing values and lack of compatibility between NUTS versions

The exploration of data at NUTS 2 level shows the lack of compatibility of NUTS between datasets. Indeed, the collected data is a mixture of nuts versions from 1995 to 2006. If we look at the NUTS 2 level, we can observe many kinds of change. We illustrate some of them by the following tables.

ID (CR2)	name	ID (CR3)	name
IT11	PIEMONTE	ITC1	Piemonte
IT12	VALLE D'AOSTA	ITC2	Valle d'Aosta/Vallée d'Aoste
IT13	LIGURIA	ITC3	Liguria
IT20	LOMBARDIA	ITC4	Lombardia
IT31	TRENTINO-ALTO ADIGE	IT31	<i>Trentino-Alto Adige</i>
		ITD1	Provincia Autonoma Bolzano/Bozen
		ITD2	Provincia Autonoma Trento
IT32	VENETO	ITD3	Veneto
IT33	FRIULI-VENEZIA GIULIA	ITD4	Friuli-Venezia Giulia
IT40	EMILIA-ROMAGNA	ITD5	Emilia-Romagna
IT51	TOSCANA	ITE1	Toscana
IT52	UMBRIA	ITE2	Umbria
IT53	MARCHE	ITE3	Marche
IT60	LAZIO	ITE4	Lazio
IT71	ABRUZZO	ITF1	Abruzzo
IT72	MOLISE	ITF2	Molise
IT80	CAMPANIA	ITF3	Campania
IT91	PUGLIA	ITF4	Puglia
IT92	BASILICATA	ITF5	Basilicata
IT93	CALABRIA	ITF6	Calabria
ITA	SICILIA	ITG1	Sicilia
ITB	SARDEGNA	ITG2	Sardegna

**Figure 23:** The lack of compatibility between Italians NUTS2 units between 1999 and 2003.

ID (CR2)	name	ID (CR3)	name
FI14	VÄLI-SUOMI	FI18	Etelä-Suomi
FI15	POHJOIS-SUOMI	FI19	Länsi-Suomi
FI16	UUSIMAA (SUURALUE)	FI1A	Pohjois-Suomi
FI17	ETELÄ-SUOMI		

**Figure 20:** A complicated territorial change of NUTS2 level in Finland

Concerning the missing values, we can distinguish between cases caused by changes of NUTS and cases of non-available data. Depending of these types we use the adequate framework of estimation.





Codes_Unit	Indicator 1	.....	Indicator 13
AT1-CR2 AT11-CR2 AT12-CR2 AT13-CR2 AT2-CR2			
AT1-CR3 AT11-CR3 AT12-CR3 AT13-CR3 AT2-CR3			
AT1-CR4 AT11-CR4 AT12-CR4 AT13-CR4 AT2-CR4			

**Figure 25:** Theoretical structure of the complete temporal database based on 2, 3 and 4 cohesion reports.

CR	id	NUTS_VER	GDP_head_av	Emp_agri	Emp_ind	Emp_serv	Unemp	Unemp_LT	Pop_tot
CR2	BE10	2006	170,3	0,21	13,42	86,38	14	62,0	952
CR2	BE21	2006	137,5	2,15	28,07	69,79	6,5	53,2	1637
CR2	BE22	2006	109,4	2,09	34,52	63,39	7	53,7	782
CR3	BE10	238,5	0,1	13,1	86,9	14,5	55,1	17,8	65,4
CR3	BE21	135,9	1,2	29,7	69,2	5,5	44,0	17,2	65,9
CR3	BE22	98,7	1,6	32,9	65,5	5,3	32,5	17,4	68,8
CR4	BE10	248,3	0,2	11,1	88,7	16,3	56,4	18,2	66,1
CR4	BE21	144,5	1,7	28,4	69,9	6,2	44,0	16,8	65,7
CR4	BE22	101,5	1,8	32,2	66,0	7,1	44,2	16,5	68,3
CR4	BE23	111,0	2,0	28,4	69,6	4,9	37,1	16,4	65,8

**Figure 26:** An extract of the complete temporal database based on 2, 3 and 4 cohesion reports.

### 4.3 Spatio-temporal analysis of the European regional cohesion

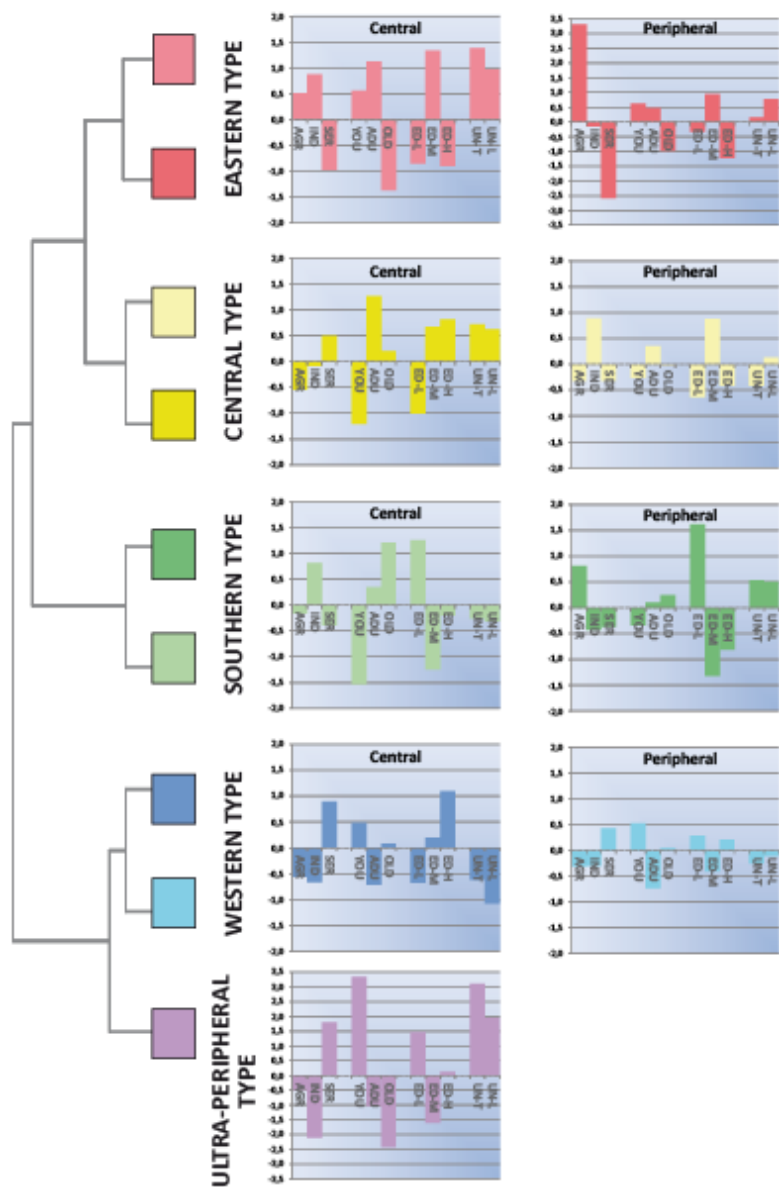
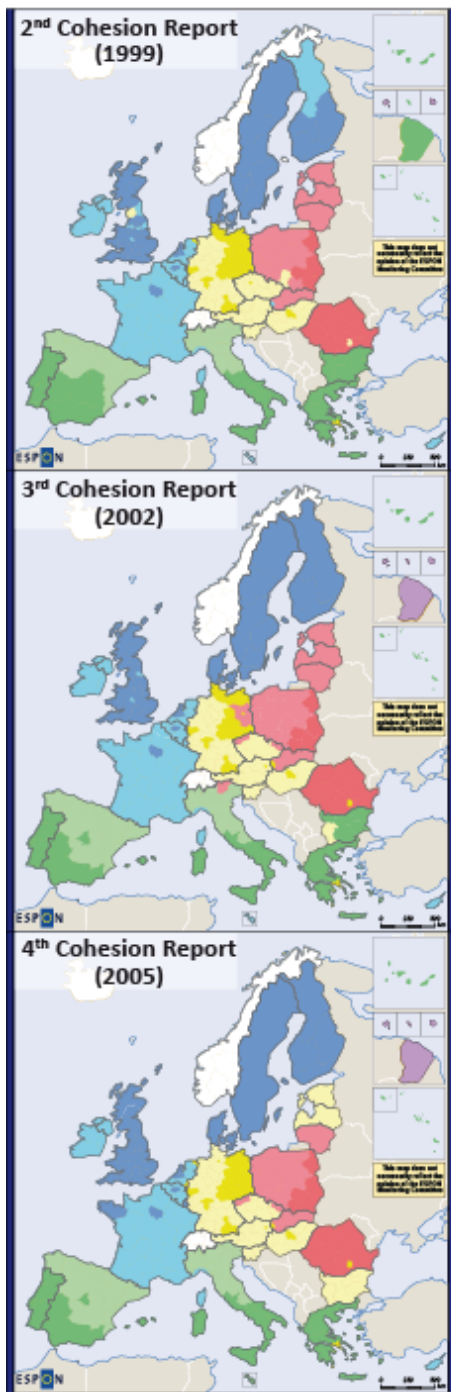
This section shows that it is possible to analyze the evolution of spatial patterns of regional inequalities with changing territorial units, through the example of a cross analysis of statistical annexes of 2nd, 3rd and 4th Cohesion Reports. The built temporal dataset allows implementing a temporal hierarchical cluster analysis.

The territorial cohesion of the European regions is organised along four main groups. Each group can be either central or peripheral (Figure 27):

- North-western European type:
  - Central: it is characterized mainly by: a very high education level, a very high rate of the service sector employments and a high rate of elderly people.

- Peripheral: it is characterised by a relative decrease in the level of education and service sector employments.
- Ultra-peripheral: it is characterised by a very high rate of young population. There is also an important decrease in education and service sector employments levels.
- Southern European type:
  - Central: it is characterized mainly by a high rate of elderly people, an important rate of industry sector employment and a low education level.
  - Peripheral: it is characterised by a relative decrease in the rate of elderly people and an increase in the rate of the agriculture sector employment.
- Central European type:
  - Central: it is characterized mainly by an important rate of adult population, a high education rate and the primacy of service sector employments rate.
  - Peripheral: it is characterised by the decrease the rate of high education level and an increase in the rate of the industry sector employment.
- Eastern European type:
  - Central: it is characterised by an important rate of the industry and the agriculture sectors employment and an important rate of adult population.
  - Peripheral: it is characterised by the primacy of the rate of the agriculture sectors employment and the increase of the rate of young population rate.

This organization is very stable through time. Indeed, no major changes happened.



Indicators contented in the classification, from the left to the right of the profile:

- AGR: Employment in agriculture sector
- IND: Employment in industry sector
- SER: Employment in service sector
- YOU: Share of youngs (0-14)
- ADU: Share of actives (15-64)
- OLD: Share of old (65+)
- ED-L: Share of active with low education level
- ED-M: Share of active with medium education level
- ED-H: Share of active with high education level
- UN-T: Unemployment rate
- UN-L: Long term unemployment rate

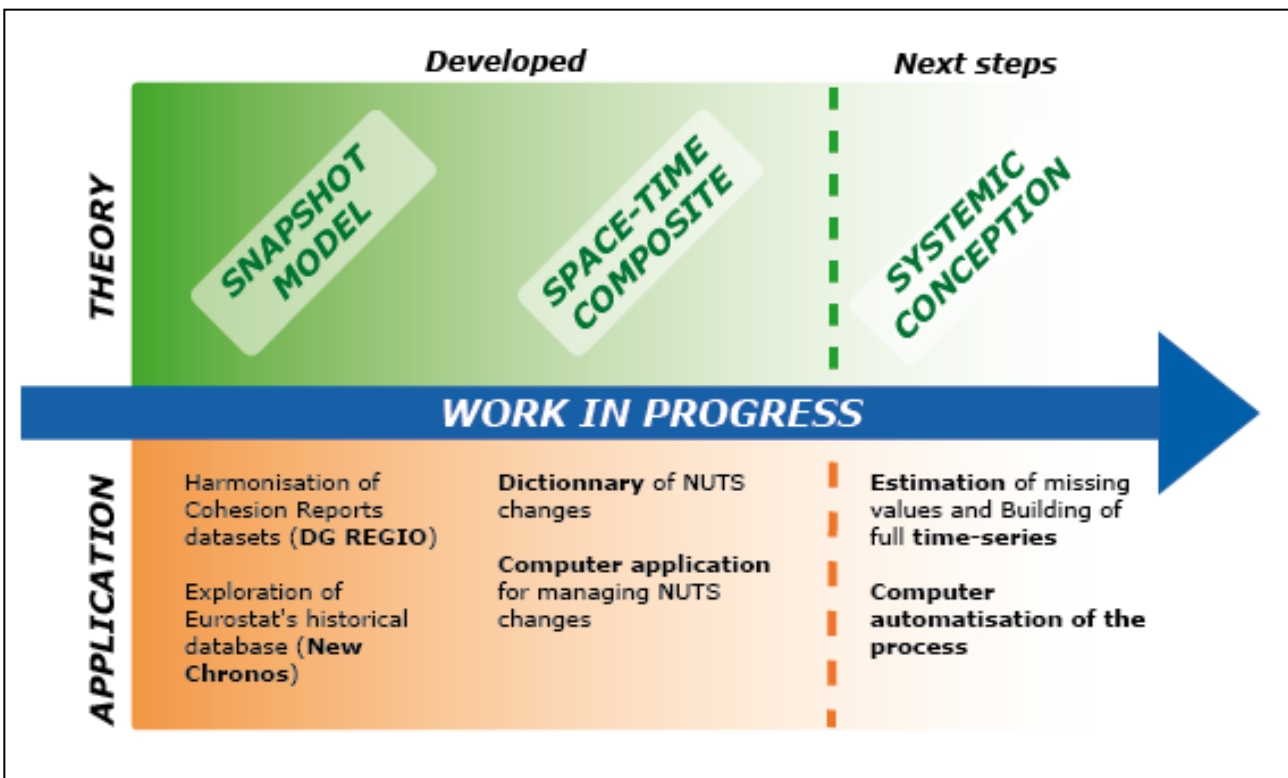
**Figure 27:** An extract of the complete temporal database based on 2, 3 and 4 cohesion reports.

## Conclusion

The contribution of the change analysis based-approach to the topic of time-series is very important. The theoretical formalization and modelling allow the building of a complete temporal database (figure 24). They provide a key of conversion between NUTS versions. This formalization is not a normative approach, but rather it requires an improvement.

The improvement perspectives could be oriented in the following directions:

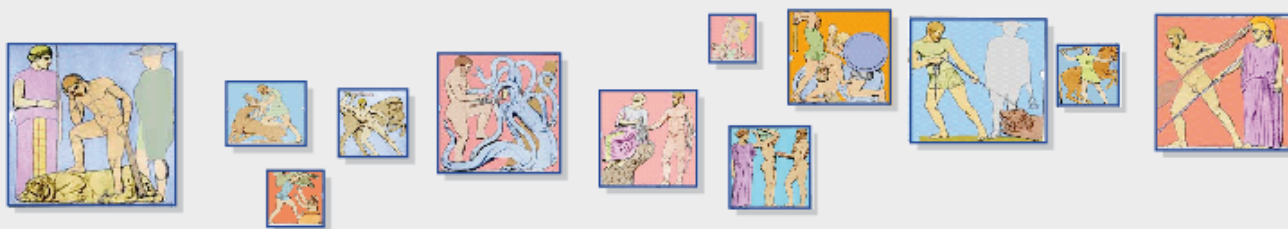
- The enlargement of the temporal period to the beginning of 1980. The archived New-Cronos database of Eurostat is a very relevant case study.
- The integration of a various geographic objects like cities and grids could have a very potential contribution to the time series building process. It could also affine and improve the thematic analysis
- The automation of the dictionary of changes and the estimation methods



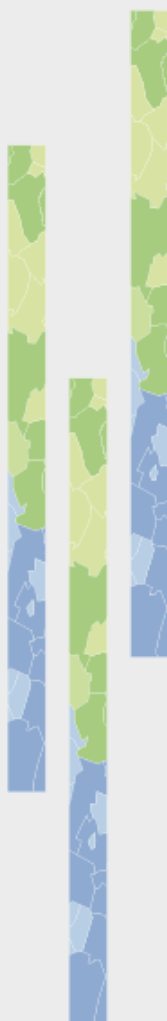
**Figure 28:** Synthesis of the work of the "time-series" group.

## Bibliography

- [Andrienko 01]** Natalia Andrienko, Gennady Andrienko & Peter Gatalisky. Exploring change in census time series with interactive dynamic maps and graphics. *Computational Statistics*, vol. 16, no. 3, pages 417–433, 2001.
- [Andrienko 03]** Natalia Andrienko, Gennady Andrienko & Peter Gatalisky. Exploratory spatio-temporal visualization: an analytical review. *Journal of Visual Languages & Computing*, vol. 14, no. 6, pages 503 – 541, 2003. *Visual Data Mining*.
- [Andrienko 06]** Natalia Andrienko & Gennady Andrienko. *Exploratory analysis of spatial and temporal data*. Springer, 2006.
- [Ben Rebah 08]** Maher Ben Rebah. *La cartographie dynamique comme outil d'investigation territoriale : le cas du découpage administratif en Tunisie*. PhD thesis, Université de Paris VII, 2008.
- [Cheylan 93]** Jean-Paul Cheylan & Sylvie Lardon. Towards a Conceptual Data Model for the Analysis of Spatio-Temporal Processes: The Example of the Search for Optimal Grazing Strategies. In *COSIT*, pages 158–176, 1993.
- [Claramunt 95]** Christophe Claramunt & Marius Thériault. Managing Time in GIS: An Event-Oriented Approach. In *Proceedings of the International Workshop on Temporal Databases*, pages 23–42, London, UK, 1995. Springer-Verlag.
- [Galton 04]** Antony Galton. Fields and Objects in Space, Time, and Space-time. *Spatial cognition and computation*, vol. 4, no. 1, pages 39–68, 2004.
- [Gregory 02]** Ian Gregory. Time-variant GIS Databases of Changing Historical Administrative Boundaries: A European Comparison. *Transactions in GIS*, vol. 6, no. 2, pages 161–178, 2002.
- [Langran 92]** Gail E. Langran. *Time in geographic information systems*. Taylor and Francis, Seattle, WA, USA, 1992.
- [Langran 98]** Gail E. Langran & Nicholas R. Chrisman. A Framework for Temporal Geographic Information. *Cartographica: The International Journal for Geographic Information and Geovisualization*, vol. 25, no. 3, pages 11–14, 1998.
- [Peuquet 94]** Donna Peuquet. It's About Time: A Conceptual Framework for the Representation of Temporal Dynamics in Geographic Information Systems. *Annals of the Association of American Geographers*, vol. 83, no. 3, pages 441–461, 1994.
- [Peuquet 95]** Donna Peuquet & Niu Duan. An event-based spatiotemporal data model (ESTDM) for temporal analysis of geographical data. *International Journal of Geographical Information Science*, vol. 9, no. 1, pages 7–24, 1995.
- [Peuquet 02]** Donna Peuquet. *Representations of time and space*. Guildford Press, New York, NY, USA, 2002.
- [Plumejeaud 09]** Christine Plumejeaud, Jerome Gensel, Marlène Villanova-Oliver, Maher Ben Rebah & Guillaume Verognaud. Modélisation de hiérarchies territoriales multiples - Vers la gestion d'informations spatio-temporelles évolutives. In *Colloque International de Géomatique et d'Analyse Spatiale (SAGEO 2009)*, Paris, France, 2009.



## Disaggregation of socioeconomic data into a regular grid and combination with other types of data



### CONTENT

- **Introduction and methodology description.** This section describes the background details and methodological solution adopted with regard to the disaggregation of data.
- **Integration of socio-economic and environmental information.** This part presents the OLAP Cube solution implemented in order to combine data from several data sources and formats, mainly socioeconomic statistics together with continuous data, such as land cover.
- **Future steps.** Taking into account the results from the challenge along the ESPON DB projects this section drafts the following steps that could be developed as a continuation.



**ESPON 2013 DATABASE**

MARCH 2011



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

**37 PAGES (incl. annexes)**

# LIST OF AUTHORS

Roger Milego, Universitat Autònoma de Barcelona, UAB 08193 Bellaterra

Maria José Ramos, Universitat Autònoma de Barcelona , UAB 08193 Bellaterra

## **Contact**

roger.milego@uab.cat

mariajose.ramos@uab.cat

tel. + 34 93 581 35 46



# TABLE OF CONTENT

1	Introduction.....	3
2	Methodology description .....	7
2.1	Integration methods.....	8
3	Integration of socio-economic and environmental information .....	10
4	Future steps .....	12
5	Conclusions .....	14
6	References .....	16
ANNEX 1.	Testing the methodology .....	18
Testing the maximum area criterion: Urban dominance (2000) based on the Urban Morphological Zones.....		18
Testing the “Proportional calculation” method: Unemployment rate total (2001) .....		23
Testing the “Proportional and weighted calculation” method: GDP – Wealth and Production (2002) .....		26
ANNEX 2.	ESPON OLAP Cube User Manual.....	33

# 1 Introduction

The ESPON 2013 DB project has been structured in several challenges in order to fulfil its objectives. The challenge to which this technical report refers to is the challenge number 5: "Combining socio-economic data measured for administrative zoning (NUTS level) and environmental data defined on a regular grid (like Corine Land cover)". The UAB (Universitat Autònoma de Barcelona) is the responsible partner with regard to this challenge.

Most of the socioeconomic variables or indicators are typically given by administrative unit, i.e. NUTS regions, whereas the environmental data is usually not following those boundaries, but given by natural units or regular grid cells.

The aim of this challenge is to define a suitable methodology for integrating and making comparable data coming from statistical sources (e.g. EUROSTAT) and measured by administrative unit, together with environmental data stored by natural unit or regular grid structure (e.g. Corine Land Cover).

The ESPON 2006 program developed some indicators in which the environmental data was transposed to NUTS division by means of GIS tools, in order to make them comparable to socioeconomic data. The results from this integration strategy, not always convincing, make clear the necessity of implementing a new integration process based on grid methods as it is said in the tender of the Espo 2013 Database project and in the Modifiable Areal Unit Problem study.

## ESPON 2013 Database challenges

According to the E.S.T.I (space, Source, Time, Indicator) framework presented in the "Handbook for data collection" (ESPON 3.2, Final Report, Annex) four main objectives were identified in the tender of the ESPON 2013 DB Programme.

This challenge is included in the second key-question: "**Combination of heterogeneous sources- balancing Eurostat data**", that emphasizes the need of the integration between different types and sources of data. The harmonization of the database in a fixed spatial division (NUTS3) solution that was chosen by many ESPON 2006 projects presented some doubts and not always convincing results. It is in this scenario that **a new integration methodology based on the reverse operation** is mentioned, **a projection of socio-economic information into units elaborated for the monitoring of natural resources.**

## Modifiable Areal Unit Problem (ESPON 3.4.3)

The MAUP study, in its chapter 4 "*Exploration of gridding methods*", highlights the integration of heterogeneous databases as one of the most promising application of gridding methods for ESPON.

Two potential fields of applications are distinguished for gridding methods:

- **Time harmonisation of changing territorial units.**

“The use of grid help to build an harmonised territorial framework where all changing territorial divisions are harmonised and can further be used for the analysis of time variation” MAUP study (ESPON 3.4.3)

- **Thematic harmonisation and combination of heterogeneous spatial sources.**

The ESPON 2006 integration strategy, called “Eurostat oriented” by MAUP study, based on transferring all the information that it is not delivered on the basis of administrative units (NUTS 2 or NUTS 3) toward administrative units, is questioned, and the use of a new strategy is proposed.

“Information of good quality (as CLC) is therefore transformed into information of bad quality when projected in spatial units which are not adapted” MAUP study (ESPON 3.4.3).

“Eurostat oriented” strategy could be replaced by another strategy that could be called the “EEA oriented” where all data would be transformed into grid and integrated on this basis” MAUP study (ESPON 3.4.3).

## Methodology Proposal

In the First Interim Report of the project (Feb 2009), a methodology proposal was made on the basis of existing applications made by other institutions, such as:

- **“A Downscaled Population Density Map of the EU from Commune Data and Land Cover Information”** by Javier Gallego, JRC-ISPRA.

A combination of commune population data with Corine Land Cover to produce an EU-wide grid with 1 ha resolution of downscaled population density<sup>1</sup>.

- **G-Econ Research project of the University of Yale to develop a geophysically based data set on economic activity.**

Estimation of gross output at a 1-degree longitude by 1-degree latitude resolution at a global scale for virtually all terrestrial grid cells based on spatial rescaling settled on proportional allocation<sup>2</sup>.

- FARO-EU (Foresight Analysis of Rural areas Of Europe)

The project is aimed to analyse Rural Development in Europe by analysing patterns and trends of a selection of territorial indicators specific for rural areas within a Spatial Regional Reference Framework<sup>3</sup>.

---

1

[http://epp.eurostat.ec.europa.eu/portal/page/portal/research\\_methodology/documents/S14P3\\_JAVIER\\_GALLEGO\\_DO\\_WNSCALED\\_POPULATION\\_DENSITY.pdf](http://epp.eurostat.ec.europa.eu/portal/page/portal/research_methodology/documents/S14P3_JAVIER_GALLEGO_DO_WNSCALED_POPULATION_DENSITY.pdf)

<sup>2</sup> “New Metrics for Environmental Economics: Gridded Economic Dats” by William D. Nordhaus.

<http://www.oecd.org/dataoecd/44/7/37117455.pdf>

<sup>3</sup> [www.faro-eu.org](http://www.faro-eu.org)

- **"Transforming Population Data for Interdisciplinary Usages: From census to grid"** by Deborah Balk & Greg Yetman from Columbia University.

Creation of the Gridded Population of the World (GPW) data base implementing **a proportional allocation** of population from administrative units to grid cells

<sup>4</sup>

The objectives established in the tender of the ESPON 2013 DB, the MAUP study results and recommendations, the bibliography research on existing methodologies and our experience at the UAB, as European Topic Centre on Land Use and Spatial Information, supporting the EEA in monitoring the land use/land cover change in Europe and analyzing the environmental consequences; lead us to the conclusion that the best way to downscale socioeconomic data and make them comparable with other kind of data, is **using a regular grid structure**, in which each cell takes a figure of the indicator or variable.

It was also concluded that depending on the nature of each variable, a different integration method should be applied. In other words, the way of calculating the actual figure for each grid cell might differ between different types of data, according to their definition.

## The European Reference Grid

The EEA recommends the use of EEA reference grids for projection ETRS89-LAEA 52N 10E. The recommendation is based on proposal at the 1<sup>st</sup> European Workshop on Reference Grids<sup>5</sup>.

The 1<sup>st</sup> Workshop on European Reference Grids was organized by the Joint Research Centre of the European Commission following a request of the EEA and the request of the INSPIRE Implementing Strategies Working Group that recommended the adaptation of a Europe-wide reference grid to facilitate the management and analyses of spatial information. The interest of the creation of a common coordinate reference system and a common equal-area grid to represent EU and Pan-Europe was also expressed by the National Statistical Institutes.

Taking into account this recommendations and our UAB/ETC-LUSI experience under some EEA projects such as LEAC (Land and Ecosystem Accounting), we proposed to disaggregate socioeconomic data into the **1 km European Reference Grid**<sup>6</sup>, as it is the way in which valuable data for the ESPON projects, such as the Corine Land Cover changes, are stored as well.

## Testing process

After making all these decisions, we have carried out several tests with different data using different integration methods, trying to achieve useful results, on the one hand,

<sup>4</sup> <http://sedac.ciesin.columbia.edu/gpw-v2/GPWdocumentation.pdf>

<sup>5</sup> [http://eusoiils.jrc.ec.europa.eu/projects/alpsis/Docs/ref\\_grid\\_sh\\_proc\\_draft.pdf](http://eusoiils.jrc.ec.europa.eu/projects/alpsis/Docs/ref_grid_sh_proc_draft.pdf)

<sup>6</sup> <http://dataservice.eea.europa.eu/dataservice/metadetails.asp?id=760>

but aiming at preparing the basis for the automation of the processes and integration with environmental data in an OLAP (On-Line Analytical Processing) cube.

Whenever an ESPON project would like to make the comparison of any kind of socioeconomic data together with land cover or environmental data not measured by administrative unit, the ESPON 2013 DB project, through the challenge 5 outcomes, will be ready to facilitate this task and provide them with the expected results or methodological tools to be applied.

The goal of the results presented in this report is to highlight the high potential of the three methodologies proposed. At this point the layers distributed by 1km grid resulting from applying the different methods are not available; their future distribution will be based on the building of OLAP's cubes using the most updated data.

## **Objectives**

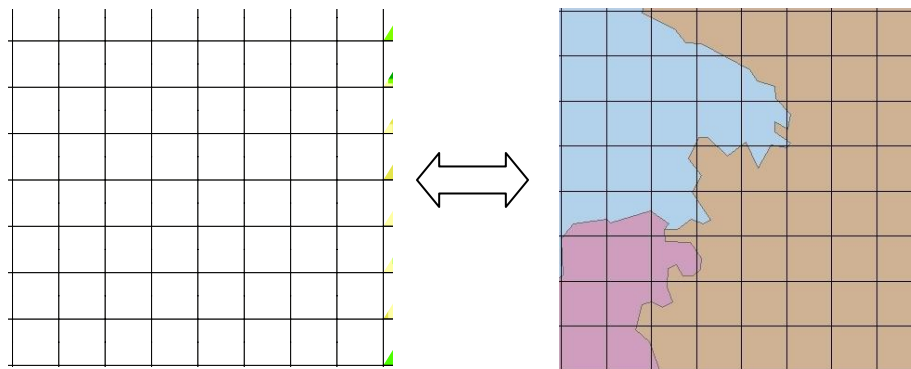
This technical report is aimed at explaining the details of the methodology and tests undertaken regarding the challenge 5 of the ESPON 2013 DB and it has, in particular, the following objectives:

- Review and summarise the background of the challenge 5.
- Give a detailed description of the methodology to be applied in order to downscale socioeconomic data.
- Describe the different disaggregation methods.
- Explain the tests undertaken and the results that have been achieved so far.
- Make some conclusions about all the processes undertaken so far.
- Define the next steps to be carried out.

## 2 Methodology description

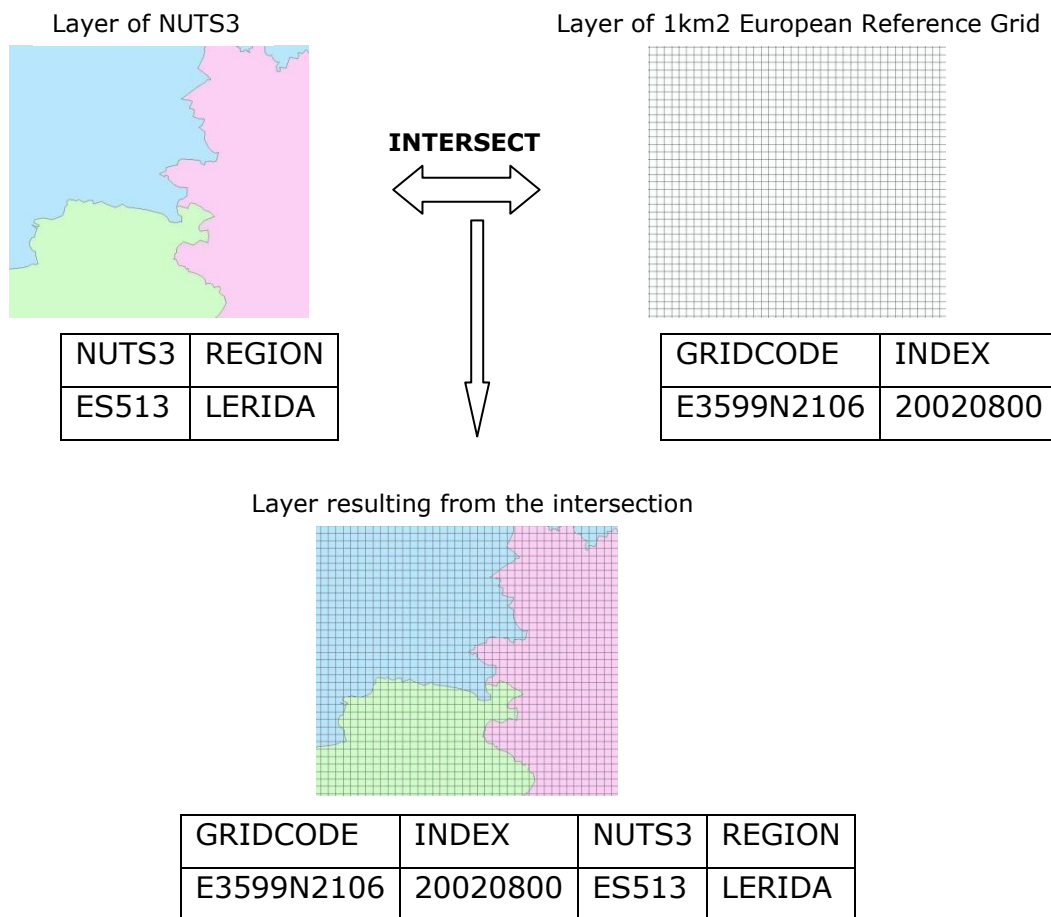
This section describes the details of the methodology proposed to face the challenge of combining data measured using different reporting units.

According to what it has been explained in the introduction, and after reviewing several studies and taking into account our experience at the UAB (ETC-LUSI) and the EEA, we propose to integrate socioeconomic data in the 1 km European Reference Grid, because, besides some other reasons, this unit is used to summarise land cover data and other types of environmental data processed at the EEA.



**Figure 1.** The 1 km European Reference Grid will hold both environmental and socioeconomic information.

Therefore, the first step to be carried out should be the **intersection** between the 1 km European Reference Grid and the administrative units by which the indicator is given. This is done by a physical overlay of both layers in vector format, by means of the ArcGIS tool Intersect. This tool creates a new layer holding both the geometries and the attributes of the source layers.

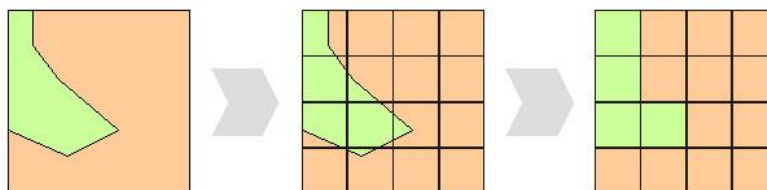


**Figure 2.** Example of the intersection tool.

Once the intersection has been computed, a way of computing a single figure by each reference grid cell should be defined. It has been stated that depending on the nature of each indicator or variable, a different kind of integration procedure should be defined. In this regard, we have defined and tested with different data the following three integration methods:

## 2.1 Integration methods

**Maximum area criteria:** the cell takes the value of the unit which covers most of the cell area. It should be a good option for uncountable variables.



**Proportional calculation:** the cell takes a calculated value depending on the values of the units falling inside and their share within the cell. This method seems very appropriate for countable variables.



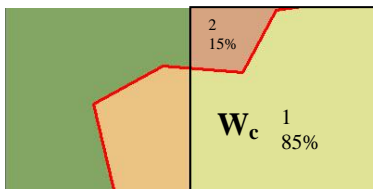
$$\text{Cell value} = \sum ( V_i * \text{Share}_i )$$

Where:  $V_i$  = Value of unit i

$\text{Share}_i$  = Share of unit i within the cell

In the example:  $V_1 * 0.85 + V_2 * 0.15$

**Proportional and weighted calculation:** the cell takes also a proportionally calculated value, but this value is weighted for each cell, according to an external variable (e.g. population). This method can be applied to improve the territorial distribution of a socioeconomic indicator. For instance, a GDP indicator can be redistributed by 1 km grid and weighted by the population figures of each cell (coming from the 1 km population density dataset produced by JRC).



$$\text{Cell value} = W_c \sum ( V_i * \text{Share}_i )$$

Where:  $V_i$  = Value of unit i

$\text{Share}_i$  = Share of unit i within the cell

$W_c$  = weight assigned to cell c

In the example:  $W_c * (V_1 * 0.85 + V_2 * 0.15)$

Depending on each type of indicator or variable to be integrated within the reference grid, a different type of integration should be decided and tested. Besides the method finally chosen to integrate, it is important to highlight that indicator figures given by area unit, e.g. by square kilometre, should be converted considering that each cell has a total area of 1 km<sup>2</sup>.

Whenever it is possible, the third method has been used and should be used to disaggregate the data, as it gives an added value to the source data, providing more interesting results when different data are put together on a cell-by-cell basis.

**Annex 1** shows the test results for the three integration methods.

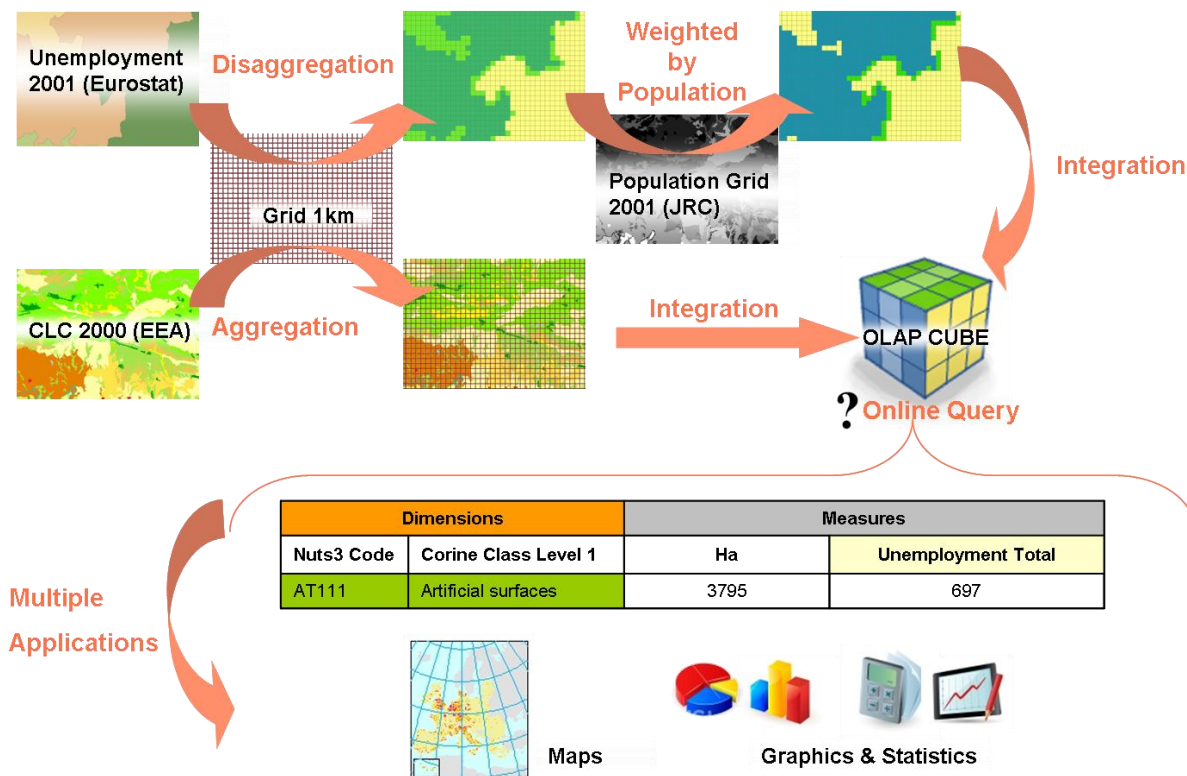


### 3 Integration of socio-economic and environmental information

Once the variable has been distributed by 1 km cell, it can be compared to other data on a cell-by-cell basis, and it can be integrated into an **OLAP (Online Analytical Processing) cube**.

The OLAP technology<sup>7</sup> use a multidimensional data model, allowing complex analytical and ad-hoc queries with a rapid execution time. This solution facilitates the integrated analysis of several types of geographic and statistical data to users with or without GIS knowledge.

The next schema presents the general process to make possible the integration within an OLAP cube



**Figure 13.** Simplified schema of data integration in an OLAP Cube

In the case of ESPON OLAP cube, it consists on the **ESPON socio-economic** variables as numerical attributes or measures that will be aggregated using a set of dimensions. The **dimensions** or topics of interest for the user are generally represented by different types: **spatial dimensions**, usually represented by

<sup>7</sup>Some OLAP information resources:

- 1- [http://en.wikipedia.org/wiki/Online\\_analytical\\_processing](http://en.wikipedia.org/wiki/Online_analytical_processing)
- 2- <http://www.cs.sfu.ca/CC/459/han/papers/chaudhuri97.pdf>
- 3- [http://es.wikipedia.org/wiki/Cubo\\_OLAP](http://es.wikipedia.org/wiki/Cubo_OLAP)

administrative units for Europe (NUTS), a number of **thematic dimensions**: land use data or dominant land cover types; and a third type, which is the **temporal dimension** that shows the difference between two years time.

In this way, we are able to ask to the ESPON-OLAP cube some questions taking into account the socioeconomic variables or indicators. For example, having the GDP and CLC changes in the ESPON OLAP cube, we could analyse which land cover flows occur by different GDP ranges, and, in the end, get the results on a NUTS3, NUTS2 or country (NUTS0) basis.

The OLAP Cube delivered together with the ESPON 2013 DB Final Report is the version 3.0, and holds the following dimensions and measures:

- Socioeconomic variables
  - GDP 2003
  - GDP 2006
  - Active population 2003
  - Active population 2006
  - Unemployment 2003
  - Unemployment 2006
- Land cover
  - Corine Land Cover 1990
  - Corine Land Cover 2000
  - Corine Land Cover 2006
- Land cover changes
  - Land Cover Flows 1990-2000
  - Land Cover Flows 1990-2006
  - Land Cover Flows 2000-2006
- Measures
  - Population density 2001 (inhab/km<sup>2</sup>)
  - Area (ha)
- Geographical dimensions
  - Elevation Breakdown
  - Biogeographic Regions
  - Large Urban Zones and City Names
  - Massifs
  - NUTS 2006
  - NUTS 2003
  - River Basin Districts UE

The ESPON OLAP Cube can be accessed either online or offline. Online connection has not been implemented yet in the framework of the ESPON DB. So far, a CUB file has been provided in order to use the ESPON OLAP Cube offline. **Annex 2** explains how to use this file and make queries to the Cube by means of MS-Excel.

## 4 Future steps

There are several improvements and future implementations we have in mind to make this methodological approach and derived tools grow up in the future. We list below some of them:

- a) Methodological improvements: Some aspects that have to be deeply analysed are:
  - Treatment of administrative units with no data values.
  - Differences between geographical extents, for example between Nuts 3 2006 layer and Corine Land Cover.
  - Improvement of the disaggregation performance in terms of time and manageability of the final layer.
  - Calculation of deviations with respect to the source data.
- b) Extent increase: It is foreseen to increase the number of neighbours, countries and areas to process. The possible candidates will comprise Maghreb and part of Saharan Africa together with a portion of the CIS<sup>8</sup>.
- c) Follow-up of the EFGS: It might be very useful to keep following up the outcomes of the European Forum for Geostatistics<sup>9</sup> in order to contrast the proposed method, and, eventually, improve it.
- d) Questionnaire: As the ESPON OLAP Cube is a standalone database we would like to explain to the ESPON projects the practical possibilities of the tool and ask them feedback about the data they would like to have combined with other data sources. For example, many projects might like to have some of their datasets combined with Corine Land Cover. It is important for us to identify such datasets in order to think how aggregate/disaggregate them into the grid and, eventually, integrate them within the ESPON OLAP Cube.
- e) Integration with the ESPON 2013 Database: It should be deeply studied the most suitable solution to integrate grid data in the ESPON 2013 Database and to make the ESPON OLAP Cube available to users through the database interface. The main idea is to facilitate the comparison and analysis of socio-economic data with environmental data that usually not follows an administrative distribution. At this stage, the ESPON OLAP Cube is provided as a standalone file which is accessed from the Database by means of a link. In order to use it, a connection from MS Excel should be established as it is explained in the ESPON OLAP Cube User Manual (Annex 2 of this Technical Report).
- f) Visualisation tool: At this stage, the ESPON OLAP Cube queries are made through MS Excel Pivot Tables application, which allows to get the results in table format. A suitable web visualisation tool could be adapted and integrated into the ESPON database framework allowing the visualisation of grid and socio-

---

<sup>8</sup> Commonwealth of Independent States

<sup>9</sup> <http://www.efgs.ssb.no/>

economic ESPON OLAP data and, eventually, the other ESPON database data. Two specific environments could be defined depending on the specific needs of the user: one more simple, designed for a basic user just for visualisation purposes, and a more advanced one, being able to show comprehensive results like interactive maps, graphs or tables.

- g) A set of ESPON OLAP Cubes: The ESPON OLAP Cube, as a standalone database, is built by means of the integration of specific datasets within the 1 km<sup>2</sup> European Reference Grid, which should be previously selected according to the analytical purposes. Therefore, if several purposes are identified, there should be several thematic Cubes to fit them. Moreover, some ESPON OLAP Cubes could also be built on the basis of a better resolution grid only for some specific region, such as urban areas to better analyse urban processes.
- h) in order to integrate the gridded data and make them ready to be queried and analysed together with other types of data, like land cover, natural units, protected areas, and so on, for different scales (i.e. Urban Cube).

## 5 Conclusions

After having defined a methodology in order to be able to put together and analyse socioeconomic and environmental data, and having made several tests using different datasets, we can make now some conclusions about the main outcomes of the work done and the things we have learned so far:

- Disaggregating socioeconomic data by a regular grid is the best solution in order to downscale such information reported by administrative areas.
- The 1 km European Reference Grid is a good option to undertake the disaggregation because:
  - It has an European coverage
  - It follows Inspire specifications
  - It is used for several institutions as the reference grid
  - Its resolution is optimal in order not to lose data precision
- For uncountable data (non-numeric values), the best aggregation method is the “maximum area criterion”.
- For countable data, the best method is the proportional one, which calculates the final value according to the area share of each of the values.
- Whenever it is possible, it is better to weight the final figures when using a proportional method, e.g. by population.
- The “proportional and weighted” aggregation method is the one that gives better results, plus some added value to the downscaling.
- The different methods are independent from the source data format and can be applied to vector and raster format.
- In order to achieve good results following this methodology it is important to use data sources which follow the same spatial and temporal specifications (extent, spatial resolution, temporal resolution...).
- This methodology allows the integration of socio-economic in an OLAP cube, which facilitates the comparison and analysis of such data together with land cover data, for example.
- The benefits of having the ESPON OLAP Cube built up has been proven, although the pre-processing and data selection steps take quite a long time and resources. Furthermore, once the Cube is ready, it is not possible for the users to add data dynamically.

To sum up, it can be added that any kind of socioeconomic data can be processed using the methodology proposed and tested, in order to have them downscaled and stored by 1 km grid, facilitating their comparison with many other data not reported by administrative units.

Our next steps will be aimed at improving the performance of the methodologies proposed but also to analyse the introduction of some changes based on our results and new projects at European context.



## 6 References

### • Literature

Arévalo J., *Land and Ecosystem Accounting. Technical Procedure, Internal Report v.2*, 2009, ETC-LUSI, European Environmental Agency.

Chaudhuri S., Dayal U., *An overview of Data Warehousing and OLAP Technology*, Simon Fraser University Canada (SFU.CA).

Deichmann U., Balk D., Yetman G., 2001, *Transforming Population Data for Interdisciplinary Usages: From census to grid*, NASA Socioeconomic Data and Applications Center (SEDAC).

Gallego J., *A Downscaled Population Density Map of the EU from Commune Data and Land Cover Information*, JRC-Ispra.

Gallego J., *Downscaling population density in the European Union with a land cover map and a point survey*, JRC-Ispra.

Gallego J., *Population density grid of EU-27+, version 4. Summary of the downscaling method*, JRC-Ispra.

Malinowski E., Zimányi E., 2009, *Advanced Data Warehouse Design- From Conventional to Spatial and Temporal Applications*, Springer.

*Short Proceedings of the 1<sup>st</sup> European Workshop on Reference Grids, Ispra, 27-19 October 2003*, JRC- Institute for Environmental and Sustainability, Ispra

William D. Nordhaus, 2006, *New Metrics for Environmental Economics: Gridded Economic Data*, Yale University

### • Websites

**Espon programme:** <http://www.espon.eu/>

#### - **Future Orientation for Cities (FOCI). ESPON programme**

[http://www.espon.eu/mmp/online/website/content/programme/1455/2233/2236/2239/index\\_EN.html](http://www.espon.eu/mmp/online/website/content/programme/1455/2233/2236/2239/index_EN.html)

#### - **The modifiable areas unit problem (MAUP). ESPON Scientific Support Project 3.4.3**

[http://www.espon.eu/mmp/online/website/content/projects/261/431/index\\_EN.html](http://www.espon.eu/mmp/online/website/content/projects/261/431/index_EN.html)

**European Environment Agency (EEA). European Commission:**  
<http://www.eea.europa.eu/>

#### - **Population density disaggregated with Corine land cover 2000.**

<http://www.eea.europa.eu/data-and-maps/data/population-density-disaggregated-with-corine-land-cover-2000-1>

- **EEA reference grids**

<http://www.eea.europa.eu/data-and-maps/data/eea-reference-grids>

**European forum for Geostatistics 2010.** <http://www.efgs.ssb.no/>

**Eurostat, European Commission:**

<http://epp.eurostat.ec.europa.eu/portal/page/portal/eurostat/home/>

**Foresight Analysis of Rural areas of Europe (Faro-eu.org)** <http://www.faro-eu.org>

**Geographically based Economic data (G-Econ) project, University of Yale**

<http://gecon.yale.edu/>

**Joint Research Centre (JRC). European Commission:**

<http://ec.europa.eu/dgs/jrc/index.cfm>



## **ANNEX 1. Testing the methodology**

In order to test the methodology proposed and the different methods of data integration into the reference grid, we have chosen some socioeconomic variables or other data not being measured by grid cell, but stored using administrative units or other kind of irregular delimitation. Thus, we have chosen variables such as GDP, unemployment or urban dominance. The next sections show the steps undertaken and results achieved for any kind of test done.

### **Testing the maximum area criterion: Urban dominance (2000) based on the Urban Morphological Zones.**

The maximum area criterion is suitable for uncountable data or data given by ranges. In particular, this approach can be applied to better differentiate the urban/non urban character of the land as has been highlighted by the team of **ESPON FOCI**<sup>10</sup> project. The advantage of taking percentage of Urban Morphological Zones over pure land cover classes is that UMZ indicates if certain area of urban fabric is part of a greater entity (equivalent to a city). Then, the urban dominance could be used to characterise certain areas and to relate them to other descriptors.

#### a) Source data:

In order to make this test we have used the Urban Morphological Zones 2000 (Map1. Urban Morphological Zones 2000), an EEA's dataset which is a delimitation of urban areas according to a functional definition. The UMZ2000 come from a reclassification of different land cover classes of CLC2000 following different criteria and, therefore, they are not following administrative boundaries but an artificial boundary created by addition of land cover polygons.

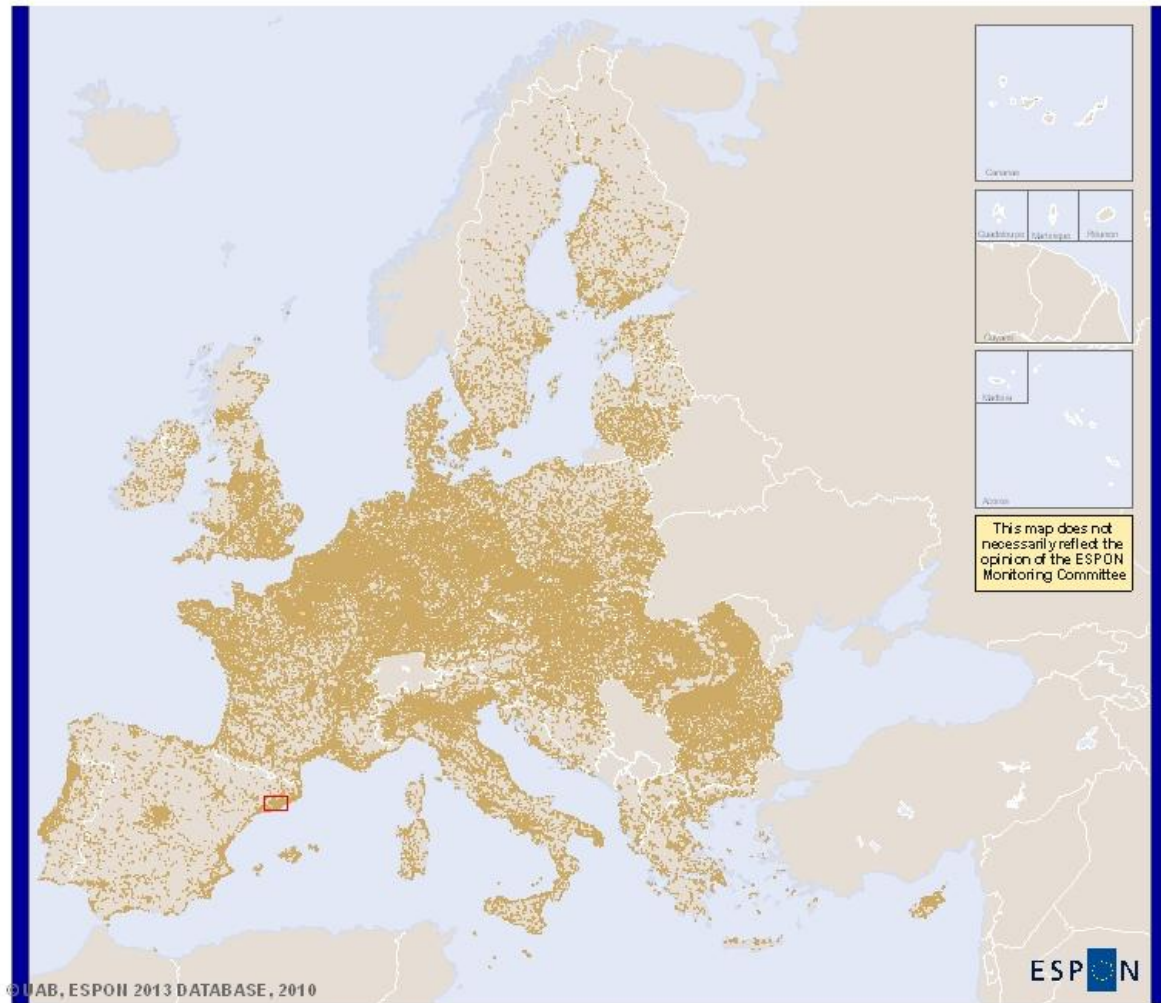
For further details about UMZ, please visit:

<http://dataservice.eea.europa.eu/download.asp?id=17335&filetype=.pdf>

---

<sup>10</sup> [http://www.espon.eu/mmp/online/website/content/programme/1455/2233/2236/2239/index\\_EN.html](http://www.espon.eu/mmp/online/website/content/programme/1455/2233/2236/2239/index_EN.html)

## Map 1 . Urban Morphological Zones 2000



© UAB, ESPON 2013 DATABASE, 2010

ESPON 2010  
 European Union Regional Development Fund  
 "PROGRESS FOR PEOPLE"

Regional level: UMZ

Source: Corine Land Cover, 2000

Origin of data: Corine Land Cover, 2000

© EuroGeographics Association for administrative boundaries

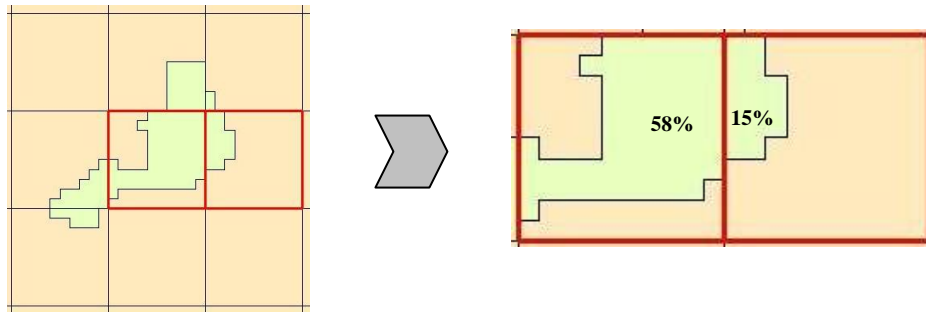
### Urban Morphological Zones 2000

Urban Morphological Zones 2000



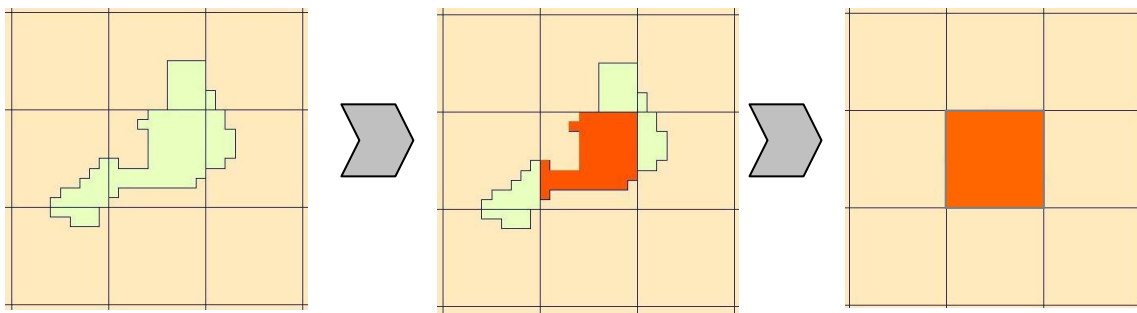
### b) Process steps:

As it has been stated in the methodology section, the first step has been the intersection between the Reference Grid and the UMZ2000. In this way, we are able to calculate which share of grid cell is occupied by an UMZ.



**Figure 3.** Intersection of UMZ2000 and the Reference Grid.

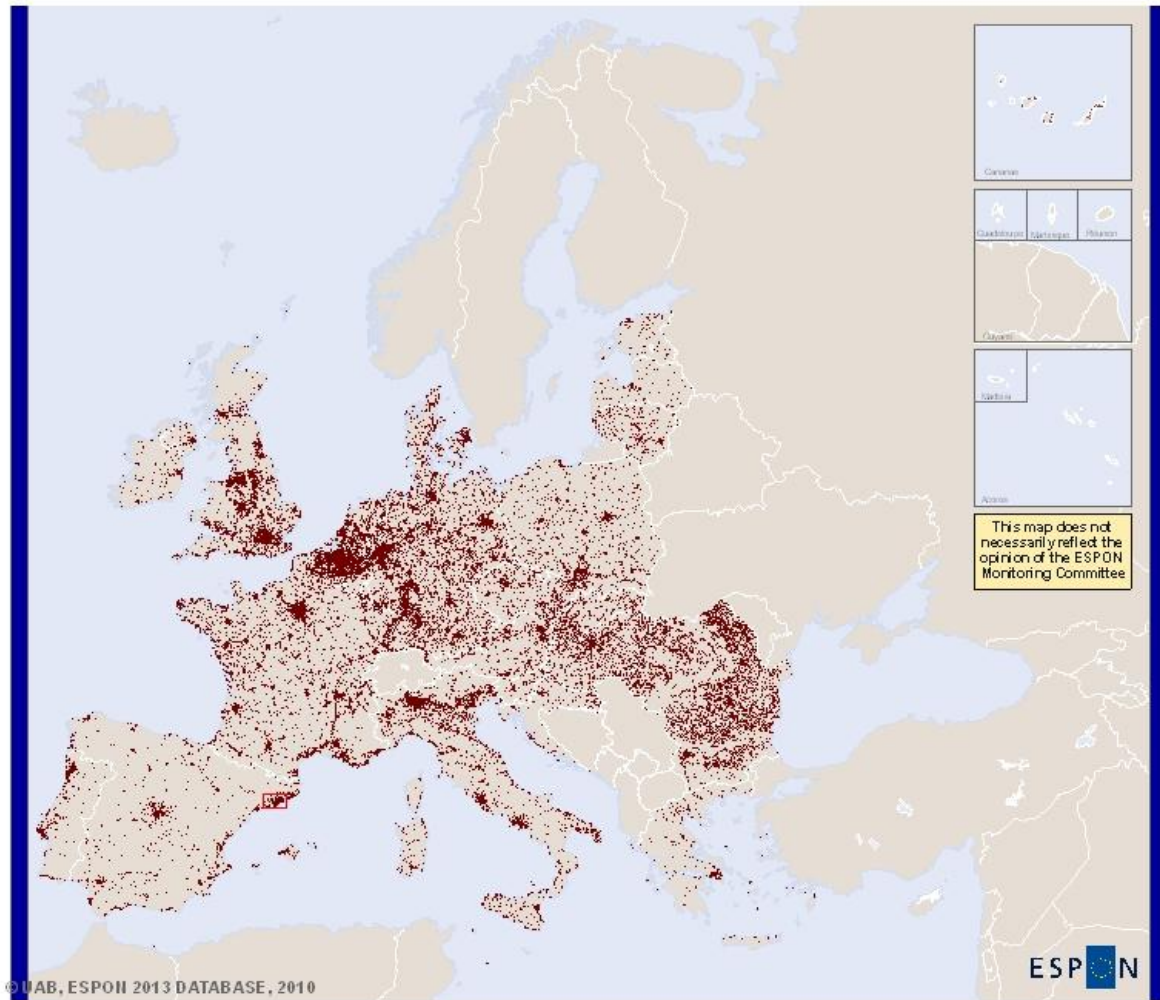
If the cell has more than its half (50%) covered by an UMZ polygon, we define it as an urban-dominant cell, whereas if it has less than 50% of UMZ inside, it is defined as a non-urban cell.



**Figure 4.** Urban dominance definition process.

Finally, this kind of map can be elaborated, where we can see the urban dominance in Europe by 1 km grid cell being able to identify quicker the main points of urban surfaces in Europe:

## Map 2 . Urban Dominance layout



© UAB, ESPON 2013 DATABASE, 2010

ESPON 2013  
 Performance of the European Regional Development Fund  
 "STRATEGIA REGIONALNA 2013-2020"

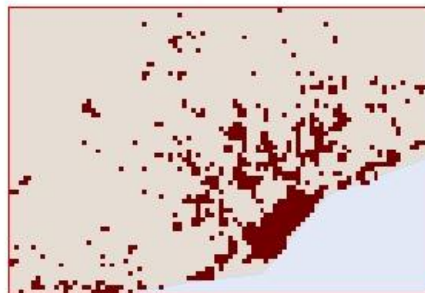
Regional level: Grid  
 Source: UMZ, 2000

Origin of data: Corine Land Cover, 2000

© EuroGeographics Association for administrative boundaries

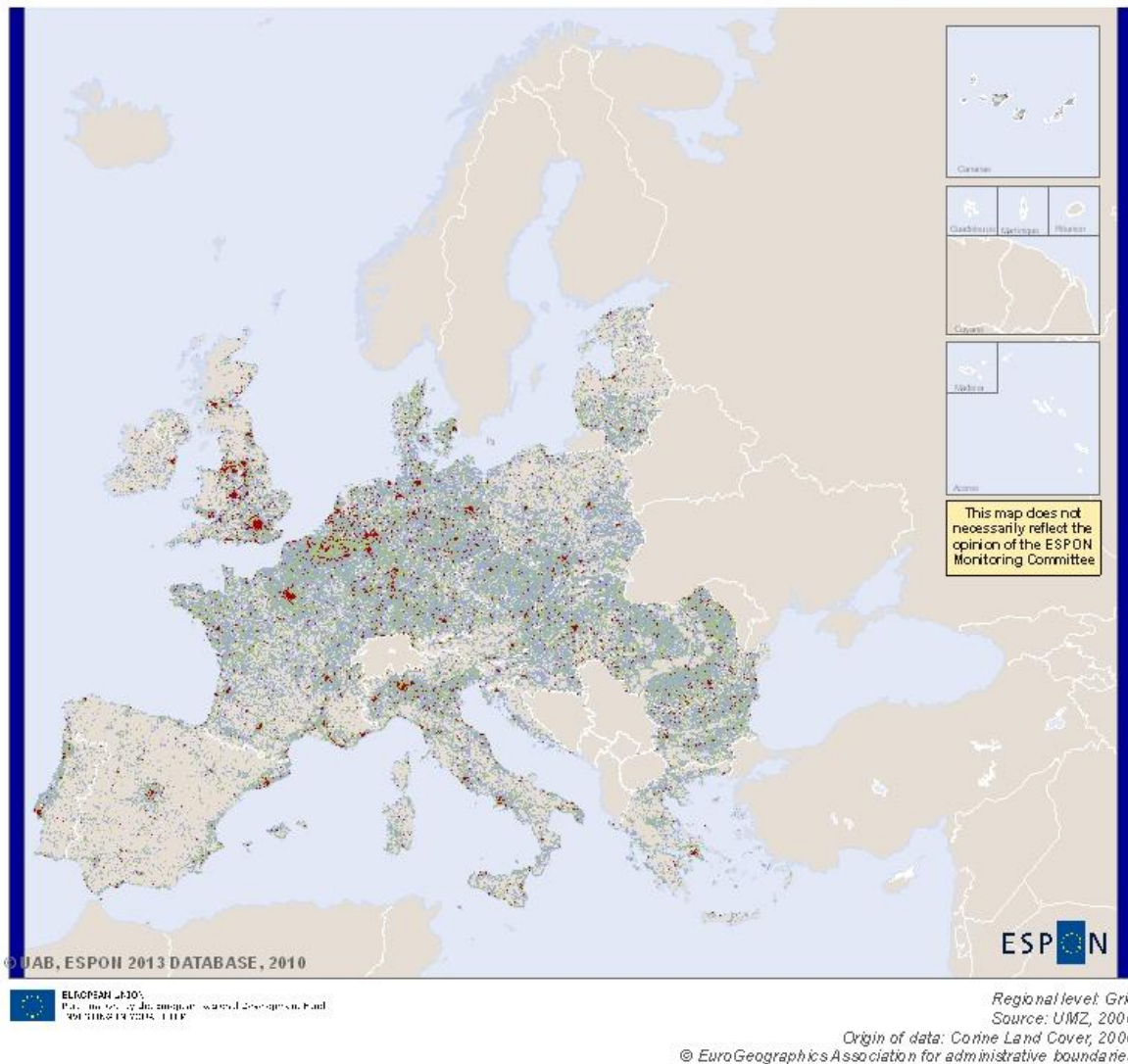
### Urban Morphological Zones 2000

■ Urban Dominance

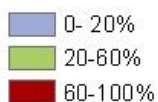


What is interesting is that now we are able to analyse land cover changes and flows by urban or non-urban cells, or any other indicator. Furthermore, we can create an urban classification by percentage ranges, e.g. 0-20%, 20-60%, 60-100% (low urban dominance, mid urban dominance, high urban dominance):

### Map 3 . Urban Classification by density ranges



#### Urban Classification



Conclusion: the maximum area criteria is useful for non-numeric values or in case we are interested in a discrete classification of grid cells, either by a thematic attribute or a value range. Moreover, the **ESPON FOCI** project found interesting this kind of approach applied to UIMZ because it can be used as a criteria to define urban dominance and integrate it in the urban-hinterland analysis.

# Testing the "Proportional calculation" method: Unemployment rate total (2001)

## a) Source data:

As for the proportional calculation method, we have chosen an indicator of unemployment: the unemployment rate total (2001) from Eurostat. It represents unemployed people as a percentage of the economically active population, and it is measured by NUTS3 region.

Although the most suitable methodology to be applied on this indicator is the third one, proportional and weighted calculation, we have selected the unemployment rate total as an example of the possible results that can be obtained by applying the second methodology.

	A	B	C	D	E	F	G	H	I	J
1	<b>Employment and Labour Market</b>									
2	<b>NUTS level 3 (version 1999)</b>									
3	<b>Subtheme:</b>		Unemployment	Unemployment	Unemployment	Unemployment	Unemployment	Unemployment	Unemployment	Unemployment
4	<b>Indicator:</b>		Unemployment rate total	Unemployment rate female	Unemployment rate male	Unemployment rate under 25 years	Development of unemployment rate 1998-2001	Development of unemployment rate, female, 1998-2001	Development of unemployment rate, male, 1998-2001	Development of unemployment rate, <25 years, 1998-2001
5	<b>Description:</b>		in %	in %	in %	in %	in percentage points	in percentage points	in percentage points	in percentage points
6	<b>Time:</b>		2001	2001	2001	2001	1998-2001	1998-2001	1998-2001	1998-2001
7	<b>Source:</b>		Eurostat, Norway and Switzerland: National Statistical Offices	Eurostat, Norway and Switzerland: National Statistical Offices	Eurostat, Norway and Switzerland: National Statistical Offices	Eurostat, Norway and Switzerland: National Statistical Offices	Eurostat, Norway and Switzerland: National Statistical Offices	Eurostat, Norway and Switzerland: National Statistical Offices	Eurostat, Norway and Switzerland: National Statistical Offices	Eurostat, Norway and Switzerland: National Statistical Offices
8	<b>Comment:</b>									
9	NUTS_3_99	Region	UNRT01N3	UNRF01N3	UNRM01N3	UNRU2501N3	UNRT98N3	UNRF98N3	UNRM98N3	UNRU2598N3
10	AT111	Mittelburgenland	3,1	4,8	2,2	3,4	-1,0	-1,9	-0,5	-0,1
11	AT112	Nordburgenland	2,3	2,8	2,0	4,0	-0,8	-1,2	-0,5	-0,6
12	AT113	Südburgenland	4,3	5,2	3,5	7,6	-0,9	-1,4	-0,5	0,0
13	AT121	Mostviertel-Eisenwurzen	2,1	2,7	1,5	3,5	-0,6	-0,8	-0,6	-0,5
14	AT122	Niederösterreich-Süd	3,4	3,8	3,0	5,2	-0,9	-1,2	-0,8	-0,1
15	AT123	Sankt Pölten	3,3	3,7	2,9	6,0	-0,8	-1,5	-0,4	0,7
16	AT124	Waldviertel	3,0	3,7	2,4	4,7	-0,9	-1,8	-0,3	-0,1
17	AT125	Weinviertel	2,9	3,5	2,4	4,1	-0,6	-1,1	-0,3	-0,5
18	AT126	Wiener Umland/Nordteil	2,3	2,7	2,1	3,6	-0,7	-1,1	-0,4	0,0
19	AT127	Wiener Umland/Südteil	2,9	3,3	2,8	4,3	-1,0	-1,2	-0,8	-0,2
20	AT13	Wien	4,9	4,9	4,9	7,2	-1,7	-1,8	-1,7	-1,9
21	AT211	Klagenfurt-Villach	4,1	5,2	3,2	8,1	-1,1	-1,3	-1,0	-1,2
22	AT212	Oberkärnten	5,6	8,5	3,6	9,4	-1,4	-2,0	-0,9	-2,4
23	AT213	Unterkärnten	3,7	5,1	2,6	5,7	-1,2	-1,7	-0,9	-2,0

**Figure 5.** Unemployment source data.

## b) Process steps:

We start joining the figures of the indicator on unemployment with the layer of NUTS3 using the unique identifier of the NUTS3 regions, this operation allows to have a geometry representation of the information.

In order to have a single dataset holding the NUTS3 and the Reference Grid geometries, including the attribute information, we carried out an overlay process.

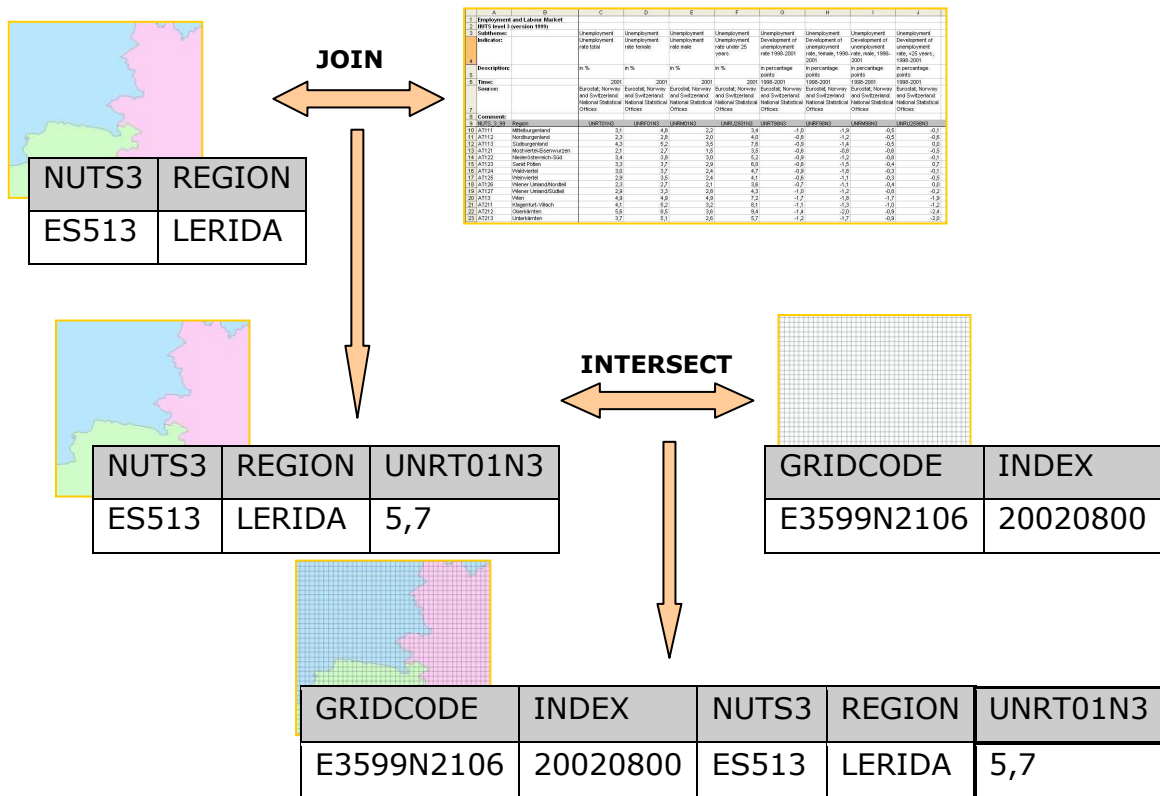
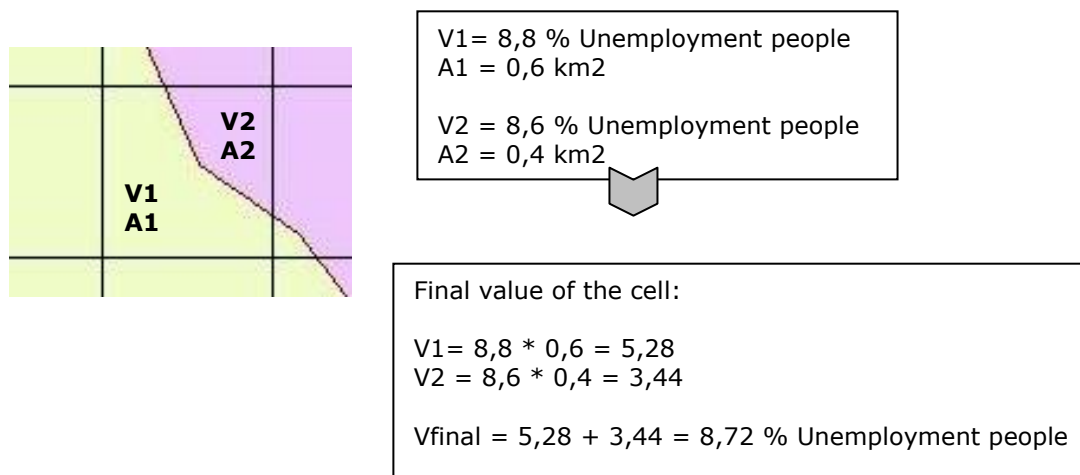


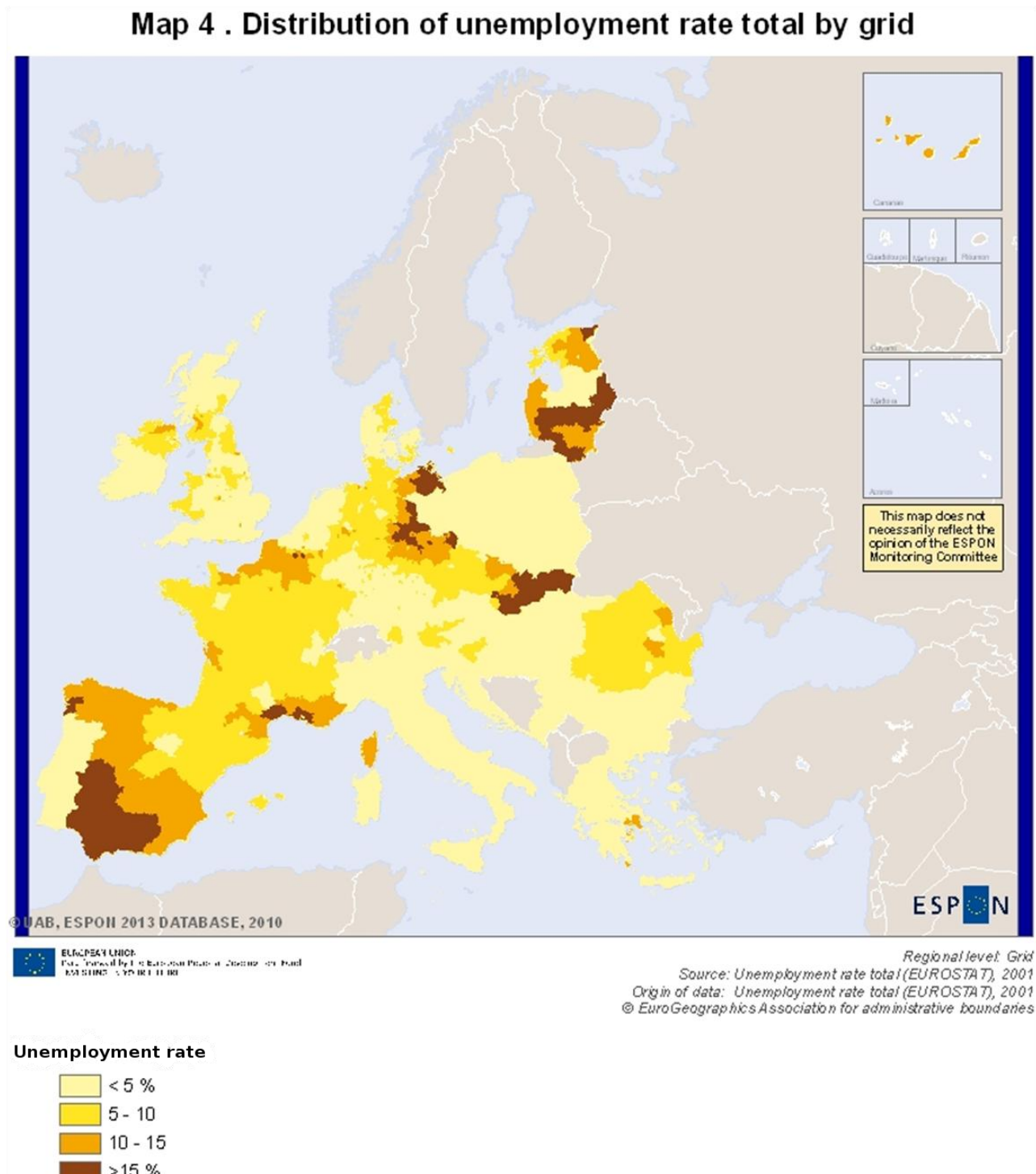
Figure 6. Creation process schema for unemployment downscaling.

After that, we should recalculate a single value for each grid cell, based on the NUTS3 value or values that fall inside. If a cell contains different NUTS3 values, the final value will be calculated depending on the share of surface of each NUTS3 region within the cell. For example, if a cell has 0.2 out of 1 covered by one NUTS3 region and 0.8 covered by a second NUTS3, the figures should be calculated accordingly (multiplying the first figure by 0.2 and the second one by 0.8).



c) Results:

The results put on a map have a look like this:



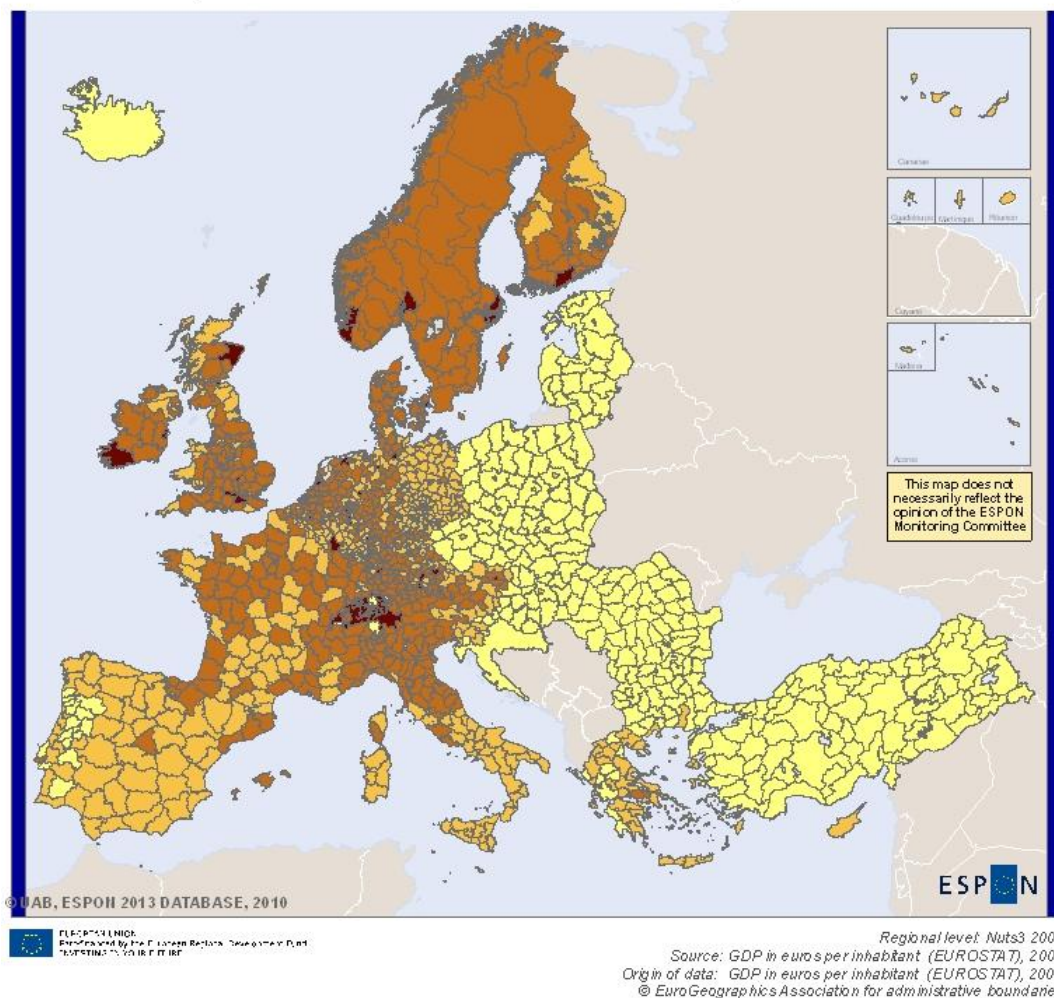
Although it seems a map where NUTS3 have been coloured as such, the data are stored by 1 km grid cell and, therefore, they can be compared with other data stored under the same grid coding, such as land cover data.



## Testing the “Proportional and weighted calculation” method: GDP – Wealth and Production (2002)

- a) Source data: In order to test the third aggregation method, we have chosen an economic variable, the GDP in euro per inhabitant 2002 (Eurostat) (Map 5. GDP €/inhab. 2002), and decided to weight its values by the population living in each 1 km grid cell. In this way, the GDP value is downscaled in a more realistic manner. As for population, we have used the JRC’s population density grid dataset<sup>11</sup> for the year 2001. In this grid, population data for communes is remapped based on Corine Land Cover classes and a quite complex algorithm<sup>12</sup>.

Map 5 . GDP €/inhab. 2002 distributed by Nuts3 2003



### Wealth and Production

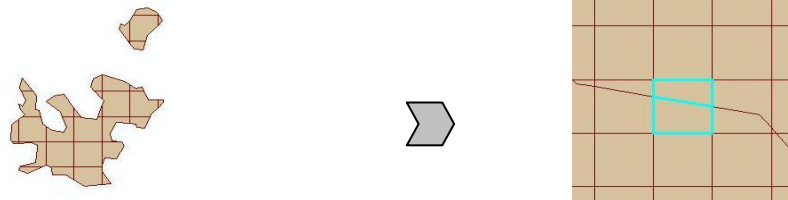
#### GDP €/inhab. 2002

	< 10000
	10000 - 20000
	20000 - 35000
	> 35000

<sup>11</sup> [http://epp.eurostat.ec.europa.eu/pls/portal/docs/PAGE/PGP\\_RESEARCH/PGE\\_RESEARCH\\_NTTS/S14P3%20-%20JAVIER%20GALLEGO%20-%20%20DOWNSCALED%20POPULATION%20DENSITY.PDF](http://epp.eurostat.ec.europa.eu/pls/portal/docs/PAGE/PGP_RESEARCH/PGE_RESEARCH_NTTS/S14P3%20-%20JAVIER%20GALLEGO%20-%20%20DOWNSCALED%20POPULATION%20DENSITY.PDF)

<sup>12</sup> <http://www.eea.europa.eu/data-and-maps/data/population-density-disaggregated-with-corine-land-cover-2000-1>

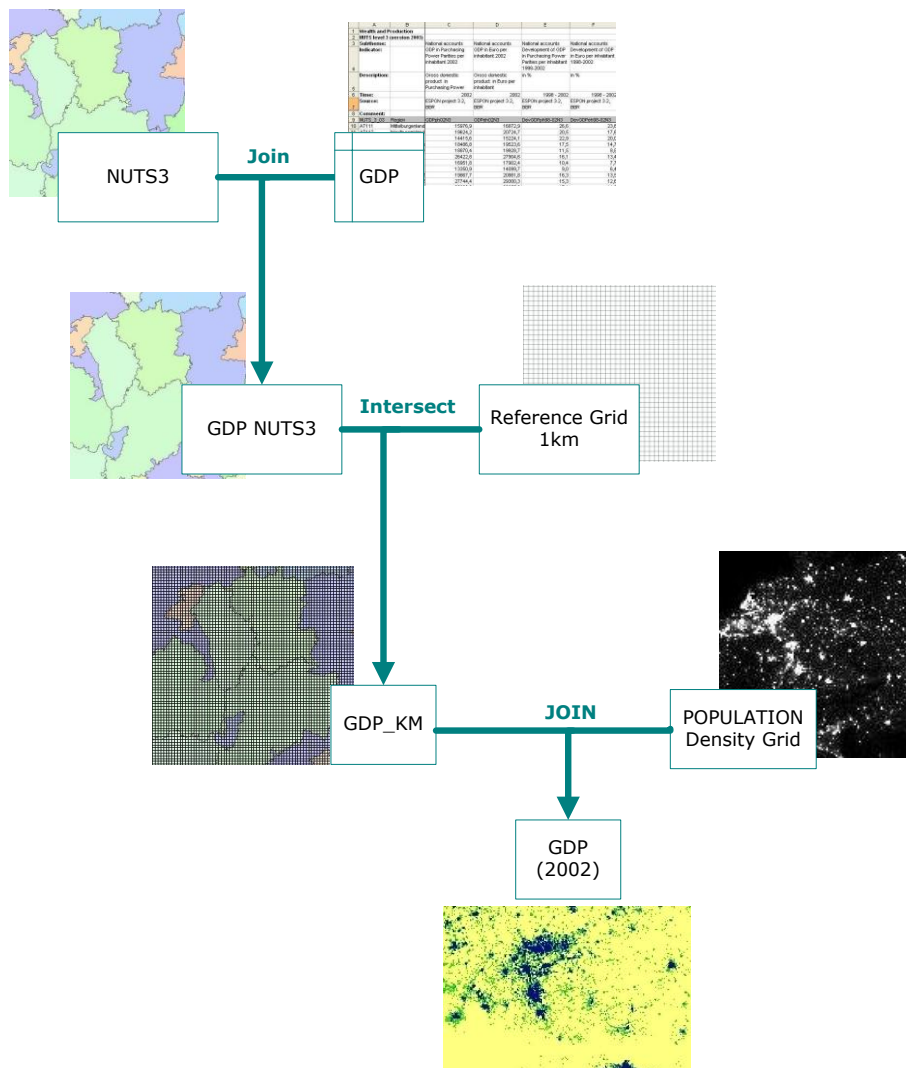
b) Process steps: In this case, the GDP is measured by NUTS3 regions. Therefore, the first step, as in the previous two cases, is overlaying the layer in which the data is given with the 1 km Reference Grid. After that, a single figure should be calculated for each 1 km grid cell, depending on the values coming from the NUTS3 regions overlaying it. If more than one value is shared by a grid cell, the final figure is calculated proportionally with regard to the area that each value occupies within the cell.



GRID CODE	AREA	GDP	GDP * AREA
GRIDCODE1	A1	GDP1	A1 * GDP1
GRIDCODE1	A2	GDP2	A2 * GDP2

GRID CODE	GDP
GRIDCODE1	A1 * GDP1 + A2 * GDP2

Finally, the value of GDP per capita that has been calculated by each grid cell is multiplied by the population figure in that cell, giving a final GDP value which reflects not only the richness of the region but also the distribution of that richness amongst the inhabitants. The next figure presents the general schema followed to calculate the indicator.



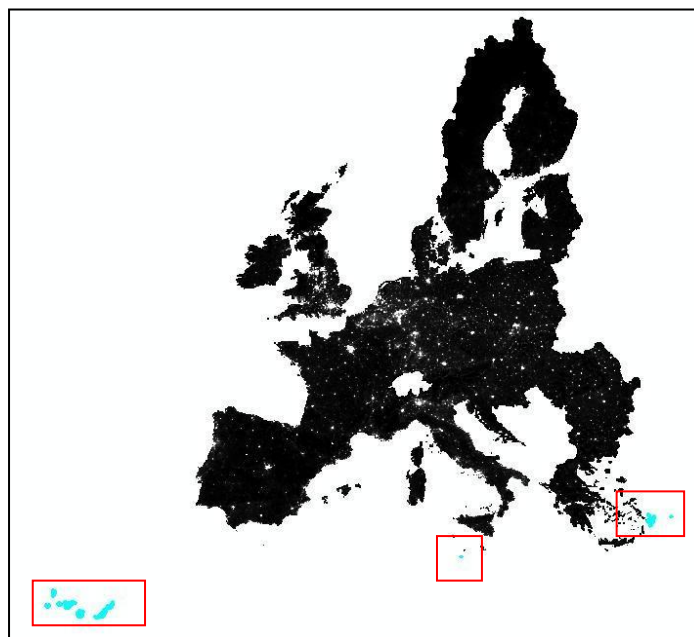
**Figure 7.** Creation process schema for GDP downscaling.

**Exceptions:** To obtain the final GDP (2002) weighted by the population it has been necessary to use the population density grid but also the population 2003 distributed by NUTS3 (Eurostat). The reason is that the population density grid doesn't cover all the extension of the layer NUTS3, this mainly happened in islands zones like Canary islands. In these regions, to be able to calculate the GDP total, the weighted process has been made with the information of the population from Eurostat.

	A	B	C	D	E	F	G
1	<b>Population</b>						
2	<b>NUTS level 3 (version 2003)</b>						
3	<b>Subtheme:</b>		Population structure	Population structure	Population structure	Population structure	Population structure
4	<b>Indicator:</b>		Average Population 2003	Average male Population, share in %, 2003	Average female Population, share in %, 2003	Population density 2002	Development average population 1995-2003 in %
5	<b>Description:</b>						
6	<b>Time:</b>		2003	2003	2003	2002	2003
7	<b>Source:</b>		Eurostat; Norway and Switzerland: National Statistical Offices	Eurostat; Norway and Switzerland: National Statistical Offices	Eurostat; Norway and Switzerland: National Statistical Offices	Eurostat; Norway and Switzerland: National Statistical Offices	Eurostat; Norway and Switzerland: National Statistical Offices
8	<b>Comment:</b>		UKM + UKN = 2002	UKM + UKN = 2002	UKM + UKN = 2002		UKM + UKN = 2002
9	<b>NUTS_3_03</b>	<b>Region</b>	<b>AvgPopN303</b>	<b>AvgmPopN303</b>	<b>AvgefPopN303</b>	<b>PopdensN302</b>	<b>DavgPop9503N3</b>
25	AT222	Liezen	81.800	48,4	51,5	24,9	1,1
26	AT223	Östliche Obersteiermark	173.700	48,4	51,6	57,4	-10,6
27	AT224	Oststeiermark	268.400	49,3	50,7	77,3	4,8
28	AT225	West- und Südsteiermark	190.700	49,1	50,9	84,5	3,0
29	AT226	Westliche Obersteiermark	108.200	49,0	50,9	36,2	-4,0
30	AT311	Innviertel	273.200	49,2	50,9	96,8	1,9
31	AT312	Linz-Wels	531.300	48,5	51,5	302,4	0,6
32	AT313	Mühlviertel	202.700	50,0	50,0	75,3	4,9
33	AT314	Steyr-Kirchdorf	152.600	49,0	51,0	69,3	1,1
34	AT315	Traunviertel	227.100	48,7	51,3	89,7	2,9
35	AT321	Lungau	21.300	49,3	50,7	21,5	-1,8

**Figure 8.** Eurostat's 2003 population source data.

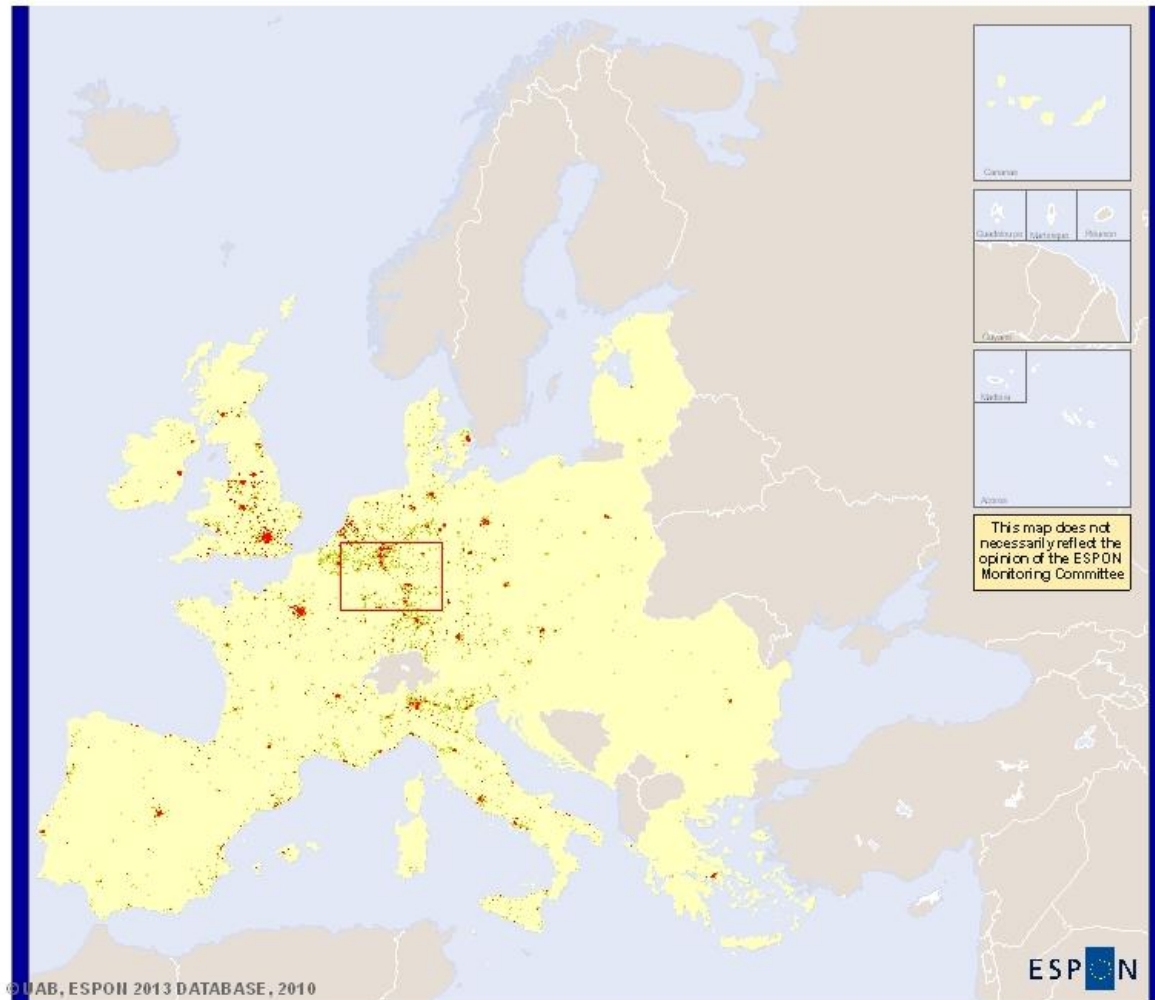
The next figure highlights in blue and red the main zones not covered by the population density grid and where it has been necessary to use the population information provided by Eurostat.



**Figure 9.** Location of areas out of population grid's scope.

Results: when we put the results on a map, we have the following layouts:

Map 6 . Distribution of GDP in Euro 2002 by grid



© UAB, ESPON 2013 DATABASE, 2010

ESPON 2013  
 Funded by the European Regional Development Fund  
 "INVESTING IN YOUR FUTURE"

ESPON

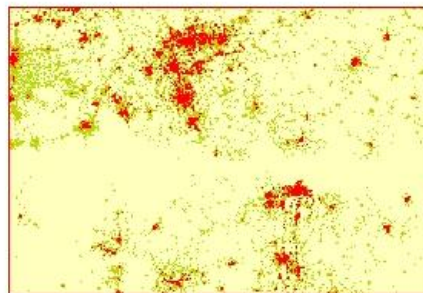
Regional level: Grid

Source: GDP in euros per inhabitant (EUROSTAT), 2002  
 Origin of data: GDP in euros per inhabitant (EUROSTAT), 2002  
 © EuroGeographics Association for administrative boundaries

**Wealth and Production**

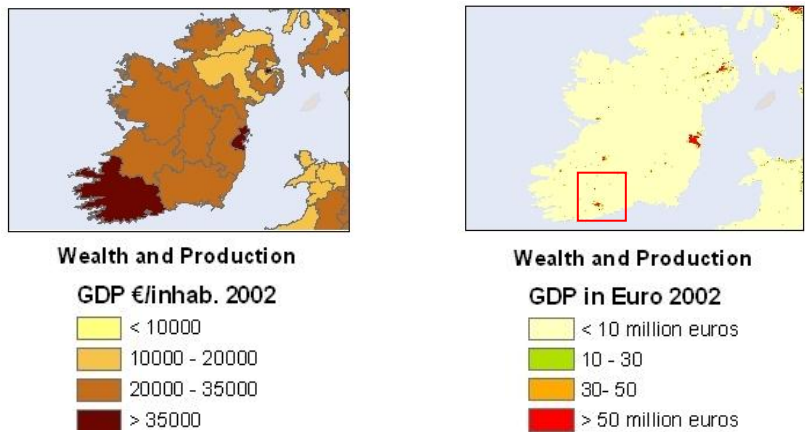
**GDP in Euro 2002**

- < 10 million euros
- 10 - 30
- 30- 50
- > 50 million euros



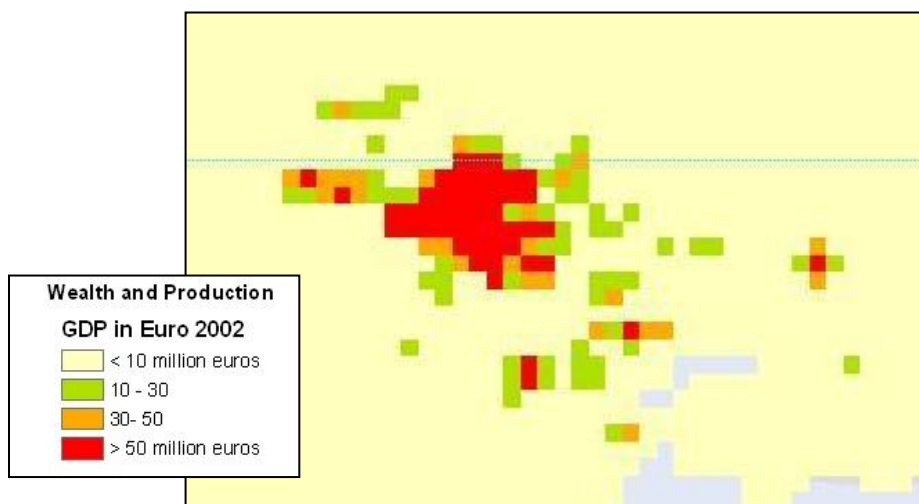
As it is obvious according to the method, the GDP is concentrated in the biggest urban areas, where most of the people are living and somehow higher in the grid cells belonging to the richest regions in Europe. Consequently, this method of redistributing and weighting data by grid cells is useful to be somehow independent of the administrative (arbitrary) divisions. This case is highlighted for example in the south-west of Ireland and in the north of Italy.

a) South-west of Ireland:



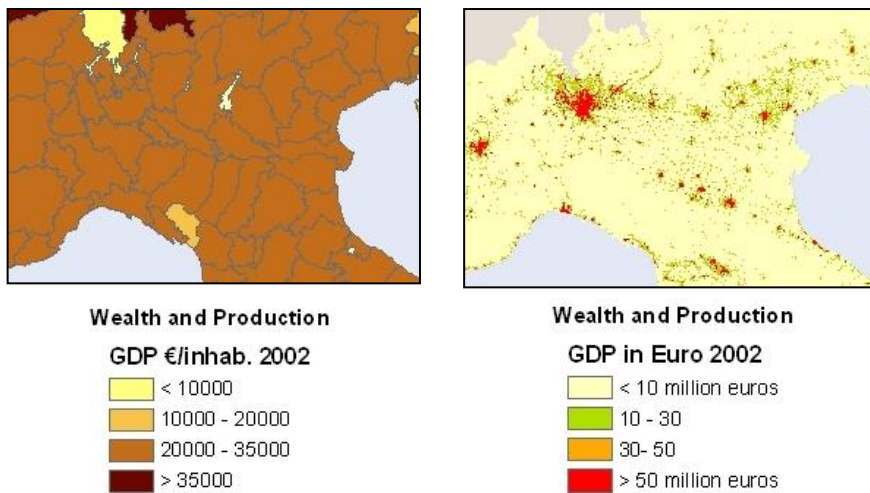
**Figure 10.** Original GDP €/inhab distributed by Nuts3 2003 vs. GDP in Euro distributed by grid at the South-west of Ireland.

In this case the Nuts3 region (IE025) is very big, but the richness is concentrated mainly around the Cork city (a small dot at the mapped scale, highlighted with a red square).



**Figure 11.** Zoom in on the Cork City at the south-west of Ireland

b) North of Italy:



**Figure 12.** Original GDP €/inhab distributed by Nuts3 2003 vs. GDP in Euro distributed by grid at the North of Italy.

In the case of the north of Italy, the regions are much smaller and have quite a high GDP/inhab. When these values are weighted by population and distributed by grid cells they better show the concentration of richness in the big cities, like Milano in this case.

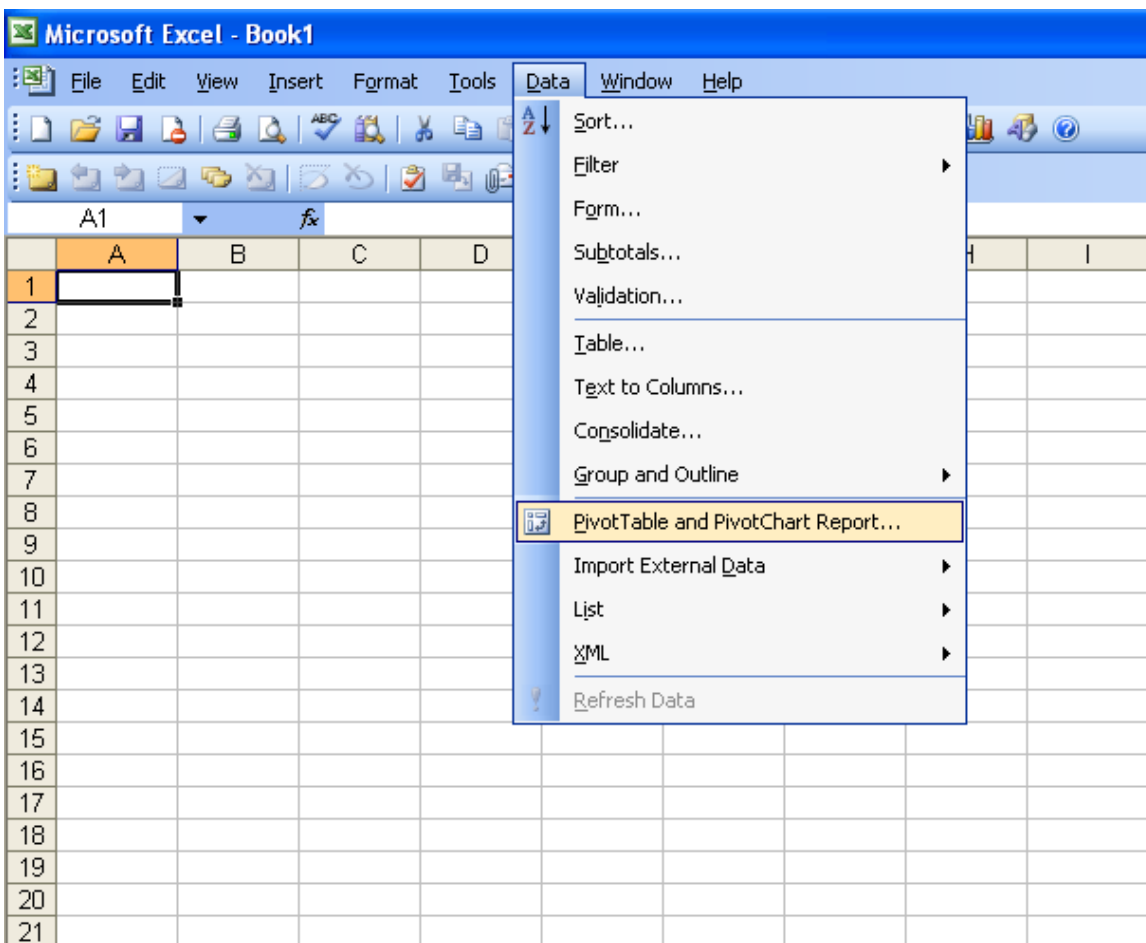
## ANNEX 2. ESPON OLAP Cube User Manual

This Annex is aimed at explaining how to query the ESPON OLAP Cube file by means of the default PivotTable application included in MS Excel.

An OLAP Cube can be queried **online** and **offline**. So far, the online connection has not been implemented. In order to use the cube, a single file **.CUB**, which works offline, has been provided.

The **.CUB** file can be connected to and queried from Microsoft Excel with a few steps detailed in the following pages:

- Select "Pivot Table..." in the Data Menu

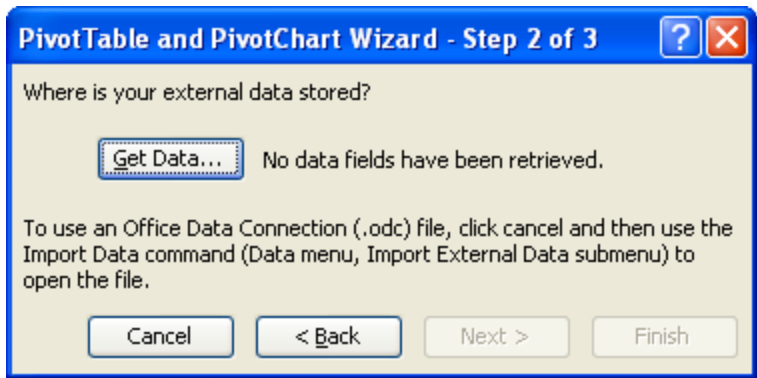




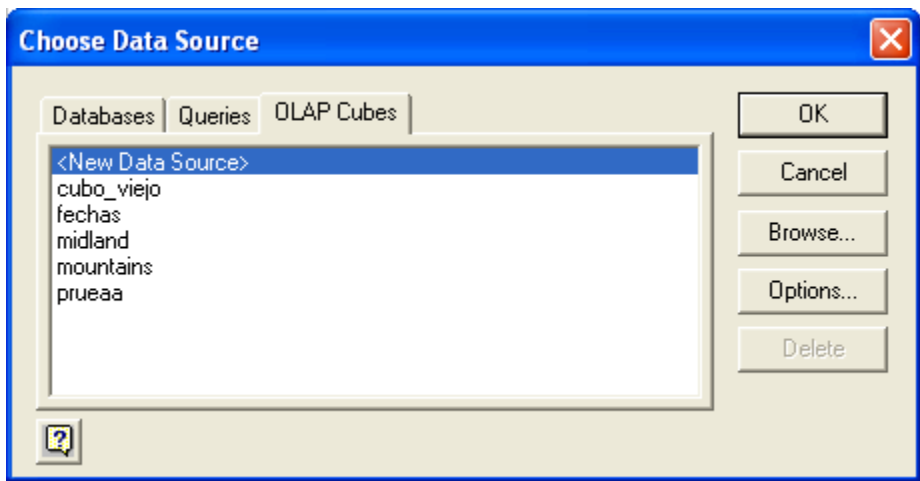
- Select “External data source”<sup>13</sup> and Pivot Table as report type



- Click on “Get Data...”

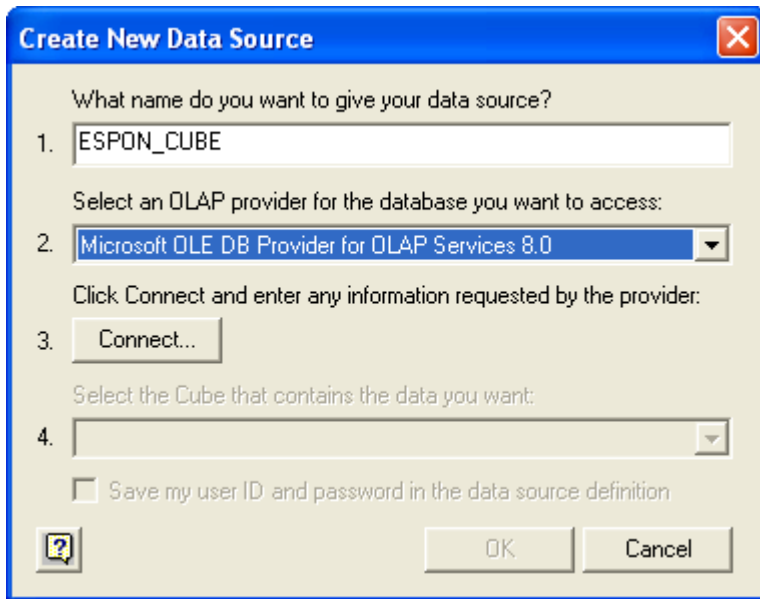


- Go to “OLAP Cubes” Tab. Choose <New Data Source>. Click OK

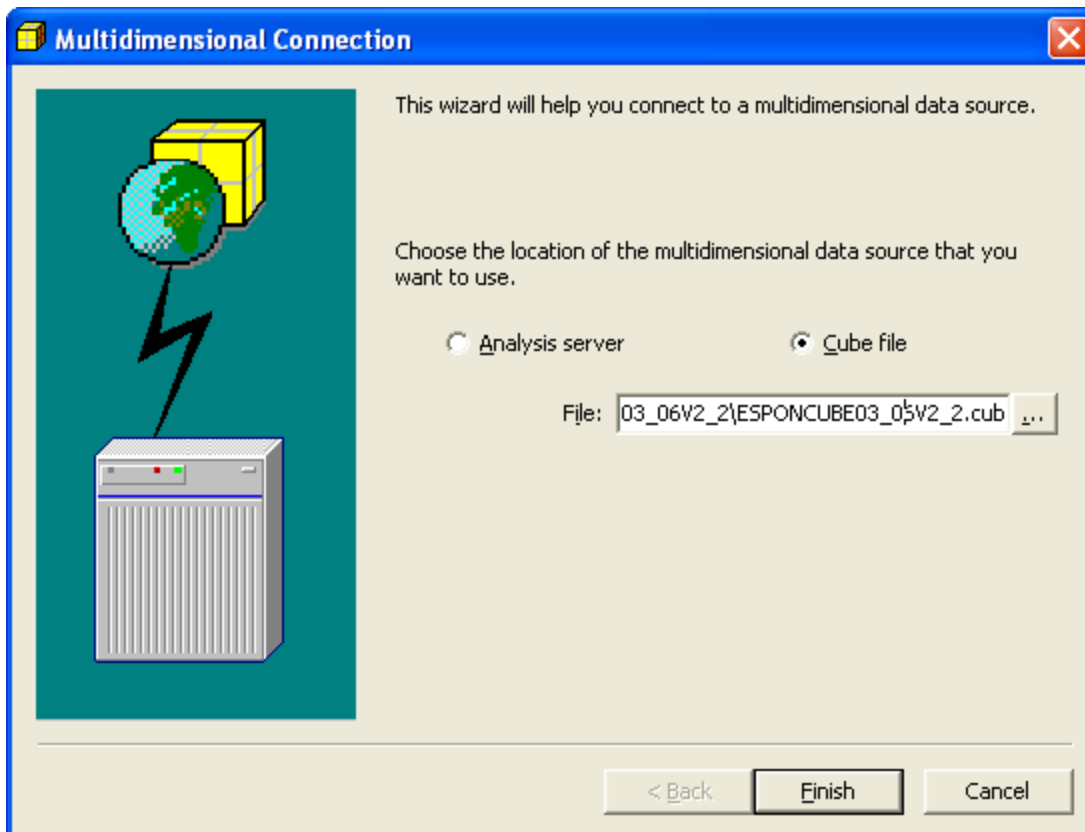


<sup>13</sup> In this case, “external data source” means that the data we are going to build the pivot table from is not coming from an Excel sheet, but from an “external” data source. In our case, this external data source is the .CUB file, i.e. the ESPON OLAP Cube.

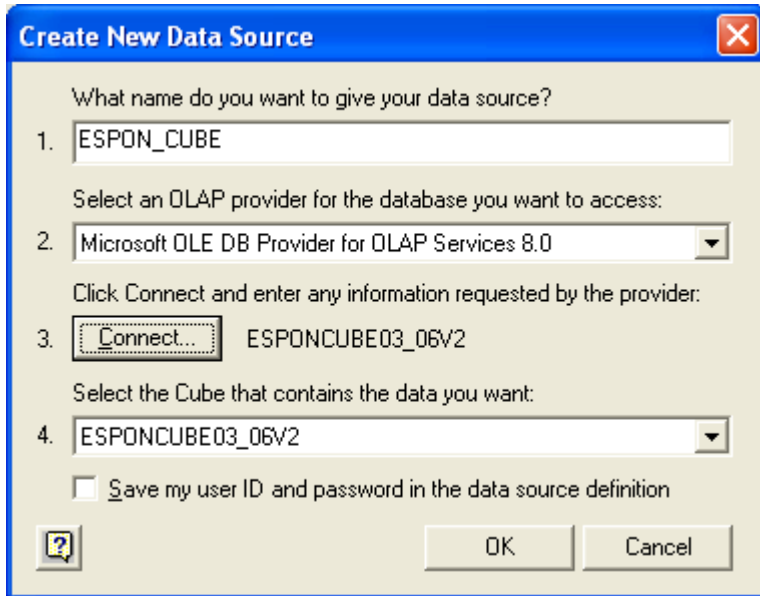
- Write down a name for your connection. Choose "MS OLE DB Provider for OLAP Services 8.0 (note: this component should be installed in order to connect to an OLAP Cube). Click Connect... button.



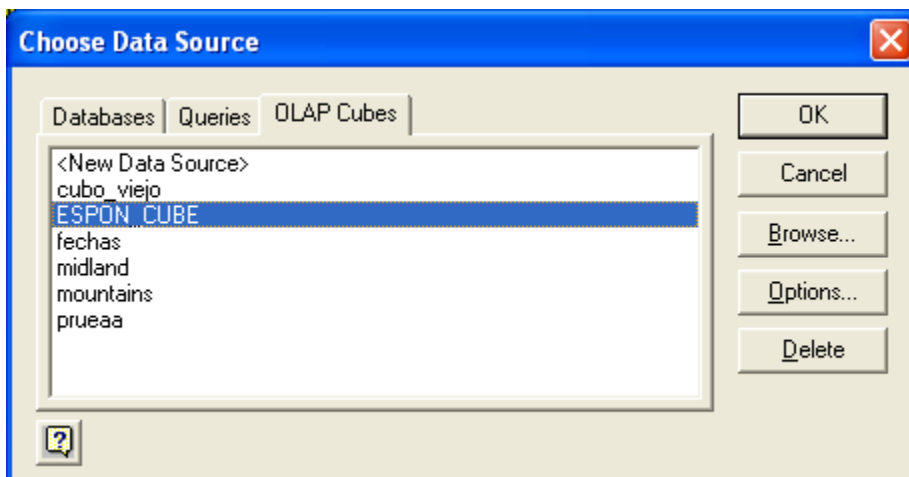
- Choose Cube file, and browse and choose the .cub file in your computer. Click Finish.



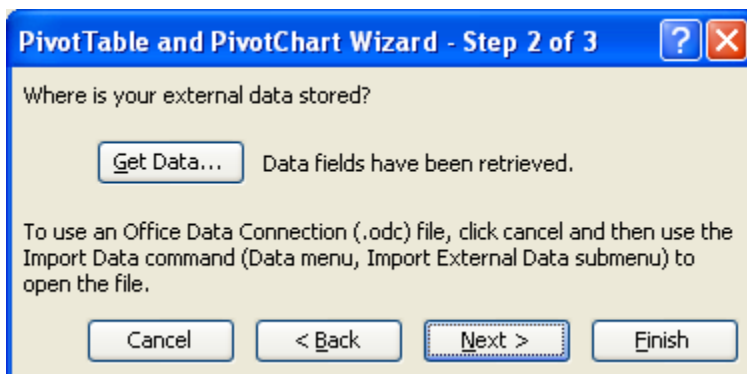
- Click OK



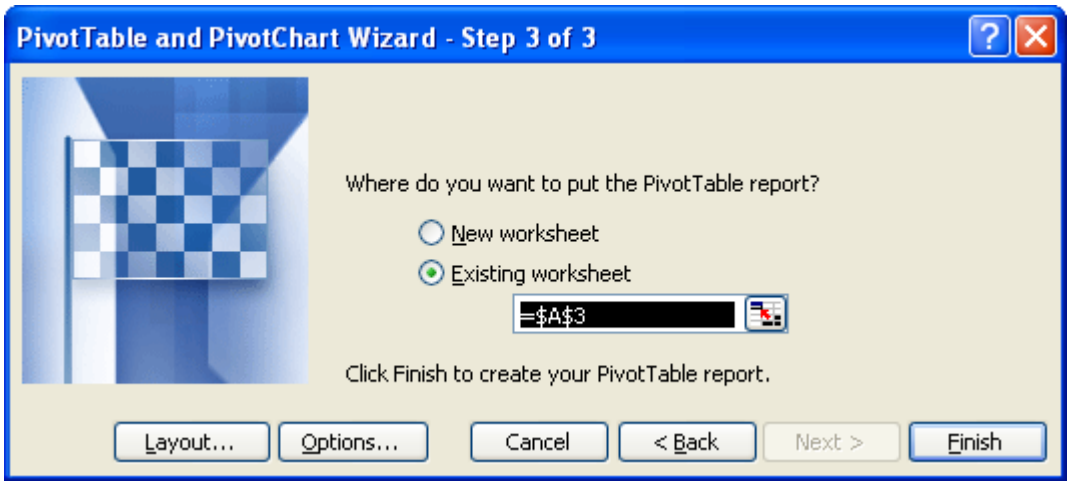
- Choose the connection just created. Click OK



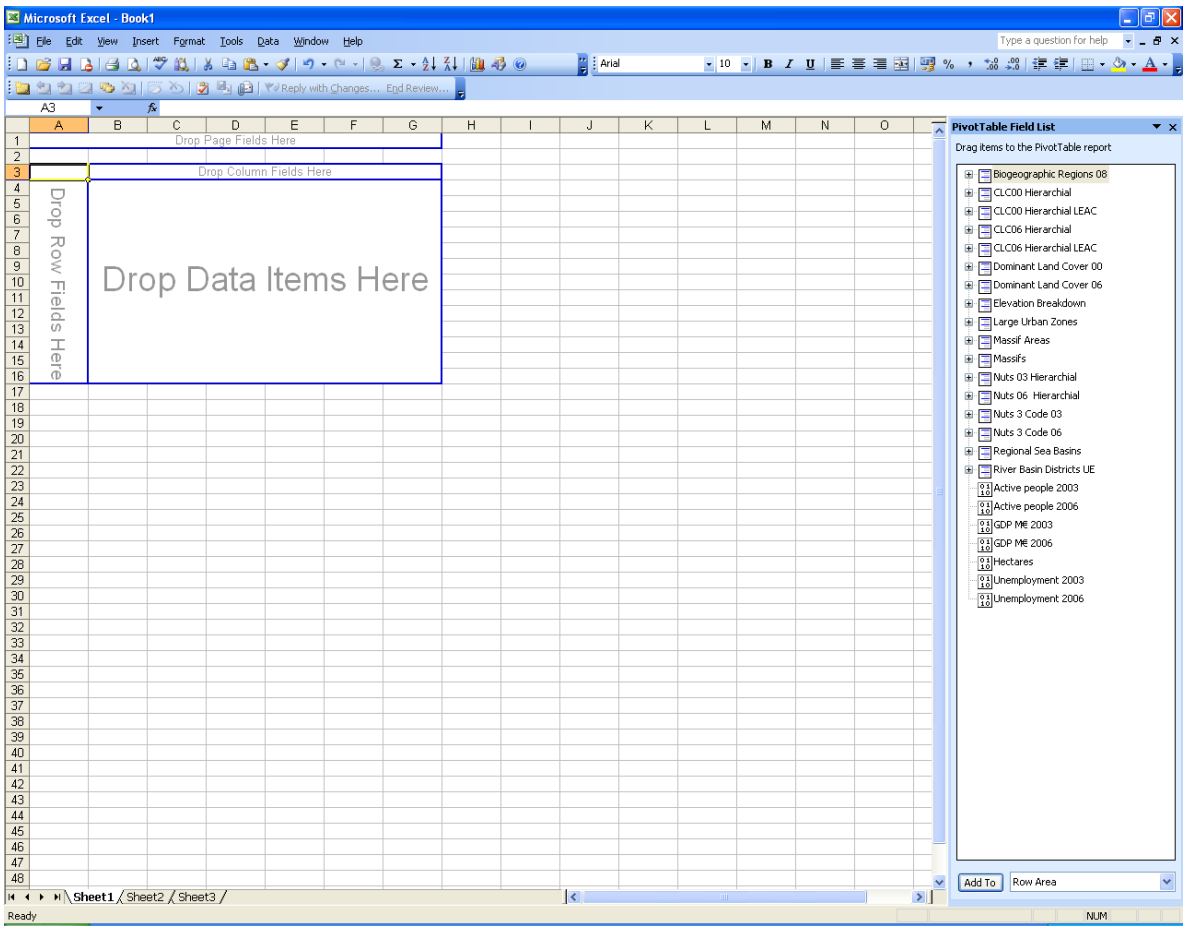
- Click "Next >"



- Choose either a new or existing worksheet. Click Finish.



- The Cube is ready!



## How to query the Cube

- Drag & Drop dimensions (e.g. Nuts 3 Code 06) in the Row Area

The screenshot shows Microsoft Excel with a PivotTable. The PivotTable is located in the range A3:G48. The Row Area contains 'Nuts 3 Code 06' and the Data Area is empty. The PivotTable Field List on the right shows 'Nuts 3 Code 06' selected and added to the Row Area. The data source is 'Regional Sea Basins'.

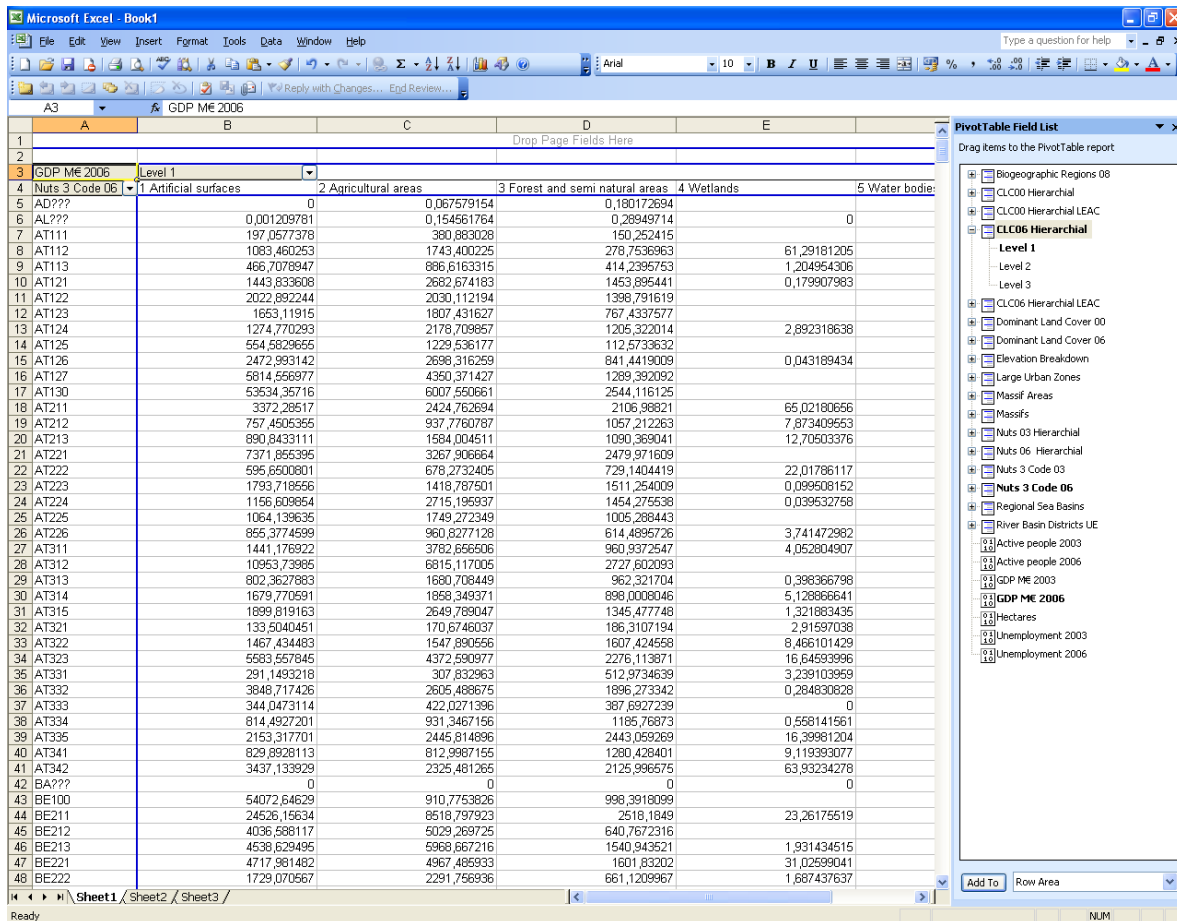
1	Drop Page Fields Here	
2		
3	Drop Column Fields Here	
4	Nuts 3 Code 06	
5	AD???	
6	AL???	
7	AT111	
8	AT112	
9	AT113	
10	AT121	
11	AT122	
12	AT123	
13	AT124	
14	AT126	
15	AT126	
16	AT127	
17	AT130	
18	AT211	
19	AT212	
20	AT213	
21	AT221	
22	AT222	
23	AT223	
24	AT224	
25	AT225	
26	AT226	
27	AT311	
28	AT312	
29	AT313	
30	AT314	
31	AT315	
32	AT321	
33	AT322	
34	AT323	
35	AT331	
36	AT332	
37	AT333	
38	AT334	
39	AT336	
40	AT341	
41	AT342	
42	BA???	
43	BE100	
44	BE211	
45	BE212	
46	BE213	
47	BE221	
48	BE222	

- Drag & Drop measures (e.g. GDP M€ 06) in the Data Area

The screenshot shows Microsoft Excel with a PivotTable. The PivotTable is located in the range A3:G48. The Row Area contains 'Nuts 3 Code 06' and the Data Area contains 'GDP M€ 2006'. The PivotTable Field List on the right shows 'GDP M€ 2006' selected and added to the Data Area. The data source is 'Regional Sea Basins'.

1	Drop Page Fields Here	
2		
3	GDP M€ 2006	
4	Nuts 3 Code 06	Total
5	AD???	0,247751848
6	AL???	0,465693633
7	AT111	728,1931807
8	AT112	3193,773684
9	AT113	1770,642889
10	AT121	5713,880087
11	AT122	5451,836716
12	AT123	4243,001456
13	AT124	4732,672501
14	AT125	1896,205201
15	AT126	6112,046373
16	AT127	11524,58537
17	AT130	63483,5594
18	AT211	8336,686231
19	AT212	2819,554547
20	AT213	3623,136297
21	AT221	13133,397441
22	AT222	2040,074413
23	AT223	4728,079903
24	AT224	5342,382776
25	AT225	3869,352016
26	AT226	2434,436218
27	AT311	6298,746306
28	AT312	21018,23391
29	AT313	3483,482771
30	AT314	4646,092557
31	AT315	6150,905421
32	AT321	493,4053386
33	AT322	4655,104042
34	AT323	12759,38914
35	AT331	1117,082467
36	AT332	9166,609866
37	AT333	1153,767175
38	AT334	2950,55961
39	AT336	7492,388255
40	AT341	2947,880575
41	AT342	8252,260815
42	BA???	0
43	BE100	56130,90852
44	BE211	36432,06288
45	BE212	9928,114352
46	BE213	12205,41349
47	BE221	11591,85324
48	BE222	4739,696783

- Drag & Drop other measures (e.g. CLC06 hierarchical Level 1) in the Column Area



- Play around with the ESPON OLAP Cube to build new queries.
- You can also build Pivot charts.
- If you are using MS Excel 2007, please check:

<http://office.microsoft.com/en-gb/excel/CH062528071033.aspx>

- Documentation about OLAP and OLAP Cubes:

[http://en.wikipedia.org/wiki/Online\\_analytical\\_processing](http://en.wikipedia.org/wiki/Online_analytical_processing)

[http://en.wikipedia.org/wiki/OLAP\\_cube](http://en.wikipedia.org/wiki/OLAP_cube)



## SPATIAL ANALYSIS FOR QUALITY CONTROL

### *Phase 1: The identification of logical input errors and statistical outliers*

#### MAIN RESULTS

- Exceptional values can arise from logical input errors and true outlying data.
- The accurate identification of an exceptional value is important as input errors should be treated differently to true outlying data.
- Input errors can usually be identified mathematically or sometimes, statistically. Outliers are identified statistically.
- Techniques to statistically identify outliers are presented using worked examples that have been coded with R open source software.

ESPON 2013 DATABASE

MARCH 2011



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

81 pages

# LIST OF AUTHORS

Paul Harris, National Centre for Geocomputation (NCG), National University of Ireland (Maynooth)

Martin Charlton, National Centre for Geocomputation (NCG), National University of Ireland (Maynooth)

## **Contact**

Paul.Harris@nuim.ie

martin.charlton@nuim.ie

tel. + 353-(0)1-7086208



# TABLE OF CONTENT

<b>Introduction.....</b>	<b>3</b>
<b>1 Exceptional values: types and identification.....</b>	<b>4</b>
1.1 Logical input errors .....	4
1.2 Aspatial statistical outliers: identification in univariate to multivariate data sets .....	4
1.3 Spatial statistical outliers: identification in univariate data sets.....	5
1.4 The use of statistical models and residual data in outlier identification .	6
1.5 The identification of spatial clusters.....	7
1.6 Summary: MAUP, temporal outliers and data imputation .....	7
1.7 Further reading.....	8
<b>2 Data for worked examples .....</b>	<b>9</b>
2.1 The full data set.....	9
2.2 Data subsets and analytical objectives.....	11
2.3 A data subset with deliberate logical input errors .....	12
<b>3 Worked examples: commented R scripts and results ....</b>	<b>19</b>
3.1 The R statistical environment .....	19
3.2 Worked example 1: univariate & residual analyses for input errors & outliers .....	19
3.3 Worked example 2: univariate & residual analyses for outliers.....	20
3.4 Worked example 3: multivariate analyses for outliers.....	21
3.5 Worked example 4: multivariate residual analyses for outliers .....	22
3.6 Worked example 5: identification of spatial clusters .....	22
3.7 Worked example 6: some consequences of MAUP .....	24
<b>4 Discussion and further developments .....</b>	<b>28</b>
References .....	29
Appendix 1 – R script for worked example 1 .....	32
Appendix 2 – R script for worked example 2 .....	41
Appendix 3 – R script for worked example 3 .....	49
Appendix 4 – R script for worked example 4 .....	57
Appendix 5 – R script for worked example 5 .....	64
Appendix 6 – R script for worked example 6 .....	72

## Introduction

The ESPON 2013 Database should be as free from errors as possible. It follows from this that detecting errors is an important activity in both data entry and data checking. This technical report is to examine how mathematical, statistical and spatial analysis tools can be applied to the ESPON 2013 Database in order to find 'logical input errors' and 'statistical outliers'. In both cases, 'exceptional values' can arise but it is not always clear if such values relate to input errors or true values that are statistically-outlying. In this respect, reliably determining the nature of an exceptional value is important, especially as input errors should be treated differently to statistical outliers. For example, input errors are usually corrected or removed, whilst suspected outliers are usually flagged for further scrutiny.

The outcome of this report is a targeted review of existing outlier-detection tools in the field of statistics, data mining and spatial analysis, and an examination how they can assist in the detection of errors/outliers in the ESPON 2013 Database for improved quality control. This methodological review has a clear focus on spatial analysis with respect to outlier-detection; and is complemented by worked examples on an ESPON-type data set, where chosen techniques are demonstrated. Worked examples are coded using open-source software so that the applied techniques are easily transferable. The list of techniques that are applied should not be considered as exhaustive, but form a cross-section of useful techniques which are appropriate for ESPON 2013 Database.

A related aim of this report is to examine the effects of the Modifiable Areal Unit Problem (MAUP) with respect to error/outlier identification. This follows previous research by NCG for the ESPON 2006 project on this topic (ESPON 2006).

# 1 Exceptional values: types and identification

## 1.1 Logical input errors

Logical input errors can arise for a number of reasons. For example, the wrong NUTS1 code could be specified; incorrect data values could be input; data could be repeated exactly but assigned to different variables; data could be displaced within or between columns; data could be swapped within or between columns. In general, the identification of an input error will follow some logical, mathematical approach. For example, if a land use class could only take a positive integer value from 1 to 9 say, then an input error of say, -2, 4.5 or 10 would be easily identified.

An input error may also be identified statistically. For example, if the number 27 is inadvertently entered as 72 for a region's unemployment rate, the value 72 may lie in the extreme tail of this variable's distribution and as such, is statistically-outlying. A difficulty here would be to distinguish between an input error of 72 and a true value of 72.

In this respect, when dealing with errors/outliers, most input errors can be either be corrected or removed, whilst most outliers should be flagged as: (i) suspected outliers and (ii) potential (undetected) input errors. Flagged observations would then require further scrutiny, which should ascertain whether the observation should be: (a) replaced; (b) removed; or if specifically an outlier, (c) retained or possibly down-weighted in some way (so as to provide some robust model fit or statistic of the data).

## 1.2 Aspatial statistical outliers: identification in univariate to multivariate data sets

A simple, graphical tool for the detection of outliers in univariate data sets is the boxplot (e.g. Frigge et al. 1989). Central to the creation of the boxplot is the inter-quartile range (Q3-Q1) around the median value Q2. Commonly, at the upper end of the distribution, the *inner fence* is defined as the value given by  $Q2 + 1.5(Q3 - Q1)$  and the *outer fence* as the value given by  $Q2 + 3(Q3 - Q1)$ ; and there are corresponding values for the lower end of the distribution. Observations whose values lie between the inner and outer fences are usually referred to as *outside* and those whose values lie beyond the outer fence are usually referred to as *far out*. In either case, such observations can be flagged as outlying, however most attention should be placed on observations that lie beyond the outer fence. In this report, we not only demonstrate the use of the standard boxplot but also an adjusted boxplot for skewed distributions (Hubert and Vandervieren 2008). For bivariate data sets, a simple extension of the boxplot, the bagplot (Rousseeuw et al. 1999) can be constructed.

---

<sup>1</sup> NUTS stands for "nomenclature of territorial units for statistics".

To detect outliers in multivariate data sets, we first demonstrate a technique where outliers are observations that have a *large* squared Mahalanobis Distance ( $MD^2$ ), where the MD itself is estimated in a robust manner (Filzmoser et al. 2005). MDs are used as they take into account the covariance matrix from which the shape and size of the multivariate data set can be quantified. In this outlier detection technique, robust  $MD^2$  values are related to some pre-determined (upper) quantile of a chi-square distribution (e.g. the 97.5<sup>th</sup> percentile), where *large* robust  $MD^2$  values lie above this pre-determined threshold. Furthermore, to address subjectivity in choosing the threshold, the technique automatically adjusts the pre-determined threshold (downwards or upwards) via simulation reflecting specific properties of the sample data. The technique (called here RMD2-AQ-outlier) is applied incorporating useful graphical displays of suspected outliers.

We also demonstrate two further multivariate techniques that each use principal component analysis (PCA) to reduce the dimensions of the multivariate data set, where in the resultant transformed space, outliers may be more readily observable. Of the many PCA-based techniques for outlier detection that have been proposed (e.g. see Rousseeuw et al. 2006; Daszykowski et al. 2007; Filzmoser et al. 2008), we demonstrate: (a) the 'sign' approach of Locantore et al. (1999) (call this technique, PCA-outlier-1) and (b) the 'PCOut' approach of Filzmoser et al. (2008) (call this technique, PCA-outlier-2). Both techniques are computationally fast and thus suited to large, high dimensional data sets (see the comparisons given in Filzmoser et al. 2008).

### 1.3 Spatial statistical outliers: identification in univariate data sets

Commonly outlier detection techniques ignore any spatial element to the data. Data not observed as an outlier when an *aspatial* technique is used, may nevertheless be a *spatial* outlier. Therefore it is important to consider spatial aspects if false negatives (i.e. outliers undetected by an aspatial technique) are to be avoided. In this respect, we demonstrate a technique of Hawkins (1980) to detect spatial outliers in univariate data sets<sup>2</sup>. This technique has much in common with the more recent techniques of Lui et al (2001); Kou et al. (2005).

For this technique, all observations  $z_i$  are suspected a priori as spatial outliers, where  $z_i$  is a spatial outlier if

$$\left| \frac{z_i - m_l}{\bar{s}_l} \right| \geq \chi_{crit-1}^2 \quad (1)$$

Here,  $i = 1, \dots, n$ ;  $\mathbf{x}$  is spatial location;  $N$  is the number of neighbouring values of  $z_i$ ;  $m_l$  is the local mean;  $\bar{s}_l^2$  is the average variance for equivalently sized neighbourhoods across the sample area (i.e. the average local variance) and  $\chi_{crit-1}^2$  is

<sup>2</sup> We only present a technique to identify spatial outliers in a univariate sense. Extensions to bivariate and multivariate spatial data sets are not considered here. However our current research in this area concerns the development of geographically weighted PCA techniques with respect to outlier identification (see Charlton et al. 2010), which should allow the identification of multivariate spatial outliers in the ESPON database.

a critical value of the chi-squared distribution for 1 degree of freedom. As there is no objective function for cross-validation, then neighbourhood definitions (for the local mean and variances) are chosen subjectively for this test statistic. In this report, the local mean and variances are found using a geographically weighted approach (see sections 2.4 and 2.5), with 95%, 99% and 99.9% critical levels chosen as appropriate cut-offs.

## 1.4 The use of statistical models and residual data in outlier identification

In a statistical analysis, it is common to identify outliers via large (positive or negative) prediction errors (or residuals) from some predictive model fit. Observations that are poorly predicted produce large residuals when compared with the actual data, and are therefore deemed as outlying. The key drawback to this approach is the need to specify a model in the first place, where different models may produce different outlying observations. However if several prediction models are applied, then it is reasonable to expect that the most influential outlying observations should be repeatedly identified.

In this respect, we first identify outliers (in a univariate sense) simply using the key component of expression 1, where a spatial outlier relates to a large (absolute) value of the error  $z_i \rightarrow m_i$ . Here our prediction model is simply the one chosen to find the local mean  $m_i$ , which in this case is some simple spatial predictor using geographical weights (which we shall call the local mean predictor, LM). The widely-used inverse distance weighting model would be one example of such an LM model.

Furthermore, we also identify outliers (via residual data) using univariate and multivariate regressions in both aspatial and spatial forms. In particular we apply: (a) standard multiple linear regression (MLR) models, (b) attribute-space local regression (LR) models (see Loader 2004) and (c) geographic-space local regression models (Fotheringham et al. 2002) (i.e. geographically weighted regression, GWR). Here LR accounts for nonstationarity and nonlinearity in attribute-space, whilst GWR accounts for nonstationary and nonlinearity in geographic-space. Both LR and GWR are nonparametric in design. The conventional MLR model assumes stationarity and linearity in both attribute- and geographic-space; and is parametric in design. Consequently, each of the three regression forms will identify outliers (or possibly groups of outliers, see section 2.5) according to their particular specification (or set of modelling assumptions).

The investigation of residual data plays a central role in the formulation of a robust regression model, where the influence of outlying data on the regression fit is reduced (e.g. see Faraway 2004, p98-106; Cruz Ortiz et al. 2006). MLR, LR (see Loader 2004) and GWR (Fotheringham et al. 2002, p73-82; Harris et al. 2010) all have robust forms. Commonly, a robust regression will identify outliers as observations with large standardised (or studentised) residuals via a leave-one-out approach. However, in this report we only identify outliers simply, via the raw residuals and without the benefit of a leave-one-out fit.

## 1.5 The identification of spatial clusters

A group of observations identified as outliers may actually be spatially clustered with a substantive reason for their 'unusualness' (i.e. false positives are to be avoided as well). In this respect, it is worthwhile applying techniques that identify local (or regional) changes in the spatial process according to some key moment or relationship<sup>3</sup>.

Furthermore, seemingly significant clusters can be sometimes be attributable to only a few (influential and outlying) observations; so although the local techniques described below are not specifically designed to identify spatial outliers, they sometimes do so. Indeed, a corresponding robust form of the given local technique would out of necessity identify spatial outliers in order to reduce their influence.

Thus in the first instance, local summary univariate and bivariate statistics are calculated and investigated. In particular, we assess changes in the mean, standard deviation and correlation across space, where these (spatial) moments are all found in a geographically weighted form (Fotheringham et al. 2002)<sup>4</sup>. For the multivariate case, GWR can be applied, which complements a local correlation analysis when investigating relationship-change across space.

From a spatial autocorrelation viewpoint, a local version of Moran's I (Anselin 1995) is used. Positive spatial autocorrelation exists when neighbouring spatial units tend to have similar values of a variable; whilst negative spatial autocorrelation exists when they do not. Local Moran's I is only used to investigate univariate data, but the statistic could be adapted to investigate cross-autocorrelation in bivariate and multivariate data sets.

## 1.6 Summary: MAUP, temporal outliers and data imputation

We have presented a typology of techniques where variables are analysed singly or in combination; and aspatially or spatially. Underlying all of these techniques is the spatial structure of the reporting units, where results can be influenced not only by the level of spatial aggregation used but also by the spatial configuration of the reporting units (i.e. a MAUP; e.g. see Wong 1996). In this report we demonstrate the consequences of the MAUP for outlier identification via a worked example, where outlier-detection techniques are applied at different NUTS levels (NUTS level 3 through to NUTS level 0).

We have not addressed the identification of temporal (or by extension, spatio-temporal) outliers. This is not an oversight, as ESPON time series data is not expected to be of a sufficient length for an outlier detection technique to be reliably

---

<sup>3</sup> Brunson and Charlton (2010) assess the effectiveness of multiple hypothesis testing for detecting clusters of geographical anomalies. These tests would complement the techniques demonstrated from this section of the report.

<sup>4</sup> Robust forms of geographically weighted summary statistics (GWSS) can be found in Brunson et al. (2002) and in Harris and Brunson (2010).

applied. Instead it should suffice that the aspatial/spatial detection methods demonstrated here can be repeated at different time intervals. The consequences of the reporting units changing over time (i.e. another MAUP) are addressed elsewhere in ESPON 2013 database project.

As already discussed, once an input error has been identified the observation can either be corrected or removed (i.e. replaced with the missing value notation, NA5). On the other hand, suspected outliers (which may be an input error) can (after some additional scrutiny) be: (a) replaced; (b) removed (i.e. replaced by NA); or if indeed an outlier, (c) retained or possibly down-weighted in some way. This entails that some form of imputation or prediction of missing valued data will be required, and here the chosen regression models of section 2.4 may be of value.

## 1.7 Further reading

This report provides a brief overview to subject of error or outlier identification with respect to the task of identifying outliers in the ESPON 2013 Database. There is an extensive literature on outlier detection, where the following reading list may be useful.

- An evaluation of aspatial techniques to detect input errors and true outliers (here known as data editing), together with imputation techniques, for large scale survey data can be found in [Charlton \(2004\)](#). This and related articles arose from the EUREDIT project<sup>6</sup>. Related articles include: an outlier identification technique for multivariate data by [Béguin and Hulliger \(2004\)](#); a robust regression technique for data edits by [Chambers et al. \(2004\)](#); and a classification and regression tree technique for data edits by [Petraikos et al. \(2004\)](#).
- An aspatial Bayesian technique that both edits and imputes data in a multivariate context can be found in [Ghosh-Dastidar and Schafer \(2003\)](#).
- Reviews of aspatial outlier identification techniques from univariate to multivariate data sets can be found in [Reimann et al. \(2005\)](#); [Rousseeuw et al. \(2006\)](#); [Daszykowski et al. \(2007\)](#); [Morgenthaler \(2007\)](#).
- Further aspatial outlier identification techniques for multivariate data sets can be found in [Hoo et al. \(2002\)](#); [Jackson and Chen \(2004\)](#), where the former article also imputes data.
- Imputation (aspatial) techniques can be found in [Plaia and Bondi \(2006\)](#); [Vanden Branden and Verboven \(2009\)](#), where the former article focuses on time series data.
- Alternative spatial outlier identification techniques can be found in [D'Alimonte and Cornford \(2007\)](#); [Ainsworth and Dean \(2008\)](#); [Meiklit et al. \(2009\)](#).

---

<sup>5</sup> NA is the missing data indicator used in the R statistical computing package (see section 4).

<sup>6</sup> See <http://www.cs.york.ac.uk/euredit/>. The project website was still active as of 1/12/09.

## 2 Data for worked examples

In the worked examples, NUTS3 level data are used. Here 1351 values (with two missing) for the variable 'evolution of gross domestic product (GDP) from the years 2000 to 2005' at NUTS2006 divisions are related to sixteen contextual variables at NUTS1999 divisions (with a maximum of 1329 values for each contextual variable). As the NUTS2006 spatial units can differ to the NUTS1999 spatial units, this combining of data results in at least 438 (1351 minus 913) missing values for each contextual variable (i.e. NUTS2006 and NUTS1999 divisions have 913 reporting units in common). Thus in summary, NUTS3 level data using the NUTS2006 divisions are the spatial units that are retained.

### 2.1 The full data set

The 'evolution of GDP' variable is named EVOGDP\_2000\_2005\_2006, where the first two numbers (2000 and 2005) relate to the collection time (i.e. year) of the data and the last number (2006) relates to the NUTS division or version. Similar naming conventions were used for all other variables. EVOGDP\_2000\_2005\_2006 is itself calculated from four stock variables which are presented in [Table 1](#), together with the formula for calculating EVOGDP\_2000\_2005\_2006.

The sixteen contextual variables are presented in [Tables 2 to 6](#). These variables were selected from the basic and project indicator files posted on the ESPON website<sup>7</sup>. Contextual data include: two spatial typology variables, one unemployment variable, six land use variables, one natural hazards variable and six regional policy variables. In total, the full data set consists of twenty-three variables (plus the coordinates/centroids of each region).

Observe that as variables were collected over different time periods (from 1996 to 2005) this data set is purely used to demonstrate the outlier identification techniques of [section 2](#) via the worked examples in [section 4](#). It is essentially a fabricated data set and as such, all analytical results need to be interpreted with this in mind.

However the contextual variables were selected in expectation that if all variables were reliable (i.e. collected over the same period), then this particular set of contextual variables may help explain variation in EVOGDP\_2000\_2005\_2006 (see [sections 2.4, 4.5 and 4.6](#)).

---

<sup>7</sup> See [http://www.espon.eu/mmp/online/website/content/tools/832/850/588\\_EN.html](http://www.espon.eu/mmp/online/website/content/tools/832/850/588_EN.html) and [http://www.espon.eu/mmp/online/website/content/tools/832/873/605\\_EN.html](http://www.espon.eu/mmp/online/website/content/tools/832/873/605_EN.html)



Variable type	Variable name	Indicator	Year	Unit
STOCK (1)	GDP_2000_2006	Gross Domestic Product	2000	Million Euros
STOCK (2)	GDP_2005_2006	Gross Domestic Product	2005	Million Euros
STOCK (3)	POP_T_2000_2006	Total population (annual average)	2000	Thousands inhabit.
STOCK (4)	POP_T_2005_2006	Total population (annual average)	2005	Thousands inhabit.
RATIO (5)	GDP_POP_2000_2006	GDP per inhabit. = $\frac{\text{GDP}}{\text{POP}} \times 1000$	2000	Euros
RATIO (6)	GDP_POP_2005_2006	GDP per inhabit. = $\frac{\text{GDP}}{\text{POP}} \times 1000$	2005	Euros
RATIO	EVOGDP_2000_2005_2006	Evolution of GDP = $\frac{\text{GDP}_{2005} - \text{GDP}_{2000}}{\text{GDP}_{2000}} \times 100$	2000-2005	Percentage

**Table 1:** Description of the EVOGDP\_2000\_2005\_2006 variable

Theme Indicator	Spatial typology	Spatial typology
	Typology Settlement Structure (nine basic types defined by population density and situation regarding centres) – 1: city core region; 2: very densely populated; 3: densely populated; 4: rural region; 5: city core region; 6: densely populated region; 7: rural region; 8: more densely populated region; 9: less densely populated region	Urban-rural typology (six basic types) – 1: High urban influence, high human intervention; 2: High urban influence, medium human intervention; 3: High urban influence, low human intervention; 4: low urban influence, high human intervention; 5: Low urban influence, medium human intervention; 6: Low urban influence, low human intervention
Original variable name	Settyp99N3	URTypN3
New variable name	SPAT_TYPE_1_1999_1999	SPAT_TYPE_2_1999_1999
Min. possible	1	1
Max. possible	9	6
Unit or variable type	CLASS	CLASS

**Table 2:** Descriptions of spatial typology contextual variables

Theme Indicator	Unemployment	Land use	Land use	Land use
	Unemployment rate	Share of artificial surfaces	Artificial surfaces per 1000 inhabitants	Artificial surfaces per GDP
Original variable name	UNRT01N3	ArSu96N3	ArSc96N3	ArSg96N3
New variable name	UNEMP_R_2001_1999	LU_AS_1_1996_1999	LU_AS_2_1996_1999	LU_AS_3_1996_1999
Min. possible	0	0	0	0
Max. possible	100	100	100	100
Unit or variable type	PERCENTAGE	PERCENTAGE	PERCENTAGE	PERCENTAGE

**Table 3:** Descriptions of unemployment and three land use contextual variables

Theme Indicator	Land use	Land use	Land use	Environment - Hazards
	Share of urban fabric	Share of arable land	Share of permanent crops	Sum of all weighted hazard values
Original variable name	UFL296N3	ALL296N3	PCL296N3	smwh04
New variable name	LU_UF_1996_1999	LU_AR_1996_1999	LU_PC_1996_1999	NAT_HAZ_2004_1999
Min. possible	0	0	0	10
Max. possible	100	100	100	INFINITY
Unit or variable type	PERCENTAGE	PERCENTAGE	PERCENTAGE	INTEGER

**Table 4:** Descriptions of three land use and an environmental hazards contextual variable

<b>Theme Indicator</b>	<b>Regional policy</b> All Structural & Cohesion Fund expenditure	<b>Regional policy</b> Structural Fund expenditure related to Regional Development & Productive Infrastructure	<b>Regional policy</b> Structural Fund expenditure related to Social Integration & Human Resources
<b>Original variable name</b>	SFT99N3	SFR99N3	SFS99N3
<b>New variable name</b>	SF_CF_1999_1999	SF_R_1999_1999	SF_S_1999_1999
<b>Min. possible</b>	0	0	0
<b>Max. possible</b>	INFINITY	INFINITY	INFINITY
<b>Unit or variable type</b>	REAL NUMBER	REAL NUMBER	REAL NUMBER

**Table 5:** Descriptions of three regional policy contextual variables

<b>Theme Indicator</b>	<b>Regional policy</b> Structural Fund expenditure related to Agriculture, Rural Development & Fishery	<b>Regional policy</b> Cohesion Fund expenditure related to Transport	<b>Regional policy</b> Cohesion Fund expenditure related to Environment
<b>Original variable name</b>	SFA99N3	SFCT99N3	SFCE99N3
<b>New variable name</b>	SF_A_1999_1999	CF_T_1999_1999	CF_E_1999_1999
<b>Min. possible</b>	0	0	0
<b>Max. possible</b>	INFINITY	INFINITY	INFINITY
<b>Unit or variable type</b>	REAL NUMBER	REAL NUMBER	REAL NUMBER

**Table 6:** Descriptions of three regional policy contextual variables

## 2.2 Data subsets and analytical objectives

Subsets of the full data set are analysed in two basic forms: (a) in their original state and (b) in a state where some commonly encountered logical input errors are deliberately introduced. Here [Tables 7 and 8](#) summarise how variable subsets of the full data set are used in each of six worked examples presented in [section 4](#).

<b>Worked example</b>	<b>1</b>	<b>2</b>	<b>3</b>
<b>Variables investigated</b>	NUTS3 code, GDP_2000_2006, GDP_2005_2006, POP_T_2000_2006 & POP_T_2005_2006	EVOGDP_2000_2005_2006 (plus the coordinate data)	Some subset of EVOGDP_2000_2005_2006, its 16 contextual variables & the coordinate data
<b>Introduced input errors?</b>	Yes	No	No
<b>Identification type: logical or statistical or both</b>	Both	Statistical	Statistical
<b>Identification type: univariate or multivariate or both</b>	Univariate	Univariate	Multivariate
<b>Identification type: aspatial or spatial or both</b>	Aspatial	Both	Aspatial
<b>Key statistical identification techniques applied</b>	Boxplots	Boxplots; Hawkins test; residuals from LM, MLR, LR & GWR fits	Bagplots; Robust MD <sup>2</sup> analysis (RMD2-AQ-outlier); & PCA for outliers (PCA-outlier-1 & PCA-outlier-2)
<b>Analysis objective</b>	Identify logical input errors so that EVOGDP_2000_2005_2006 can be investigated for statistical outliers	Identify statistical outliers	Identify statistical outliers

**Table 7:** Data subsets and analytical objectives for *worked examples 1 to 3*

<b>Worked example</b>	<b>4</b>	<b>5</b>	<b>6</b>
<b>Variables investigated</b>	EVOGDP_2000_2005_2006 in relation to some subset of its 16 contextual variables and the coordinate data	EVOGDP_2000_2005_2006 in relation to some subset of its 16 contextual variables and the coordinate data	EVOGDP_2000_2005_2006 (plus the coordinate data)
<b>Introduced input errors?</b>	No	No	No
<b>Identification type: logical or statistical or both</b>	Statistical	Statistical	Statistical
<b>Identification type: univariate or multivariate or both</b>	Multivariate	Both	Univariate
<b>Identification type: aspatial or spatial or both</b>	Both	Spatial	Both
<b>Key statistical identification techniques applied</b>	Residuals from MLR, LR & GWR fits	Data exploration with GWSS, GWR & local Moran's I	Boxplots; Hawkins test; residuals from LM, MLR, LR & GWR fits
<b>Analysis objective</b>	Identify statistical outliers	Identify statistical clusters	Investigate the consequences of MAUP with respect to outlier identification

**Table 8:** Data subsets and analytical objectives for *worked examples 4 to 6*

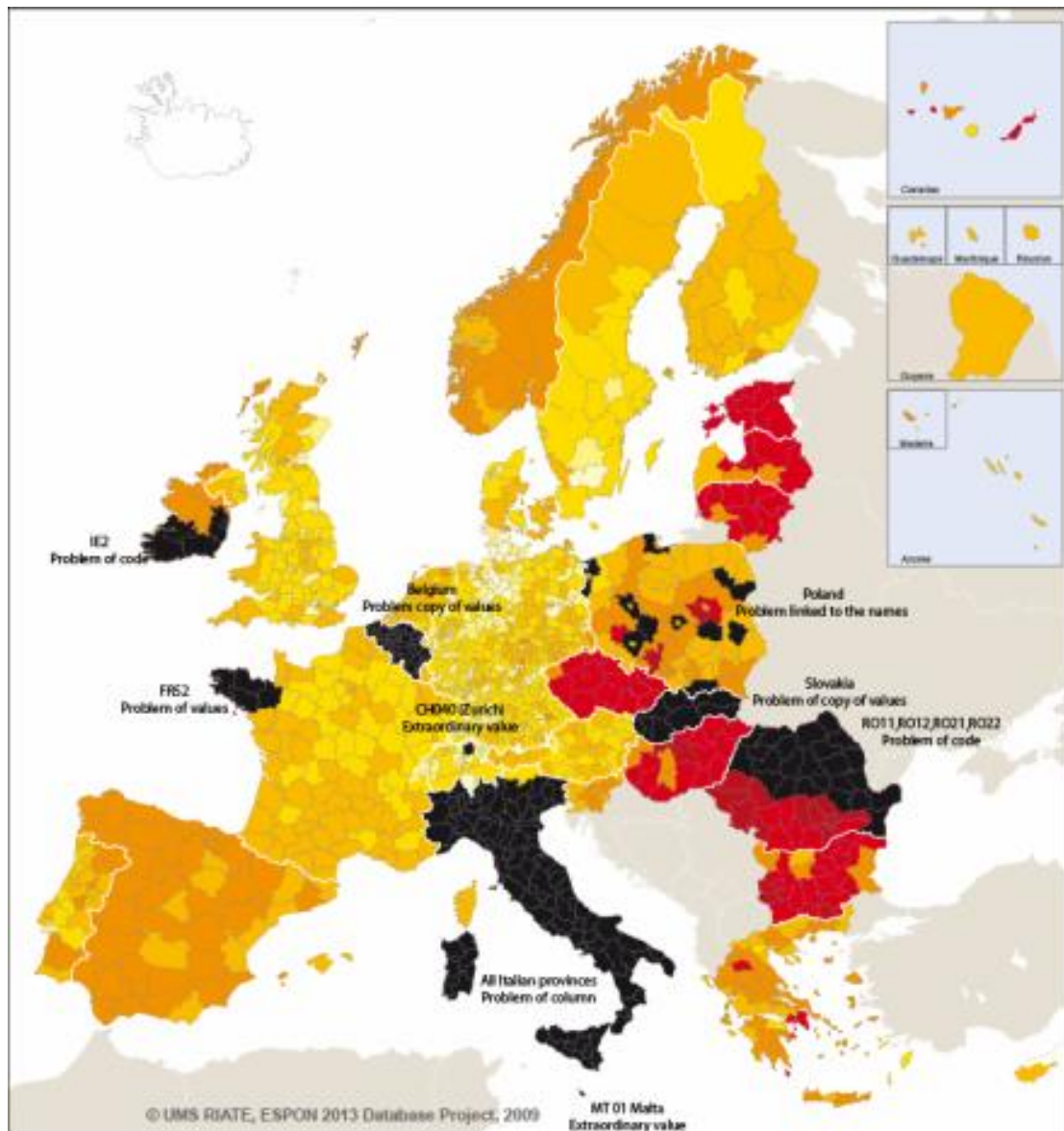
It is envisaged that when identifying exceptional values in an ESPON database data set, a first pass should identify input errors using both mathematical and statistical techniques. That is the first pass screens the data. Identified input errors should then be corrected (and as such, a revised data set can be *assumed* input error-free) before a second pass is undertaken that only uses the statistical techniques to identify outlying observations. It is essential that two passes are conducted otherwise the detection of true outliers will be compromised by input errors.

Thus from [Tables 7 and 8](#), [worked example 1](#) relates to a first pass (for input errors) before its corresponding second pass (for outliers), which is (effectively) [worked example 2](#). For [worked examples 2, 3, 4, 5 and 6](#), it should be assumed that this data has already been screened for input errors. Observe that [worked example 6](#) is the same as [worked example 2](#), but applied at different NUTS levels (i.e. spatial scales) to investigate the effects of MAUP with respect to outlier identification.

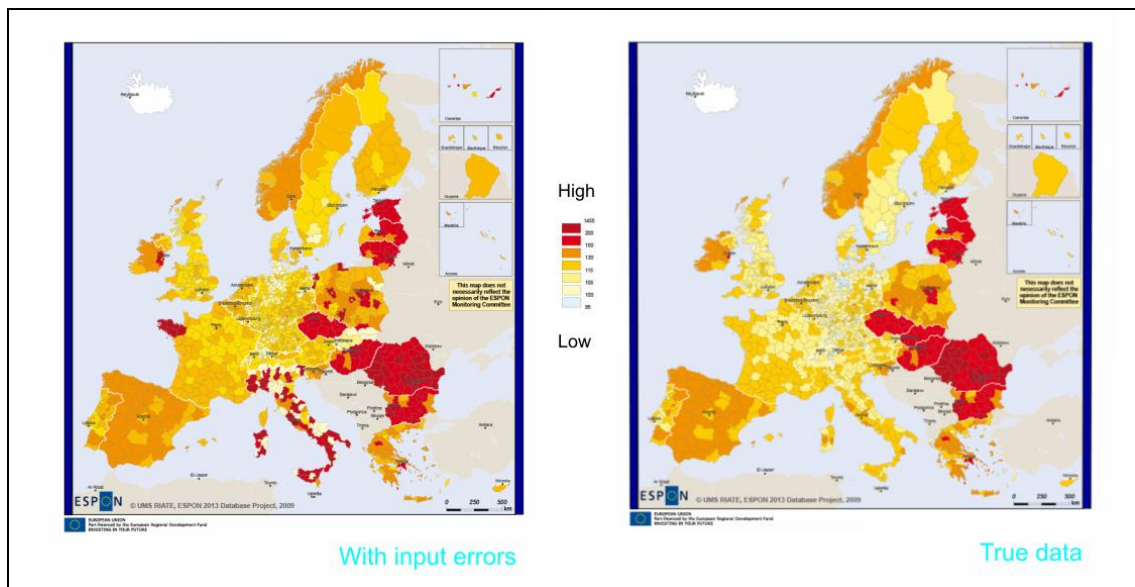
## 2.3 A data subset with deliberate logical input errors

We now present a list of logical input error-types that have been introduced to: (a) the NUTS3 codes; and (b) the variables GDP\_2000\_2006, GDP\_2005\_2006, POP\_T\_2000\_2006 and POP\_T\_2005\_2006 (i.e. only those variables used in the calculation of EVOGDP\_2000\_2005\_2006). This list of input errors is given with appropriate solutions (i.e. for [worked example 1](#)).

This list is not exhaustive, and should grow as different input error-types become apparent (i.e. at this stage, we are not expected to foresee all input error possibilities). The spatial location of the input errors is depicted in [Fig. 1](#). Consequences of input errors for the correct calculation of EVOGDP\_2000\_2005\_2006 are depicted in [Fig. 2](#).



**Figure 1:** Location of input errors (in black) overlaid on the true EVOGDP\_2000\_2005\_2006 data (see Fig. 2)



**Figure 2:** Maps of EVOGDP\_2000\_2005\_2006 with and without input errors

### Problems with NUTS code (29 input errors)

**Input error-type 1:** For Ireland, 5 wrong codes have been input at NUTS3 level. In the NUTS hierarchy, this does not imply changes at NUTS2 level (see Fig. 3a). Solution: codes can be checked by a simple relationship to the correct NUTS name and code pairs.

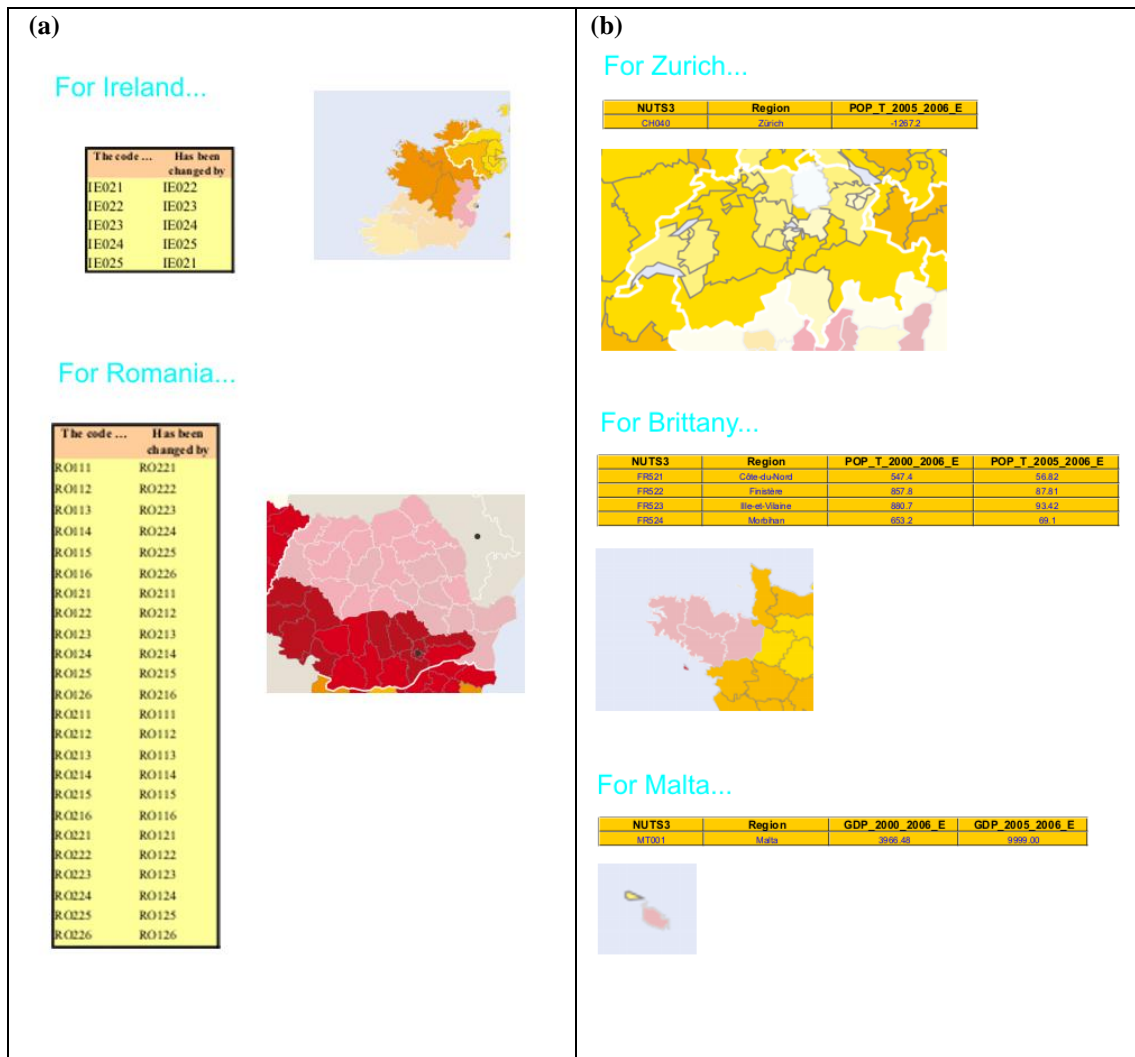
**Input error-type 2:** For Romania, 24 wrong codes have been input at NUTS3 level. In the hierarchy, this does imply changes at NUTS2 level (see Fig. 3a). Solution: codes can be checked by a simple relationship to the correct NUTS name and code pairs.

### Problems with values (6 input errors)

**Input error-type 3:** For Zürich (NUTS3 - CH040), the total population in 2005 (POP\_T\_2005\_2006) has been multiplied by -1 (see Fig. 3b). This value is impossible for this variable and as such, should be easily identified.

**Input error-type 4:** For Brittany (NUTS2 - FR52), the total population in 2005 (POP\_T\_2005\_2006) has been divided by 10 at the NUTS3 level (all 4 of them, see Fig. 3b). These values are possible, but should be easily identified by a simple subtraction of both population variables (POP\_T\_2005\_2006 minus POP\_T\_2000\_2006) and looking for unusually large (negative) differences. Large (negative) differences could be identified as statistically outlying (which upon further scrutiny would indicate *potential* input errors).

**Input error-type 5:** For Malta (NUTS3 - MT001), the total GDP in 2005 (GDP\_2005\_2006) has been incorrectly entered as 9999. 9999 is sometimes used to denote a missing value (see Fig. 3b). This value is possible, but should be identified when all *potential* missing values (e.g. values of -99, -999, -9999, 99, 999, 9999, etc.) are identified for further scrutiny.



**Figure 3 (a):** incorrect NUTS code entries (b) incorrect value entries

### Problems with copied or repeated data (52 input errors)

**Input error-type 6:** For Belgium (44 entries), the total population in 2000 (POP\_T\_2000\_2006) is repeated exactly for the total population in 2005 (POP\_T\_2005\_2006) (all at NUTS3 level, see Fig. 4). These values are possible, but should be easily identified by a simple subtraction of the two population variables and looking for (exact) zero values. It can be assumed that equal populations for the two years are highly unlikely. Also observe that differences of zero are unlikely to be statistically outlying. A difficulty here would be to decide whether the values for POP\_T\_2000\_2006 or the values for POP\_T\_2005\_2006 were inputted incorrectly. This would require further scrutiny.

**Input error-type 6:** For Slovakia (8 entries), the total GDP in 2000 (GDP\_2000\_2006) is repeated exactly for the total GDP in 2005 (GDP\_2005\_2006) (all at NUTS3 level, see Fig. 4). These values are possible, but again should be easily identified by a simple subtraction of the two GDP variables and looking for zero values. As with the population data, it can be assumed that equal GDP data for the two years is highly unlikely. Again, it is difficult to know whether the values for GDP\_2000\_2006 or the values for GDP\_2005\_2006 were inputted incorrectly.

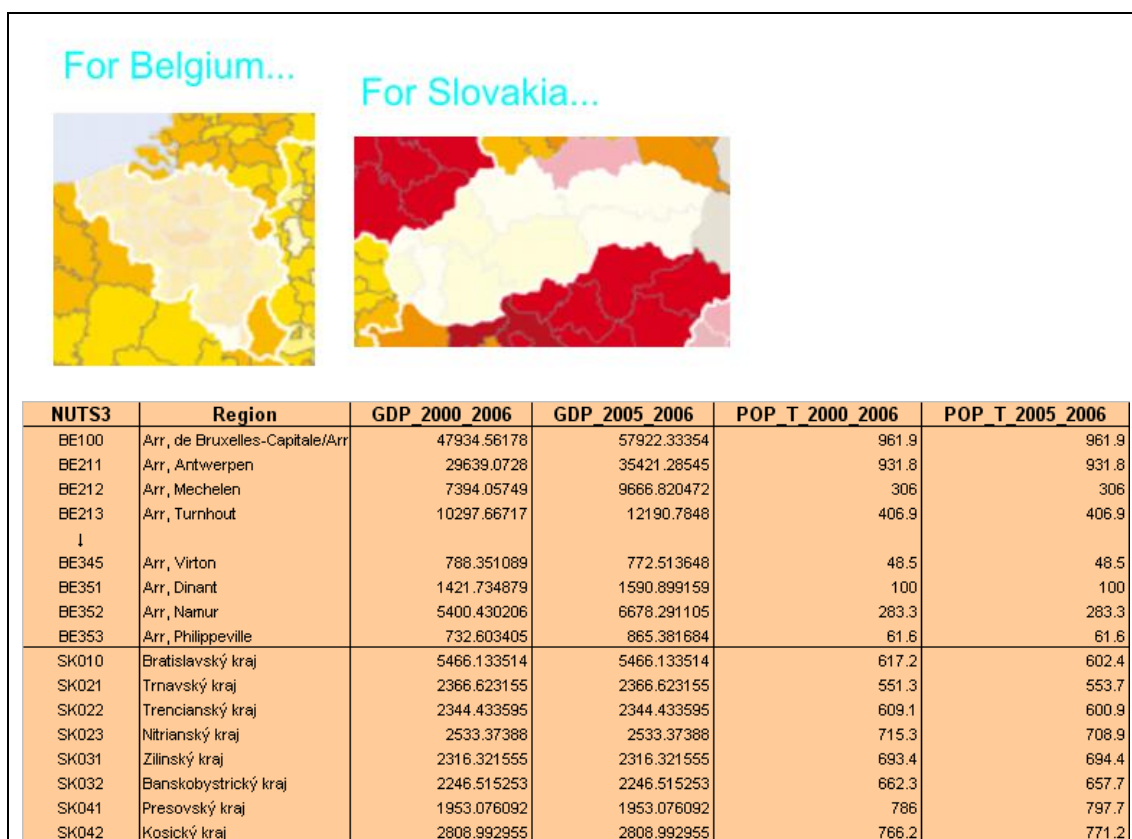


Figure 4: Problems of copied or repeated data

**Shift in data values (up one or down one line in its data column) (107 input errors)**

**Input error-type 7:** For Italy (107 entries), the total population in 2005 (POP\_T\_2005\_2006) has been shifted up by one line (all at NUTS3 level, see Fig. 5). These values are possible, but most values (not all) should be statistically identified as *potential* input errors. Again a subtraction of the two population variables should for most cases, produce unusually large positive or unusually large negative values which should give rise to suspicion.

Observe that this error-type has created an extra missing value (NUTS3 - ITG2C) and one value has effectively been lost (NUTS3 - ITC11).

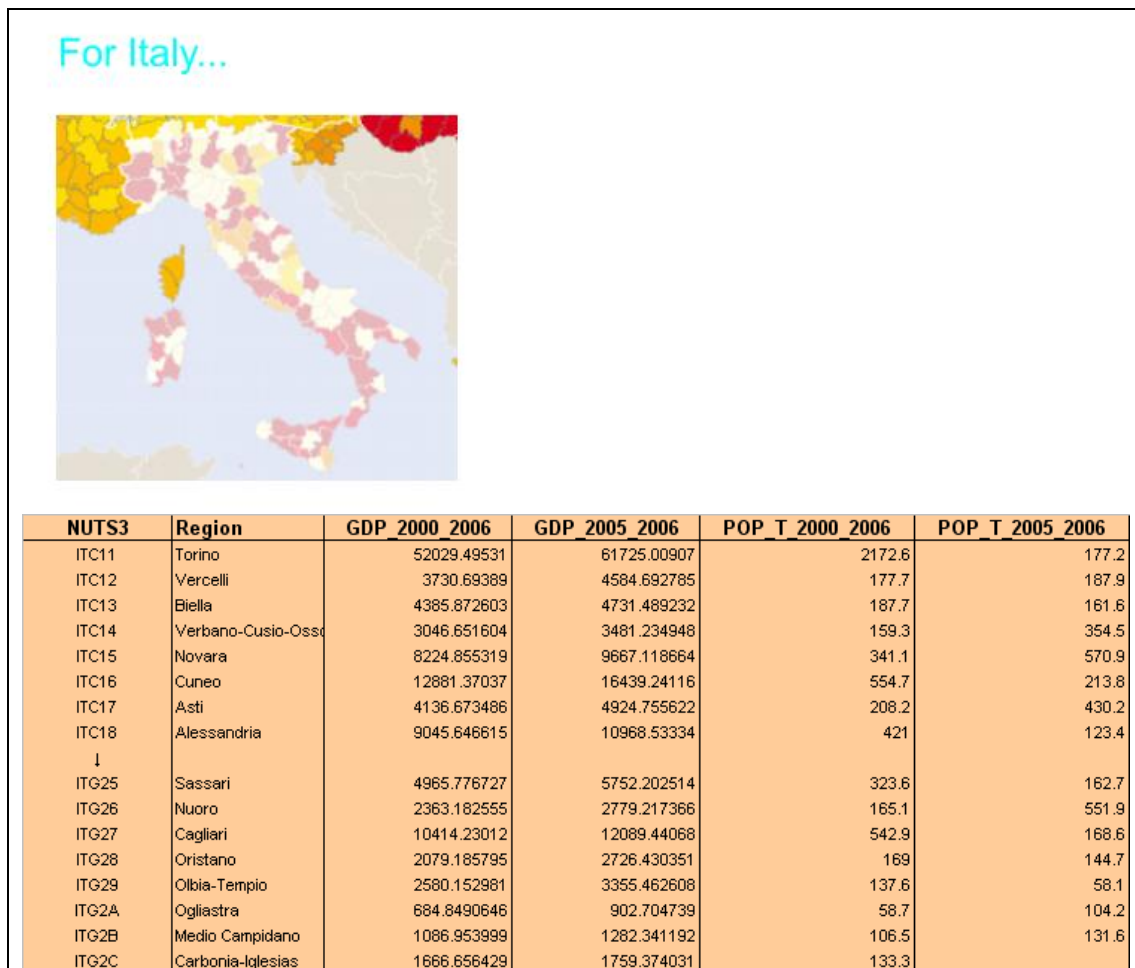


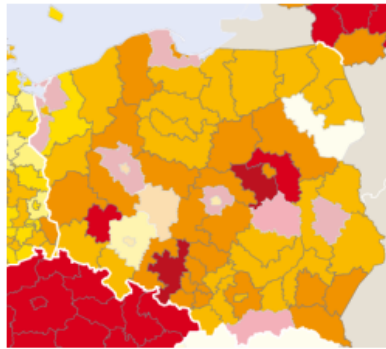
Figure 5: Problems of shifted data

### Problems with NUTS codes and names (11 input errors)

**Input error-type 8:** For some regions of Poland (11 entries), the total population in 2000 (POP\_T\_2000\_2006) has been estimated by the total population in 2003 using NUTS2003 divisions (i.e. POP\_T\_2003\_2003). Such estimations are fine provided the NUTS3 codes are used to relate the regions and not the region names. In this case, the region names have been erroneously used (see Fig. 6). Here the region names have not changed but the geometries for the regions have (and thus the sensible use of different NUTS codes for such instances). The resultant (erroneous) population values are all possible, and in this case, may not be easily identified. They may be identified by a subtraction of POP\_T\_2005\_2006 from POP\_T\_2000\_2006 provided the subtraction happens to result in large outlying differences.



For Poland...



Code_v2003	Code_v2006	Name_v2003	Name_v2006	POP2003_v2003	POP2006_v2006
pl111	pl114	Łódzki	Łódzki	940,7	379,7
pl124	pl128	Radomski	Radomski	736	607,9
pl212	pl215	Nowosadecki	Nowosadecki	1099,1	742,4
pl313	pl314	Lubelski	Lubelski	1216,5	720,9
pl342	pl344	Lomzynski	Lomzynski	311,4	420,5
pl413	pl416	Kaliski	Kaliski	800,9	720,74
pl412	pl418	Poznanski	Poznanski	1140	603,39
pl421	pl425	Szczedinski	Szczedinski	1103,1	314,9
pl513	pl518	Wlodawski	Wlodawski	433,8	537,1
pl520	pl522	Opolski	Opolski	1058,3	649,1
pl632	pl634	Gdanski	Gdanski	952,7	466,7

**Figure 6:** Problems with NUTS codes and names.

## 3 Worked examples: commented R scripts and results

### 3.1 The R statistical environment

All worked examples are coded in the R statistical computing environment (Ihaka and Gentleman 1996), which is open source<sup>8</sup>. In particular we use version 2.9.0 of the base system. For each worked example, only contributed packages are used (i.e. can be downloaded from the R website) except for a useful R mapping package, GISTools (Brunsdon pers. comm.), which is currently under development and will be made available on the R website shortly. The (unsupported) version of GISTools used here (version 0.5-4) is posted on ESPON 2013 database extranet, together with all other relevant materials that are needed to repeat each worked example. The GISTools package is not essential for outlier detection and maps could be constructed using other R packages or outside of R in a GIS.

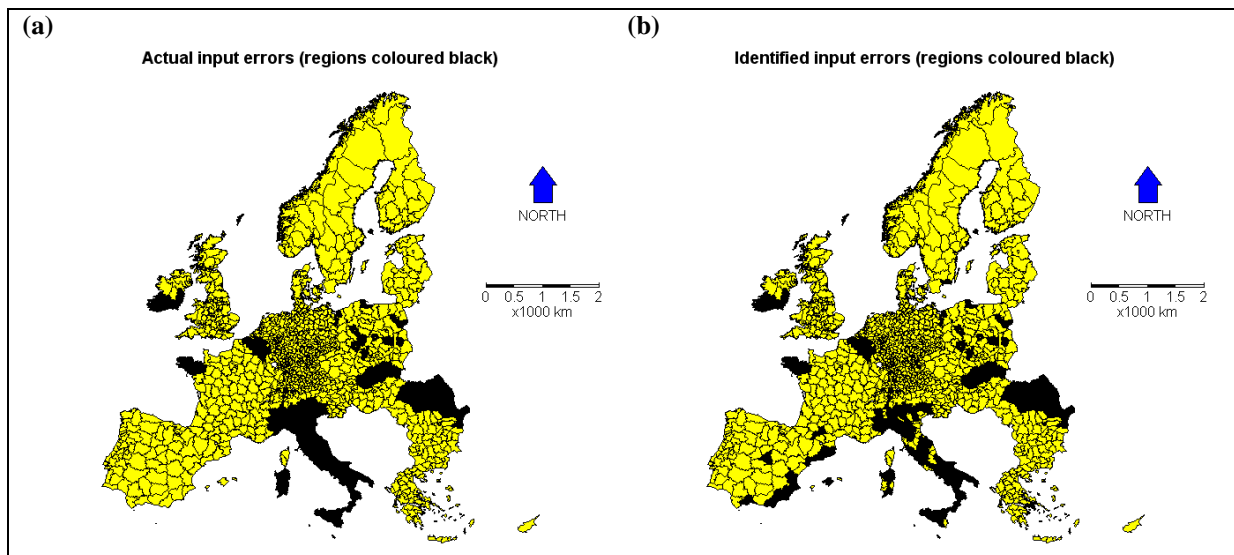
### 3.2 Worked example 1: univariate & residual analyses for input errors & outliers

The R script for worked example 1 is given in Appendix 1. The results are summarised in Fig. 7, where the rates of false negatives, false positives and overall misclassification with respect to input error identification were found to be 13.2%, 2.0% and 3.7% respectively. These rates are promisingly low, where their existence, is in part, a reflection of the automated nature of the identification procedures undertaken. Rates should tend to zero upon further (manual) scrutiny of the input errors that only have the *potential* to be so. For example, it would be expected that the rate of false negatives would reduce upon a manual scrutiny of the data in Italy, where the shift in data values (input error-type 7) should be quickly identified.

Observe also that there are many instances of false positives in Spain. This may reflect (unknown) input errors that were already present in this data set (i.e. before our deliberate introduction of input errors) or may reflect true (but unusual) data values (i.e. outliers). Either way, the corresponding data should be scrutinised and checked.

---

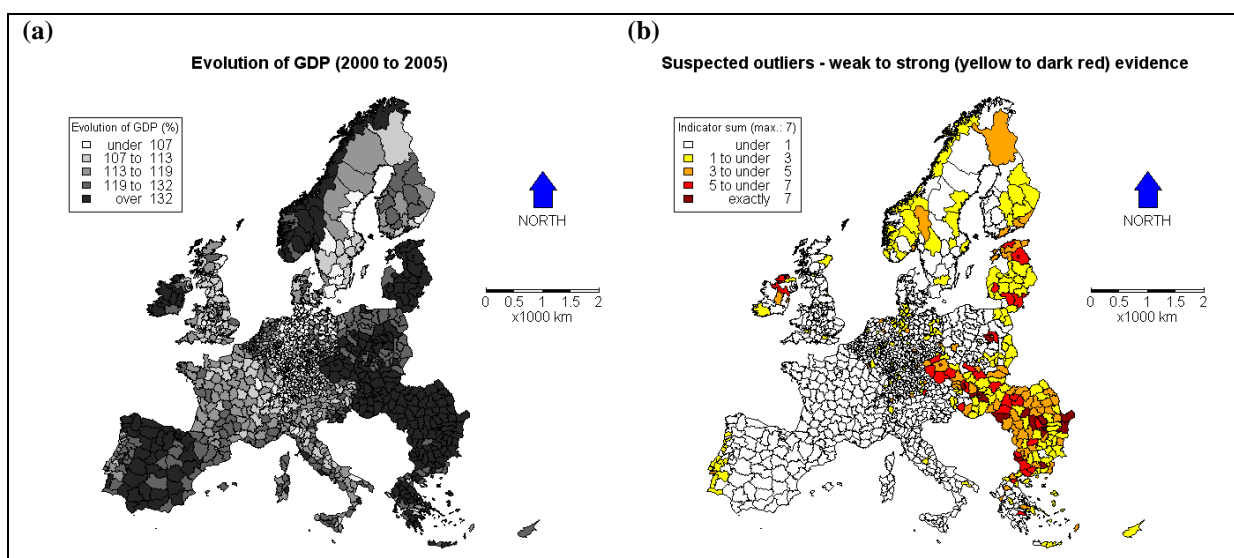
<sup>8</sup> The R website is <http://www.r-project.org/>



**Figure 7:** Location of (a) true (deliberate) input errors versus (b) identified input errors. Rates of false negatives, false positives and overall misclassification are 13.2%, 2.0% and 3.7% respectively

### 3.3 Worked example 2: univariate & residual analyses for outliers

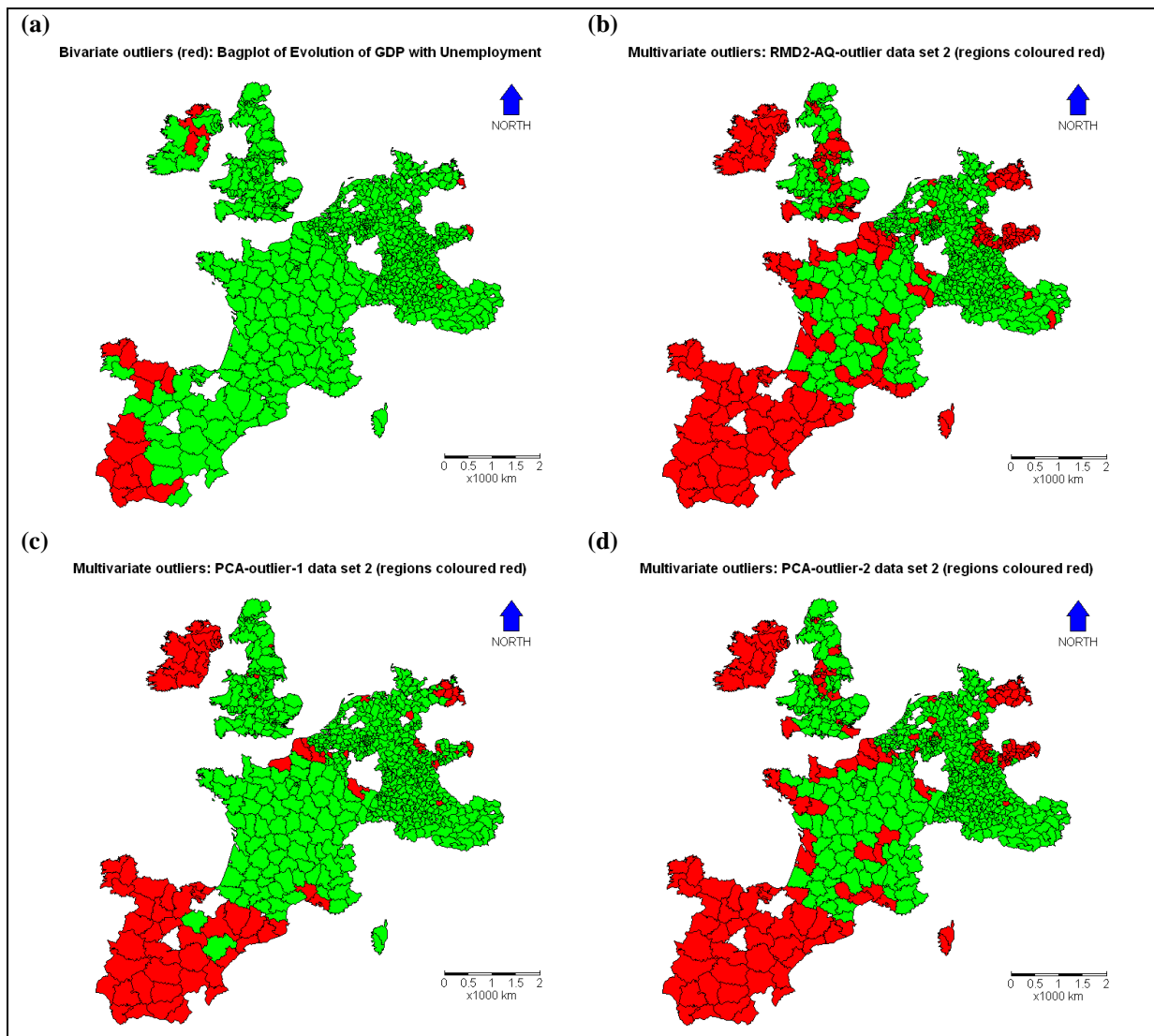
The R script for worked example 2 is given in [Appendix 2](#). The results are summarised in [Fig. 8](#), where the spatial distribution of `EVOGDP_2000_2005_2006` is compared with the spatial distribution of (suspected) outliers for this variable. In total, seven indicators were used to gauge whether or not an observation is outlying: (1) standard boxplot statistics; (2) adjusted boxplot statistics; (3) Hawkins' test for spatial outliers; and (4 to 7) large (absolute and raw) residuals from LM/MLR/LR/GWR fits (each calibrated with the coordinate data). These indicators are summarised in [Fig. 8b](#), where a strong case for an outlier relates to an observation that has positive results for all seven outlier identification analyses. It appears that the most outlying `EVOGDP_2000_2005_2006` observations lie in the south-east of the ESPON region.



**Figure 8:** Spatial distribution of (a) `EVOGDP_2000_2005_2006` and (b) suspected outliers for `EVOGDP_2000_2005_2006` (seven univariate indicators)

### 3.4 Worked example 3: multivariate analyses for outliers

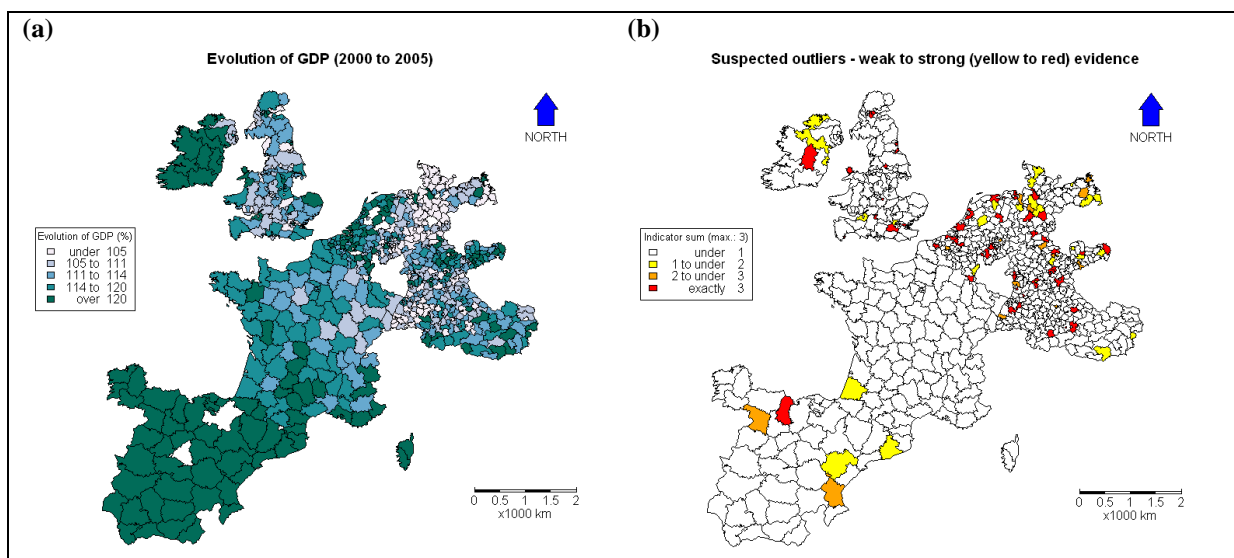
The R script for worked example 3 is given in [Appendix 3](#). The results are summarised in [Fig. 9](#), where only a much reduced data set of 731 regions could be used in this set of analyses (a consequence of a considerable amount of missing data). In [Fig. 9](#), potential outliers are found according: (a) a bagplot of EVOGDP\_2000\_2005\_2006 with UNEMP\_R\_2001\_1999; (b) the technique, RMD2-AQ-outlier; (c) the technique, PCA-outlier-1; and (d) the technique, PCA-outlier-2. The bagplot identifies outliers in a bivariate-sense whilst the other three techniques identify outliers in a multivariate-sense (in this case, with respect to outlying or unusual relationships amongst EVOGDP\_2000\_2005\_2006, UNEMP\_R\_2001\_1999 NAT\_HAZ\_2004\_1999 and SF\_CF\_1999\_1999). Clearly, the bivariate approach is missing key information. The three multivariate approaches give broadly similar results, but where the relatively large number of potential outliers (as many as 201 observations using RMD2-AQ-outlier) suggests multiple (statistical) populations rather than one population with many outlying observations. This would require further investigation.



**Figure 9:** Spatial distribution of suspected (bivariate and multivariate) outliers according to (a) a bagplot of EVOGDP\_2000\_2005\_2006 with UNEMP\_R\_2001\_1999; (b) the RMD2-AQ-outlier technique; (c) the PCA-outlier-1 technique; and (d) the PCA-outlier-2 technique

### 3.5 Worked example 4: multivariate residual analyses for outliers

The R script for worked example 4 is given in [Appendix 4](#). The results are summarised in [Fig. 10](#), where the spatial distribution of EVOGDP\_2000\_2005\_2006 is compared with the spatial distribution of (suspected) outliers for this variable. Again, only a much reduced data set of 731 regions could be used for this multivariate analysis. In total, three indicators were used to gauge whether or not an observation is outlying: (1) large (absolute and raw) residuals from an MLR fit; (2) large (absolute and raw) residuals from an LR fit; and (3) large (absolute and raw) residuals from a GWR fit. All three models were calibrated using the coordinates, SF\_CF\_1999\_1999 and SPAT\_TYPE\_2\_1999\_1999 as independent contextual data. These indicators are summarised in [Fig. 10b](#), where a strong case for an outlier relates to an observation that has positive results for all three outlier identification analyses.

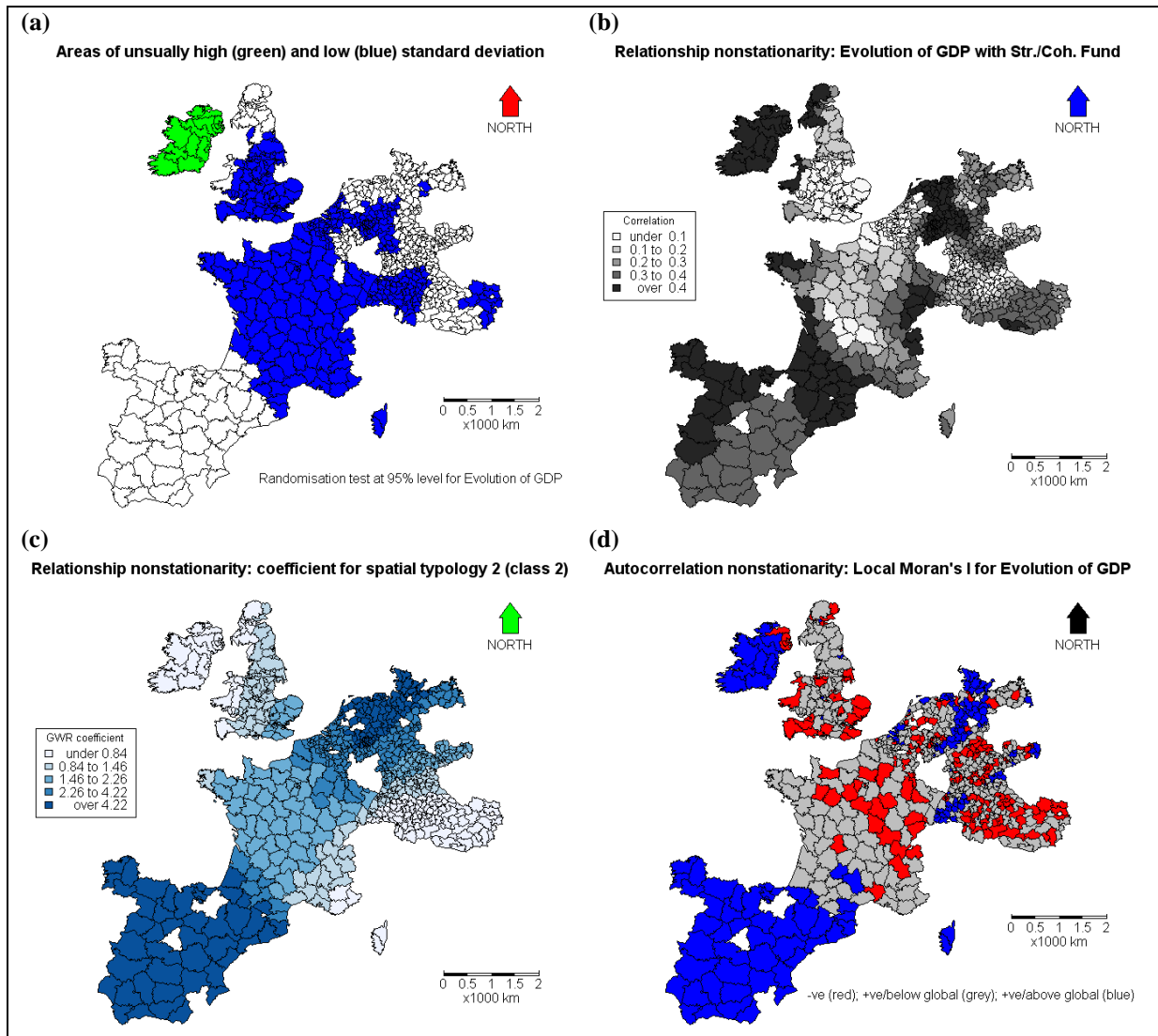


**Figure 10:** Spatial distribution of (a) EVOGDP\_2000\_2005\_2006 and (b) suspected outliers for EVOGDP\_2000\_2005\_2006 (three multivariate indicators).

### 3.6 Worked example 5: identification of spatial clusters

The R script for worked example 5 is given in [Appendix 5](#). The results are summarised in [Fig. 11](#), where the aim is to identify 'unusual' clusters in EVOGDP\_2000\_2005\_2006 with respect to (a) its local variability (using GW standard deviations); (b) its local relationship to SF\_CF\_1999\_1999 (via a GW correlation analysis); (c) its local relationship to class 2 of SPAT\_TYPE\_2\_1999\_1999 (via a GWR analysis); and (d) its local spatial autocorrelation (via local Moran's I statistic). Again, only a much reduced data set of 731 regions could be used for this combined

univariate and multivariate analysis. Observe that the shown local relationships for EVOGDP\_2000\_2005\_2006 are examples, as different local relationship can be investigated.



**Figure 11:** Identification of 'unusual' clusters in EVOGDP\_2000\_2005\_2006 with respect to (a) its local variability (using GW standard deviations); (b) its local relationship to SF\_CF\_1999\_1999 (via a GW correlation analysis); (c) its local relationship to class 2 of SPAT\_TYPE\_2\_1999\_1999 (via a GWR analysis); and (d) its local spatial autocorrelation (via local Moran's I statistic)

Briefly and focusing on EVOGDP\_2000\_2005\_2006 for Ireland and Northern Ireland only; Fig. 11a indicates that these regions tend to have unusually high levels of variation in EVOGDP\_2000\_2005\_2006; Fig. 11b suggests that these regions tend to have an unusually strong relationship between EVOGDP\_2000\_2005\_2006 and SF\_CF\_1999\_1999; Fig. 11c suggests that these regions tend to have an unusually weak relationship between EVOGDP\_2000\_2005\_2006 and class 2 of SPAT\_TYPE\_2\_1999\_1999; and Fig. 11d suggests that some regions of Northern Ireland can have an unusual negative spatial autocorrelation for EVOGDP\_2000\_2005\_2006 (i.e. neighbouring values of EVOGDP\_2000\_2005\_2006 tend to be dissimilar).

### 3.7 **Worked example 6: some consequences of MAUP**

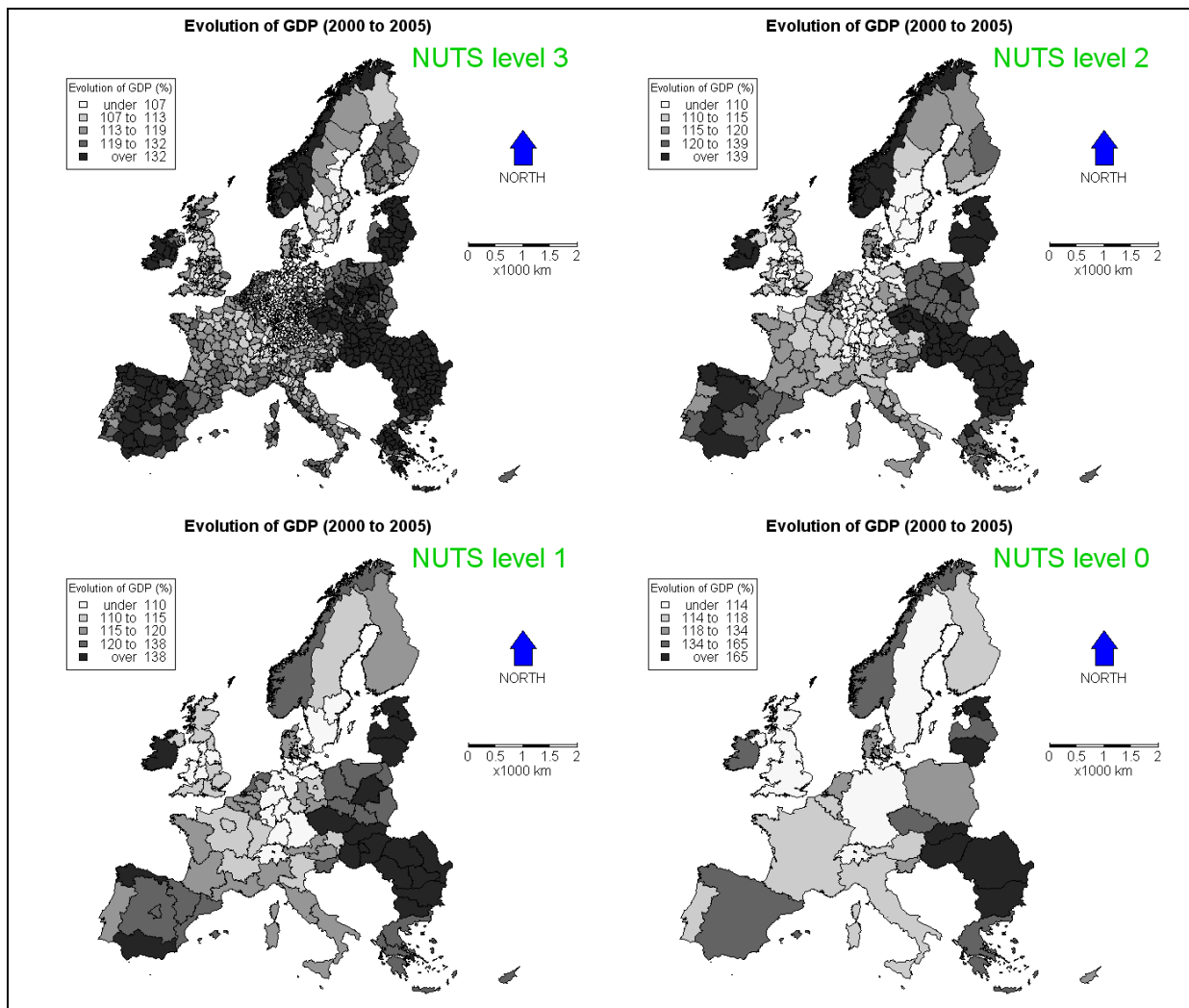
The R script for worked example 6 is given in [Appendix 6](#). The results are summarised in [Figs. 12 and 13](#), where the spatial distribution of EVOGDP\_2000\_2005\_2006 and the spatial distribution of (suspected) outliers for EVOGDP\_2000\_2005\_2006 are shown at four different NUTS levels (i.e. 3, 2, 1 and 0), respectively. At each NUTS level, the same seven indicators are used to gauge whether or not an observation is outlying (as in worked example 2).

Scatterplots and correlations are given in [Fig. 14](#), where the "Strongest indication of an outlier for any constituent NUTS level 3 region" is related to the "Indication of an outlier in a corresponding aggregated NUTS level 2/1/0 region". If the effects of MAUP on outlier identification are minimal, then a strong relationship (and correlation) would be expected.

From [Figs. 12 to 14](#), we can observe that:

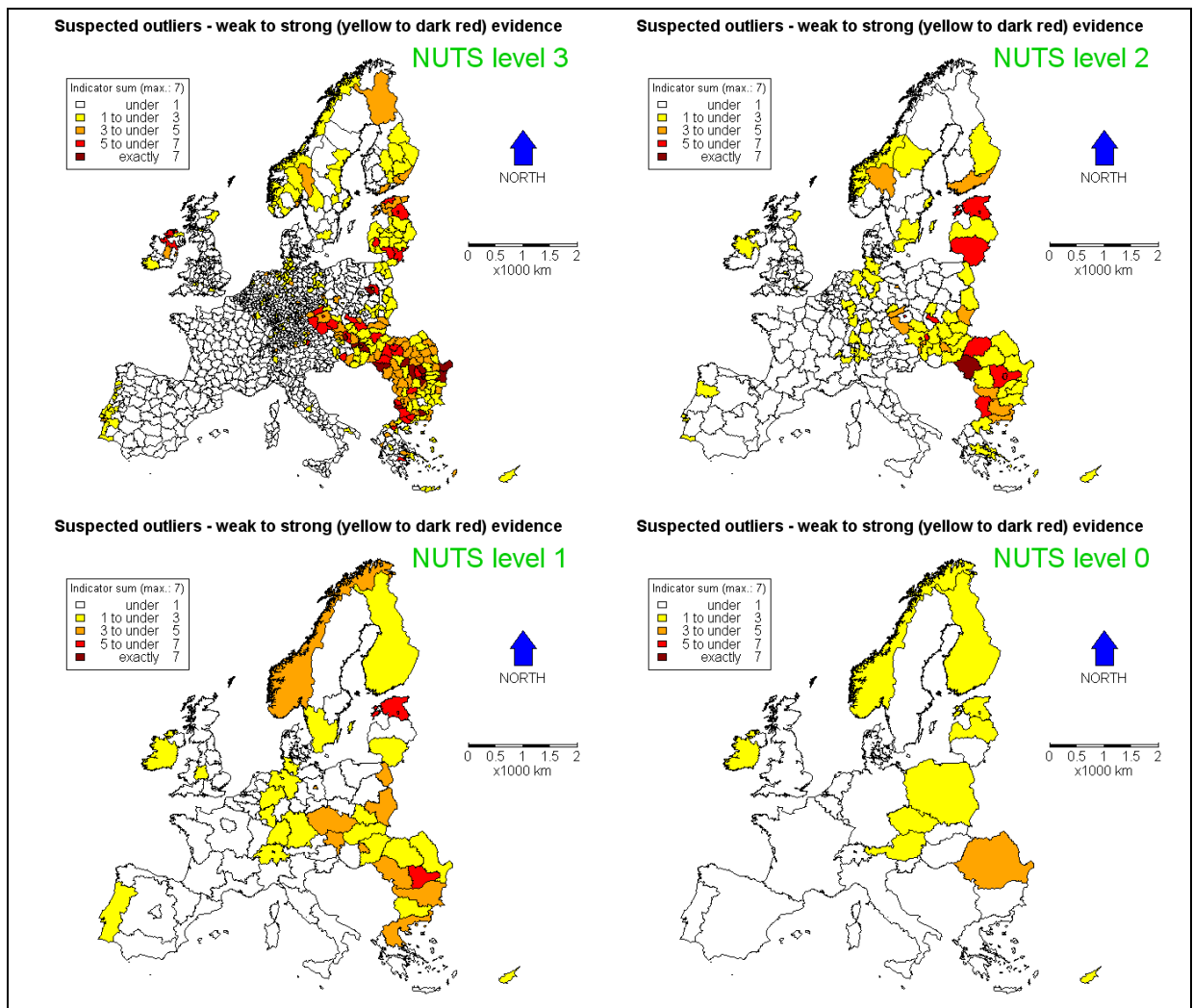
- A NUTS level 3 region that is an outlier does not imply that the NUTS level 2/1/0 region that contains it will also be an outlier.
- Several adjacent NUTS level 3 regions that are outliers, which belong to two or more adjacent NUTS level 2 regions, do not imply that those NUTS level 2 regions will be outliers (and so forth down the NUTS levels).
- A NUTS level 0/1/2 region that is an outlier is likely to contain one or more NUTS level 3 regions that are outliers.
- Evidence of an outlier weakens as the level of spatial aggregation increases.

In summary, it is recommended that outliers should be identified at the smallest spatial scale.

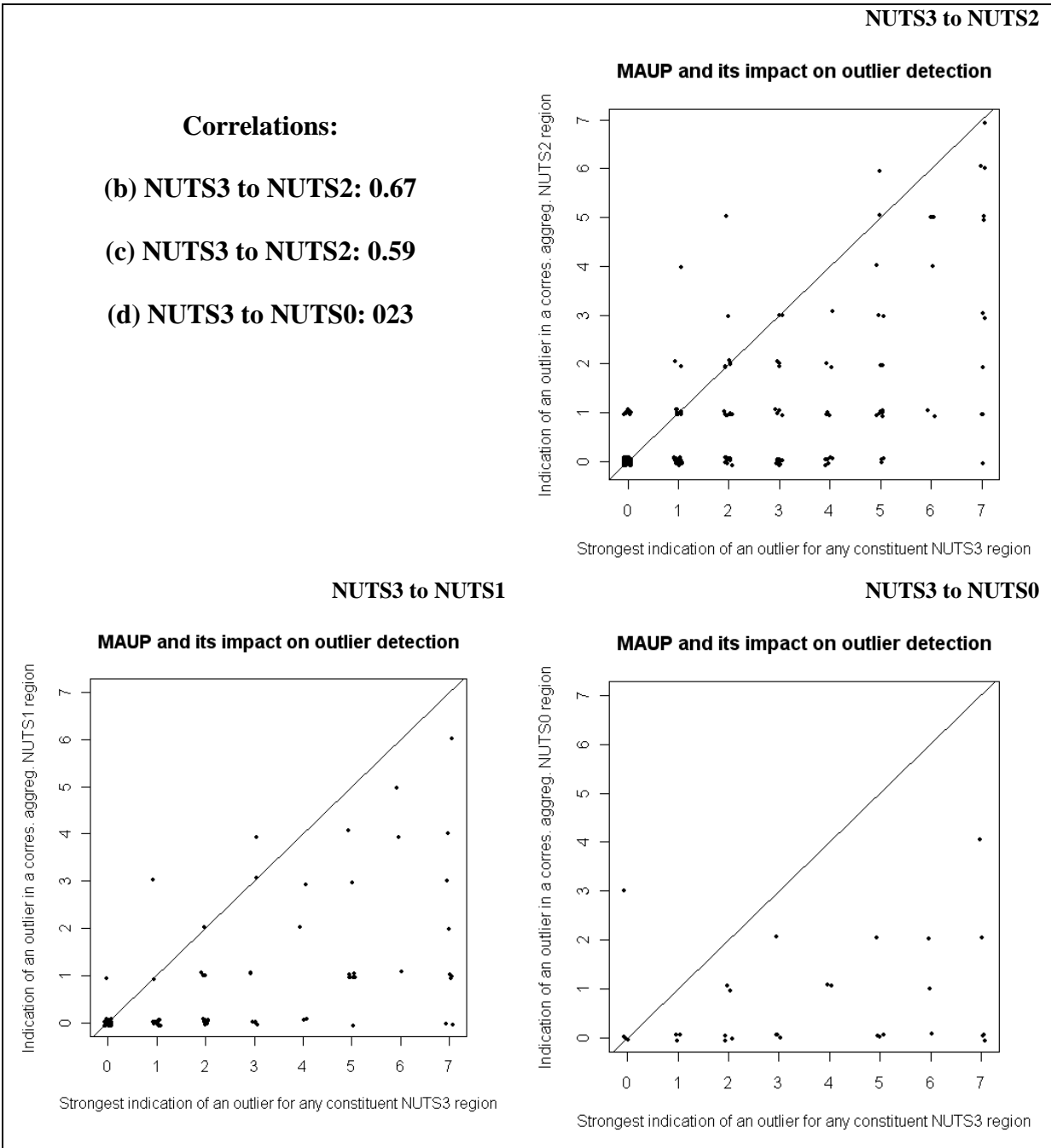


**Figure 12:** Spatial distribution of EVOGDP\_2000\_2005\_2006 at four different NUTS levels (3, 2, 1 and 0)





**Figure 13:** Spatial distribution of suspected outliers for EVOGDP\_2000\_2005\_2006 (via seven univariate indicators), where outliers are identified at four NUTS levels (3, 2, 1 and 0)



**Figure 14:** Scatterplots and correlations, where the "Strongest indication of an outlier for any constituent NUTS level 3 region" is related to the "Indication of an outlier in a corresponding aggregated NUTS level 2/1/0 region". Scatterplots are jittered to aid interpretation

## 4 Discussion and further developments

This technical report provides an introduction to the detection of logical input errors and statistical outliers (i.e. exceptional values) for data sets of the ESPON 2013 Database. Some important aspatial and spatial techniques have been introduced and demonstrated within the R statistical computing environment.

The field of robust statistics and outlier detection is extremely large and diverse, and as such can not be comprehensively reviewed within the terms of reference of this report. However, outlier detection techniques applicable (or designed for) *spatial* data sets are not as developed as those for *aspatial* applications.

In this respect, our current research is focused on this specific area of model development. Here robust versions of geographically weighted summary statistics (GWSS), geographically weighted regression (GWR) and geographically weighted principal component analysis (GWPCA) are to the fore, as they allow the detection of outliers in both univariate and multivariate spatial data sets.

Our expected deliveries for the final report of this phase of the ESPON project will be firmly based on the analytical techniques described and applied here. However we will now hone these procedures using a concrete, real-life data set rather than the fabricated data set used here. This new data set will no doubt present some new analytical challenges that have not been considered. This should enhance the detection methodology, which may need to include the addition of further techniques.

For the final report, we also aim to introduce a selection of the robust geographically weighted techniques that we are currently working on. An improved version of Hawkins' spatial outlier test is also under development, as is a robust version of the local Moran's I statistic (with respect to outlier identification). Here it is envisaged that our relatively advanced robust spatial methods should not be fully presented in the final report of this first phase of the ESPON project, but instead left for the next phase of the ESPON project (i.e. for the 2011 to 2013 stage), when the development of these robust spatial methods has properly matured. Work in this next phase should also include the packaging of the R code for these robust spatial methods, so that techniques are fully portable, transferable and openly documented.

### Acknowledgements

Thanks are due to Ronan Ysebaert and Claude Grasland at UMS RIATE for their work on providing us with the (univariate) test data set and the deliberate introduction of a range of challenging logical input errors to this data set.

## References

- Ainsworth LM, Dean CB (2008), *Detection of local and global outliers in mapping studies*. *Environmetrics* 19, 21-37.
- Anselin L. (1995) *Local indicators of spatial association*. *Geographical Analysis* 27, 93-115.
- Béguin C, Hulliger B (2004) *Multivariate outlier detection in incomplete survey data: the epidemic algorithm and transformed rank correlations*. *Journal of the Royal Statistical Society, Series A* 167(2), 275-294.
- Brunsdon C, Fotheringham AS, Charlton ME (2002) *Geographically weighted summary statistics - a framework for localised exploratory data analysis*. *Computers, Environment and Urban Systems* 26, 501-524.
- Brunsdon C, Charlton ME (2010) *An assessment of the effectiveness of multiple hypothesis testing for geographical anomaly detection*. Submitted to *Environment and Planning B*
- Chambers R, Hentges A, Zhao X (2004) *Robust automatic methods for outlier and error detection*. *Journal of the Royal Statistical Society, Series A* 167(2), 323-339.
- Charlton ME, Brunsdon C, Demšar U, Harris P, Fotheringham AS (2010) *Principal component analysis: from global to local*. In preparation.
- Charlton S (2004) *Evaluating automatic edit and imputation methods, and the EUREDIT Project*. *Journal of the Royal Statistical Society, Series A* 167(2), 199-207.
- Cruz Ortiz M, Sarabia LA, Herrero A (2006) *Robust regression techniques: A useful alternative for the detection of outlier data in chemical analysis*. *Talanta* 70, 499-512.
- D'Alimonte D, Cornford D (2007) *Outlier detection with partial information: application to emergency mapping*. *Stochastic Environmental Research and Risk Assessment* 22, 613-620.
- Daszykowski M, Kaczmarek K, Vander Heyden Y, Walczak B (2007) *Robust statistics in data analysis – a review Basic concepts*. *Chemometrics and Intelligent Laboratory Systems* 85, 203-219.
- ESPON (2006) 3.4.3 *The modifiable areas unit problem – Final Report* [http://www.espon.eu/mmp/online/website/content/projects/261/431/file\\_4970/](http://www.espon.eu/mmp/online/website/content/projects/261/431/file_4970/)
- Faraway J (2004) *Linear models with R*. Chapman & Hall/CRC, Boca Raton/FL
- Filzmoser P, Garrett R, Reimann C (2005) *Multivariate outlier detection in exploration geochemistry*. *Computers & Geosciences* 31, 579-587.
- Filzmoser P, Maronna R, Werner M (2008) *Outlier identification in high dimensions*. *Computational Statistics and Data Analysis* 52, 1694-1711.

- Fotheringham AS, Brunson C, Charlton ME (2002) *Geographically Weighted Regression - the analysis of spatially varying relationships*. Wiley, Chichester.
- Frigge M, Hoaglin DC, Iglewicz B (1989) *Some implementations of the Boxplot*. The American Statistician 43, 50–54.
- Ghosh-Dastidar B, Schafer JL (2003) *Multiple edit/multiple imputation for multivariate continuous data*. Journal of the American Statistical Association 98(464), 807-817.
- Harris P, Brunson C (2010) *Exploring spatial variation and spatial relationships in a freshwater acidification critical load data set for Great Britain using geographically weighted summary statistics*. Computers & Geosciences 36, 54-70.
- Harris P, Fotheringham AS, Juggins S (2010) *Robust Geographically Weighed Regression: A Technique for Quantifying Spatial Relationships Between Freshwater Acidification Critical Loads and Catchment Attributes*. To appear in the Annals of the Association of American Geographers.
- Hawkins RM (1980) *Identification of Outliers*. Chapman & Hall, London.
- Hoo KA, Tvarlapati KJ, Piovoso MJ, Hajare R (2002) *A method of robust multivariate outlier replacement*. Computers and Chemical Engineering 26, 17-39.
- Hubert M, Vandervieren E (2008) *An adjusted boxplot for skewed distributions*. Computational Statistics and Data Analysis 52, 5186-5201.
- Ihaka R, Gentleman R (1996) *R: A language for data analysis and graphics*. Journal of Computational and Graphical Statistics 5, 299-314.
- Jackson DA, Chen Y (2004) *Robust principal component analysis and outlier detection with ecological data*. Environmetrics 15, 129-139.
- Kou Y, Lu C-T, Chen D (2006) *Spatial Weighted Outlier Detection*. In proceedings of the 2006 SIAM International Conference on Data Mining No. 614 2006.
- Liu H, Jezek K, O'Kelly M (2001) *Detecting outliers in irregularly distributed spatial data sets by locally adaptive and robust statistical analysis and GIS*. International Journal of Geographical Information Science 15(8), 721-741.
- Loader C (2004) *Smoothing: Local Regression Techniques*. In Gentle J, Härdle W, Mori Y (eds) Handbook of Computational Statistics. Springer-Verlag, Heidelberg.
- Locantore N, Marron J, Simpson D, Tripoli N, Zhang J, Cohen K (1999) *Robust principal components for functional data*. Test 8, 1–73.
- Meklit T, Van Meirvenne M, Verstraete S, Bonroy J, Tack F (2009) *Combining marginal and spatial outliers identification to optimize the mapping of the regional geochemical baseline concentration of soil heavy metals*. Geoderma 148, 413-420.
- Morgenthaler S (2007) *A survey of robust statistics*. Statistical Methods & Applications 15, 271-293.
- Petrakos G, Conversano C, Farmakis G, Mola F, Siciliano R, Stavropoulos P (2004) *New ways of specifying data edits*. Journal of the Royal Statistical Society, Series A 167(2), 249-274.

Plaia A, Bondi A (2006) *Single imputation method of missing values in environmental pollution data sets*. Atmospheric Environment 40, 7316-7330.

Reimann C, Filzmoser P, Garrett R (2005) *Background and threshold: critical comparison of methods of determination*. Science of the Total Environment 346, 1-16.

Rousseeuw PJ, Ruts I, Tukey JW (1999) *The Bagplot: A Bivariate Boxplot*. The American Statistician 53, 382-387.

Rousseeuw PJ, Debruyne M, Engelen S, Hubert M (2006) *Robust and outlier detection in chemometrics*. Critical Reviews in Analytical Chemistry 36, 221-242.

Vanden Branden K, Verboven S (2009) *Robust data imputation*. Computational Biology and Chemistry 33, 7-13.

Wong D (1996) *Aggregation effects in geo-referenced data*. In Arlinghaus SL (ed) Practical Handbook of Spatial Statistics. CRC Press, Boca Raton, FL.

# Appendices

## Appendix 1 – R script for worked example 1

```
# 1. Preamble #####

# Worked example 1 - for technical report - challenge 10 - ESPON 2013 database
# NCG - P. Harris & M. Charlton
# 7/2/10

# Objective - to identify input errors in:
# "NUTS_2006" (the NUTS3 code)
# "GDP_2000_2006"
# "GDP_2005_2006"
# "POP_T_2000_2006"
# "POP_T_2005_2006"

# Methods: univariate - aspatial
# Mixture of logical & statistical methods
# Statistical methods:
# 1. Standard boxplots only

# R packages needed.....
# 1. GISTools (version 0.5-4) - depends on 2 to 11...
# 2. foreign (version 0.8-30)
# 3. gpclib (version 1.4-3)
# 4. maptools (version 0.7-16)
# 5. Matrix (version 0.999375-18)
# 6. RColorBrewer (version 1.0-2)
# 7. sp (version 0.9-28)
# 8. spam (version 0.15-2)
# 9. spdep (version 0.4-29)
# 10. spgwr (version 0.6-2)
# 11. tripack (version 1.2-11)

# Base R system version 2.9.0
# N.B. Some of the above packages may still depend on other R packages - download these from R website...

# Relevant data files (see data & ArcGIS directories):

# Excel files...
# 1. ESPON_DATA_NCG_CHALLENGE_10_original.xls
# 2. ESPON_DATA_NCG_CHALLENGE_10_subsets.xls
```

```

# Text files...
# 3. Worked example 1 true codes & new ID.txt

# ArcGIS files...
# 4. Worked_example_1a.shp - ArcGIS shapefile of the data...

# Only files 3 and 4 are needed in this worked example...

# The variables - some with deliberate input-errors...

# The following 5 variables are all suspected (i.e. in this case, known) to have input errors...
# "NUTS3_2006_E",
# "GDP_2000_2006_E",
# "GDP_2005_2006_E",
# "POP_T_2000_2006_E",
# "POP_T_2005_2006_E"

# These 3 variables are calculated from above so will be effected by an input error...
# "GDP_POP_2000_2006_E",
# "GDP_POP_2005_2006_E",
# "EVOGDP_2000_2005_2006_E"

# Remaining variables - all known to have no input errors...
# "NUTS3","NUTS23","NUTS2","NUTS1","NUTS0" - different NUTS levels
# "Error_type" - type of input error according to technical report (a number between 1 and 8)
# "New_ID" - relates to the regions name only & is purely numeric
# "Region_2006_E" - name of region (NB this does not have any errors)
# "NUTS3" - repeated (a consequence of an ArcGIS operation)
# "X","Y" - centroids of regions

# NOTE that this example dataset has been reduced to
# 1329 values from an original 1351 values (i.e. 22 values removed).
# See readme in excel files on worked example data.

# Note here that 13 of the 22 values removed relate to regions that are highly
# spatially disjoint from mainland Europe (i.e. parts of Portugal, France, and Spain
# such as the Azores, Canaries etc.). Before inclusion into the analyses, we need
# to decide on an appropriate distance metric for these regions.

# 2. Importing data as a ArcGIS shapefile & using GISTools to do some maps... ##

require(GISTools)
#help(GISTools)
# Ignore all warnings - this code is under development...

# Read in the shapefile...
data1 <- readShapePoly("Worked_example_1a.shp",
proj4string=CRS("+proj=Lambert_Azimuthal_Equal_Area+datum=D_ETRS_1989+ellps=GCS_ETRS_1989"))
colnames(data1@data)

# Renaming each variable - as they have been truncated in ArcGIS...

```



```

colnames(data1@data) <- c("NUTS3","NUTS23","NUTS2","NUTS1","NUTS0",
"Error_type","New_ID","NUTS3_2006_E","Region_2006_E",
"GDP_2000_2006_E","GDP_2005_2006_E","POP_T_2000_2006_E","POP_T_2005_2006_E",
"GDP_POP_2000_2006_E","GDP_POP_2005_2006_E","EVOGDP_2000_2005_2006_E","NUTS3","X","Y")

# Size of data set and adding an order ID...
n <- length(data1@data[,1])
Order_ID <- seq(1,n)
data1@data <- cbind(data1@data, Order_ID)
attach(data1@data)

# Creating a shading scheme and plotting a choropleth map...
shades.1 = auto.shading(GDP_2000_2006_E,5, cols=brewer.pal(5,'Greens'))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,GDP_2000_2006_E,shades.1)
title("GDP_2000_2006_E: with input errors")
choro.legend(1300000,400000,shades.1,fmt="%4.0f",title='GDP',cex=0.8)
map.scale(1800000,-950000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1000000,80000, col="blue")

# Creating a shading scheme and plotting a choropleth map...
shades.2 = auto.shading(GDP_2005_2006_E,5, cols=brewer.pal(5,'Greens'))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,GDP_2005_2006_E,shades.2)
title("GDP_2005_2006_E: with input errors")
choro.legend(1200000,400000,shades.2,fmt="%4.0f",title='GDP',cex=0.8)
map.scale(1800000,-950000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1000000,80000, col="blue")

# Creating a shading scheme and plotting a choropleth map...
shades.3 = auto.shading(POP_T_2000_2006_E,5, cols=brewer.pal(5,'Greens'))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,POP_T_2000_2006_E,shades.3)
title("POP_T_2000_2006_E: with input errors")
choro.legend(1400000,400000,shades.3,fmt="%4.0f",title='POP.',cex=0.8)
map.scale(1800000,-950000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1000000,80000, col="blue")

# Creating a shading scheme and plotting a choropleth map...
shades.4 = auto.shading(POP_T_2005_2006_E,5, cols=brewer.pal(5,'Greens'))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,POP_T_2005_2006_E,shades.4)
title("POP_T_2005_2006_E: with input errors")
choro.legend(1400000,400000,shades.4,fmt="%4.0f",title='POP.',cex=0.8)
map.scale(1800000,-950000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1000000,80000, col="blue")

# 3. Input error-types 1 and 2 - wrong NUTS code #####

# This is one approach to deal with these input error-types...

# Order our data by the New_ID (ie a numeric ID of the NUTS region name)...

```

```

data2 <- data1 @data[order(New_ID),]
attach(data2)

# Read in a data set where NUTS codes and names (again given as new_ID) are known to be correct...
data3 <- read.table("Worked example 1 true codes & new ID.txt", header=T)
colnames(data3)
attach(data3)

# Scan for input errors in the NUTS code...
# i.e. relate the "New_ID and NUTS3_2006" variables in datasets, data2 and data3...
data4 <- cbind(data2[,7],data2[,8],data3)
#fix(data4) # data spreadsheet

# Or better still - automatically identify input errors as follows...
x <- match(data4[,2], data4[,4]) # matches the New_ID values and assigns matches by position in data set
y <- seq(1,n) # sequence of numbers from 1 to the size of data set (same as Order_ID)
z <- y-x # should be a data set of zeros if all NUTS codes are inputted correctly
sort(-1*(abs(z))) # in this case 29 NUTS codes are inputted incorrectly...

# Updating input error information in one file - using our ordered data set...
indicator.1 <- ifelse(z==0, 0, 1)
data1.update.1 <- cbind(data2, indicator.1)
data1.update.1 <- as.data.frame(data1.update.1)
attach(data1.update.1)
#fix(data1.update.1)

# Re-order our data back to its original state...
data1 @data <- data1.update.1[order(data1.update.1[,20]),] # note using data1.update.1[,20] not Order_ID
attach(data1 @data)

# A choropleth map of input errors ...
shades.5 = shading(c(0,1,2),c("blue","white","red")) # this actually gives: white - no errors & red - errors
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1 @data[,21],shades.5) # use data1 @data[,21] not indicator.1
title("Input error-types 1 & 2 (regions coloured red)")
map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")

# Assessing the identification procedure
# Comparing "Error_types 1 & 2" with "indicator.1"
sum(indicator.1)
Assessment <- cbind(data1 @data[,6], data1 @data[,21])
Assessment.1 <- Assessment[order(-Assessment[,1]),]
#fix(Assessment.1)

# All 29 input errors correctly identified in Ireland & Romania...
# No false positives...

# COMMENT - THIS TYPE OF INPUT-ERROR IS PROBABLY BETTER DETECTED OUTSIDE OF R - i.e.
# IN A DATABASE

# 4. Input error-type 3 - impossible values #####
# This is one approach to deal with this input error-type...

```

```

# Checks for impossible values (in this case, impossible values for positive continuous data,
POP_T_2005_2006_E)

# POP_T_2005_2006_E in the ordered dataset
imp.val <- data2[,13]

# Explore the data...
summary(imp.val) # summary statistics
sort(imp.val) # ordered data
X11(width=5.3,height=5.7)
boxplot(imp.val, main="Input error-type 3", pch=19, cex=0.5) # boxplot

# Define minimum and maximums
Min_pop <- 0
Max_pop <- 10000 # This upper-limit is chosen by judgement

# Identifying & updating input error information in one file - using our ordered data set...
indicator.2 <- ifelse(Min_pop < imp.val & imp.val < Max_pop, 0, 1)
data1.update.2 <- cbind(data1.update.1,indicator.2)
data1.update.2 <- as.data.frame(data1.update.2)
attach(data1.update.2)
#fix(data1.update.2)

# Again re-order our data back to its original state...
data1@data <- data1.update.2[order(data1.update.2[,20]),]
attach(data1@data)

# A choropleth map of input errors ...
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,22],shades.5)
title("Input error-type 3 (regions coloured red)")
map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")

# Assessing the identification procedure
# Comparing "Error_type 3" with "indicator.2"
sum(indicator.2)
Assessment <- cbind(data1@data[,6], data1@data[,22])
Assessment.2 <- Assessment[order(-Assessment[,1]),]
#fix(Assessment.2)

# 1 input error correctly identified in Zurich...
# No false positives...

# 5. Input error-type 5 - potential missing value #####

# This is one approach to deal with this input error-type...

# Investigate all entries of -99, -999, -9999, 99, 999, 9999 as potential missing values...

# In this case do this for GDP_2005_2006_E

# GDP_2005_2006_E in the ordered dataset
miss.val <- data2[,11]

```

```

# Identifying & updating potential input error information in one file - using our ordered data set...
indicator.3 <- ifelse(miss.val!=abs(99) & miss.val!=abs(999) & miss.val!=abs(9999), 0, 1)
data1.update.3 <- cbind(data1.update.2,indicator.3)
data1.update.3 <- as.data.frame(data1.update.3)
attach(data1.update.3)
#fix(data1.update.3)

# Again re-order our data back to its original state...
data1@data <- data1.update.3[order(data1.update.3[,20]),]
attach(data1@data)

# A choropleth map of potential input errors...
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,23],shades.5)
title("Potential input error-type 5 (regions coloured red)")
map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")

# Assessing the identification procedure
# Comparing "Error_type 5" with "indicator.3"
sum(indicator.3)
Assessment <- cbind(data1@data[,6], data1@data[,23])
Assessment.3 <- Assessment[order(-Assessment[,1]),]
#fix(Assessment.3)

# 1 input error correctly identified in Malta...
# No false positives...

# 6. Input error-type 4,6,7 and 8 - all (relatively) unexpected values #####

# This is one approach to deal with these input error-types...

# Checking POP_T_2000_2006_E with POP_T_2005_2006_E for unusual data...
# This time four input error-types (4, 6, 7 and 8) can be investigated together...
# From section 3, impossible input error-types have already been identified for POP_T_2005_2006_E
# (i.e. we do not need to account for this error-type)
# but further input error-types can be identified if we relate/compare POP_T_2000_2006_E with
POP_T_2005_2006_E

# Intuitively, these data pairs should be broadly similar (but not exactly the same, i.e. error-type 6)
# Interest lies in the data pairs that are very different (i.e. differences are statistically outlying)
# or are identical (i.e. error-type 6 - copied or repeated data)...

# Again naming the relevant variables in the ordered dataset
x1 <- data2[,12] # POP_T_2000_2006_E
y1 <- data2[,13] # POP_T_2005_2006_E

# Exploring the data with a scatterplot (data should broadly lie on the 45 degree line)...
X11(width=5.3,height=5.7)
plot(x1,y1, main="Potential input error-types 4,6,7 or 8", pch=19, cex=0.5) # scatterplot
abline(0,1) # the 45 degree line

# Difference data...
# POP_T_2005_2006_E minus POP_T_2000_2006_E
z1 <- (y1-x1) # actual differences

```

```

#z1 <- abs(y1-x1) # absolute differences

# Exploring the difference data...
summary(z1) # summary statistics
sort(z1) # ordered data
X11(width=5.3,height=5.7)
hist(z1, main="Potential input error-types 4,6,7 or 8") # histogram
X11(width=5.3,height=5.7)
boxplot(z1, main="Potential input error-types 4,6,7 or 8", pch=19, cex=0.5) # boxplot

# Boxplot statistics...
# Change 'coef' accordingly...
# Default 'coef' is 1.5...
# The higher the 'coef' value the stricter the limits/cut-offs & vice versa...
bp <- boxplot.stats(z1, coef=6)
bp$stats
bp$stats[1] # the lower limit/cut-off - i.e. differences below are deemed outlying...
bp$stats[5] # the upper limit/cut-off - i.e. differences above are deemed outlying...
bp$conf
sort(bp$out)
length(bp$out) # number of potential outliers/errors....
# help(boxplot.stats) # for boxplot details...

# Identifying & updating potential input error information in one file - using our ordered data set...
indicator.4 <- ifelse(z1!=0 & z1>bp$stats[1]& z1<bp$stats[5], 0, 1) # i.e. identical or outlying differences...
data1.update.4 <- cbind(data1.update.3,indicator.4,z1) # note - including the difference data
data1.update.4 <- as.data.frame(data1.update.4)
attach(data1.update.4)
#fix(data1.update.4)

# Again re-order our data back to its original state...
data1@data <- data1.update.4[order(data1.update.4[,20]),]
attach(data1@data)

# A choropleth map of potential input errors...
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,24],shades.5)
title("Potential input error-types 4,6,7 or 8 (regions coloured red)")
map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")

# Observe that this methodology has also identified the impossible value for POP_T_2005_2006_E
# compare...
# indicator.2
# with
# indicators.4
# comparison.1 <- cbind(indicator.2, indicator.4) # see row 119

# We can now investigate these potential outliers more closely...

# Comparing "Error_types 4, 6, 7 & 8" with "indicator.4" and with the "difference data"
sum(indicator.4)
Assessment <- cbind(data1@data[,6], data1@data[,24], data1@data[,25])
Assessment.4 <- Assessment[order(-Assessment[,1]),]
#fix(Assessment.4)

# Results...

# Input error-type 3 - 1 out of 1 input error is correctly re-identified in Zurich...
# Input error-type 4 - 4 out of 4 input errors are correctly identified in Brittany...
# Input error-type 6 - 44 out of 44 input errors are correctly identified in Belgium...
# Input error-type 7 - 81 out of 107 input errors are correctly identified in Italy...

```

```

# i.e. 26 False negatives
# Input error-type 8 - 10 out of 11 input errors are correctly identified in Poland...
# i.e. 1 False negative

# False positives...
# For input error-types 4, 7 or 8 -
# 16 out of 1162
# i.e. 16 unusually large increases/decreases in population are actually true...

# False positives...
# For input error-type 6 -
# 6 out of 1162
# i.e. the population remained exactly the same in 6 regions...

# 7. Input error-type 6 only - repeated or copied data #####

# This is for GDP_2000_2006_E with GDP_2005_2006_E - but only for repeated data
# These data pairs should be exactly the same

# Again using the relevant variables in the ordered dataset
x2 <- data2[,10] #GDP_2000_2006_E
y2 <- data2[,11] #GDP_2005_2006_E

# Difference data...
z2 <- abs(y2-x2) # absolute differences
sort(z2) # ordered absolute data

# Identifying & updating potential input error information in one file - using our ordered data set...
indicator.5 <- ifelse(z2!=0, 0, 1) # i.e. identical differences...
data1.update.5 <- cbind(data1.update.4,indicator.5,z2) # note - including the difference data
data1.update.5 <- as.data.frame(data1.update.5)
attach(data1.update.5)
#fix(data1.update.5)

# Again re-order our data back to its original state...
data1 @data <- data1.update.5[order(data1.update.5[,20]),]
attach(data1 @data)

# A choropleth map of potential input errors...
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1 @data[,26],shades.5)
title("Potential input error-type 6 (regions coloured red)")
map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")

# Assessing the identification procedure
# Comparing "Error_type 6" with "indicator.5" and with the "difference data"
sum(indicator.5)
Assessment <- cbind(data1 @data[,6], data1 @data[,26], data1 @data[,27])
Assessment.5 <- Assessment[order(-Assessment[,1]),]
#fix(Assessment.5)

# 8 out of 8 input errors correctly identified in Slovakia...
# No false positives...

```

```

# 8. All input error-types together #####

# Put all indicator data together...
indicator.6 <- indicator.1+indicator.2+indicator.3+indicator.4+indicator.5

data1.update.6 <- cbind(data1.update.5,indicator.6)
data1.update.6 <- as.data.frame(data1.update.6)
attach(data1.update.6)
#fix(data1.update.6)

# Again re-order our data back to its original state...
data1@data <- data1.update.6[order(data1.update.6[,20]),]
attach(data1@data)

# A choropleth map of all identified input errors...
shades.6 = shading(c(0,1,3),c("blue","yellow","black")) # yellow - no errors & black - errors
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,28],shades.6)
title("Identified input errors (regions coloured black)")
map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")

# A choropleth map of actual input errors...
shades.6 = shading(c(0,1,9),c("blue","yellow","black")) # yellow - no errors & black - errors
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,6],shades.6)
title("Actual input errors (regions coloured black)")
map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")

# See Figure 7 in Technical report for above maps...

# Missclassification rates...

n # Data set size
tn.ie <- 205 # Total number of deliberate (known) input errors

tn.f <- 27 # Total number of false negatives
tn.p <- 22 # Total number of false positives

# Rate of false negatives
(tn.f/tn.ie)*100

# Rate of false positives
(tn.p/(n-tn.ie))*100

# Overall missclassification rate
((tn.f+tn.p)/n)*100

```

## Appendix 2 – R script for worked example 2

```
# 1. Preamble #####

# Worked example 2 - for technical report - challenge 10 - ESPON 2013 database
# NCG - P. Harris & M. Charlton
# 8/2/10

# Objective - to identify statistical outliers in:
# "EVOGDP_2000_2005_2006"

# Methods: univariate - aspatial & spatial
# Only statistical methods:
# 1. Standard and Adjusted boxplots,
# 2. Hawkins' test (includes the use of GWSS -
#     geographically weighted summary statistics - GW means and variances),
# 3. LM (local mean, i.e. a GW mean)
# 4. MLR (multiple linear regression),
# 5. LR (local regression) &
# 6. GWR (geographically weighted regression)

# R packages needed.....
# 1. GISTools (version 0.5-4) - depends on 2 to 11...
# 2. foreign (version 0.8-30)
# 3. gpclib (version 1.4-3)
# 4. maptools (version 0.7-16)
# 5. Matrix (version 0.999375-18)
# 6. RColorBrewer (version 1.0-2)
# 7. sp (version 0.9-28)
# 8. spam (version 0.15-2)
# 9. spdep (version 0.4-29)
# 10. spgwr (version 0.6-2) - for GWSS & GWR
# 11. tripack (version 1.2-11)
# 12. moments (version 0.11) - for skewness
# 13. robustbase (version 0.4-5) - for adjusted boxplots
# 14. locfit (version 1.5-4)- for LR

# Base R system version 2.9.0
# N.B. Some of the above packages may still depend on other R packages - download these from R website...

# Relevant data files (see data & ArcGIS directories):

# Excel files...
# 1. ESPON_DATA_NCG_CHALLENGE_10_original.xls
# 2. ESPON_DATA_NCG_CHALLENGE_10_subsets.xls

# ArcGIS files...
# 3. Worked_example_2a.shp - ArcGIS shapefile of the data...
```



```

# The 11 variables...

# "NUTS3","NUTS23","NUTS2","NUTS1","NUTS0" - 5 different NUTS levels
# "New_ID" - relates to the regions name only & is purely numeric
# "NUTS3_2006" - the 2006 NUTS3 version
# "Region_2006" - name of 2006 NUTS3 version
# "X","Y" - centroids of regions
# "EVOGDP_2000_2005_2006" - the variable of interest

# NOTE - this example dataset has been reduced to 1329 values
# from an original 1351 values (i.e. 22 values removed)
# see readme in excel files on worked example data.

# NOTE - this dataset is NOT one corrected for input errors from worked example 1.
# It is just the corresponding dataset without the introduction of deliberate input errors.

# 2. Importing data as a ArcGIS shapefile & using GISTools to do a map... #####

require(GISTools)
#help(GISTools)
# Ignore all warnings - this code is under development...

# Read in the shapefile...
data1 <- readShapePoly("Worked_example_2a.shp",
proj4string=CRS("+proj=Lambert_Azimuthal_Equal_Area+datum=D_ETRS_1989+ellps=GCS_ETRS_1989"))
colnames(data1@data)

# renaming each variable - as they have been truncated in ArcGIS...
colnames(data1@data) <- c("NUTS3","NUTS23","NUTS2","NUTS1","NUTS0",
"New_ID","NUTS3_2006","Region_2006",
"X","Y","EVOGDP_2000_2005_2006")

# Size of data set and adding an order ID...
n <- length(data1@data[,1])
Order_ID <- seq(1,n)
data1@data <- cbind(data1@data, Order_ID)
attach(data1@data)

# Creating a shading scheme and plotting a choropleth map of EVOGDP_2000_2005_2006...
shades.1 = auto.shading(EVOGDP_2000_2005_2006,5, cols=brewer.pal(5,'Greys'))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,EVOGDP_2000_2005_2006,shades.1)
title("Evolution of GDP (2000 to 2005)")
choro.legend(-2400000,2200000,shades.1,fmt="%4.0f",title='Evolution of GDP (%)',cex=0.8)
map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")
#text(1500000,2400000, "NUTS level 3", cex=2, col=3)

```

```

# 3. Boxplots #####

# Let EVOGDP_2000_2005_2006 be z1...
z1 <- EVOGDP_2000_2005_2006

# Exploring this data...
summary(z1) # summary statistics
sort(z1) # ordered data

# Histogram
X11(width=5.3,height=5.7)
hist(z1, main="Histogram: Evolution of GDP (2000 to 2005)",xlab="Evolution of GDP")

# Standard boxplot with defaults
X11(width=5.3,height=5.7)
boxplot(z1, main="Std. boxplot: Evolution of GDP (2000 to 2005)", pch=19, cex=0.5)

# Standard Boxplot statistics...
# Change 'coef' accordingly...
# Default 'coef' is 1.5...
# The higher the 'coef' value the stricter the limits/cut-offs & vice versa...
bp <- boxplot.stats(z1, coef=1.5)
bp$stats
bp$stats[1] # the lower limit/cut-off - i.e. values below are deemed outlying...
bp$stats[5] # the upper limit/cut-off - i.e. values above are deemed outlying...
bp$conf
sort(bp$out)
length(bp$out) # number of potential outliers...
# help(boxplot.stats) # for details...

# Identifying & updating outlier information in one file
indicator.1 <- ifelse(z1>bp$stats[1]& z1<bp$stats[5], 0, 1) # i.e. suspected outliers...
data1 @data <- cbind(data1 @data, indicator.1)
data1 @data <- as.data.frame(data1 @data)
attach(data1 @data)
data.1 <- data1 @data
#fix(data.1)

# A choropleth map of standard boxplot outliers
shades.2 = shading(c(0,1,2),c("blue","white","red")) # i.e. white - no & red - yes
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1 @data[,13],shades.2)
title("Std. boxplot outliers (regions coloured red)")
map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")

# Need moments package to assess skewness (before adjusted boxplots)
require(moments)
# Ignore warning message...
skewness(z1) # skew is fairly high and positive

# Package for adjusted boxplots...
require(robustbase)

# Adjusted boxplot with defaults
X11(width=5.3,height=5.7)
adjbox(z1, main="Adj. boxplot: Evolution of GDP (2000 to 2005)", pch=19, cex=0.5)

```

```

# Adjusted Boxplot statistics...
# Change 'coef' accordingly...
# Default 'coef' is 1.5...
# The higher the 'coef' value the stricter the limits/cut-offs & vice versa...
abp <- adjboxStats(z1, coef=1.5)
abp$stats
abp$stats[1] # the lower limit/cut-off - i.e. values below are deemed outlying...
abp$stats[5] # the upper limit/cut-off - i.e. values above are deemed outlying...
abp$conf
sort(abp$out)
length(abp$out) # number of potential outliers...
#help(adjboxStats) # for details...

# Identifying & updating outlier information in one file
indicator.2 <- ifelse(z1>abp$stats[1]& z1<abp$stats[5], 0, 1) # i.e. suspected outliers...
data1 @data <- cbind(data1 @data, indicator.2)
data1 @data <- as.data.frame(data1 @data)
attach(data1 @data)
data.1 <- data1 @data
#fix(data.1)

# A choropleth map of adjusted boxplot outliers
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1 @data[,14],shades.2)
title("Adj. boxplot outliers (regions coloured red)")
map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")

# 4. GW summary statistics and Hawkins' Spatial Outlier Test #####

# First need to find GW means (i.e. LMs) and GW variances for Hawkin's test...
# In this case, using the gw.cov function in spgwr to find the GW means/variances...

# Note 1. - we could define our own weighting scheme to use with the gw.cov function.
# For example, an IDW-type scheme.
# But in this case, the default bi-square weighting scheme is used.

# Note 2. - we could find an optimal bandwidth (i.e. the optimal number of nearby data)
# for a GW mean using 'leave-one-out' cross-validation.
# But in this case, a user-specified bandwidth is defined as the nearest 10% of data.
# It is not so easy to find an optimal bandwidth for a GW variance
# and as such, is commonly chosen subjectively.

# Note 3. - Hawkins' test should ideally use GW means/variances that have been
# calculated without the observation at each calibration/observation location.
# However, this oversight is not expected to adversely affect results.

# Future work can investigate the above issues...

# To re-cap...
colnames(data.1)

# Defining coordinates....
coordinates(data.1) <- c("X", "Y")

```

```

# GW summary statistics at observation locations (i.e. region centroids)...
# Calculated using 10% of nearby EVOGDP_2000_2005_2006 data.
bwd.1 <- 0.1
gwss <- gw.cov(data.1, vars=11, adapt=bwd.1)
#help(gw.cov) # for details...
names(gwss$SDF) # The GW summary statistics calculated...

# GW means and variances...
GW.mean <- gwss$SDF$mean.V1
GW.variance <- (gwss$SDF$sd.V1)^2

# Hawkins' Test for Spatial Outliers...
Hawk.N <- bwd.1*length(X) # number of neighbouring data
Hawk.lm <- GW.mean # the local mean at observation points
Hawk.alv <- mean(GW.variance) # the average local variance with same bandwidth

Hawk.test <- (Hawk.N*(EVOGDP_2000_2005_2006-Hawk.lm)^2)/((Hawk.N+1)*Hawk.alv) # test statistic
summary(Hawk.test)

# Critical values of the chi-squared distribution
chi_10 <- 2.70554
chi_5 <- 3.84146
chi_2.5 <- 5.02389
chi_1 <- 6.63490
chi_0.5 <- 7.87944
chi_0.01 <- 10.828

# Updating outlier information in one file
indicator.3 <- ifelse(Hawk.test <= chi_5, 0, 1) # change critical level accordingly...
data1@data <- cbind(data1@data, Hawk.test, indicator.3)
data1@data <- as.data.frame(data1@data)
attach(data1@data)
data.1 <- data1@data
#fix(data.1)

# A choropleth map of spatial outliers ...
shades.3 = shading(c(chi_5,chi_1,chi_0.01),c("white","yellow","orange","red"))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,15],shades.3)
title("Spatial outliers: at 5/1/0.01 % (yellow/orange/red) critical levels")
map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")

# 5. Residual analysis with LM, MLR, LR and GWR models #####

# LM...
# Using GW.mean from before...
GW.mean.r <- EVOGDP_2000_2005_2006-GW.mean # Actual minus prediction
summary(GW.mean.r)

# Identifying & updating outlier information in one file
cut.off.1 <- quantile(GW.mean.r, probs = seq(0, 1, 0.05), na.rm=T) # for 5% tails - alter accordingly...

```

```

indicator.4 <-ifelse(GW.mean.r>=cut.off.1[2] & GW.mean.r<=cut.off.1[20], 0, 1)
data1@data <- cbind(data1@data, GW.mean.r, indicator.4)
data1@data <- as.data.frame(data1@data)
attach(data1@data)
data.1 <- data1@data
#fix(data.1)

# Raw residual map for LM...
shades.4 = shading(c(cut.off.1[2],cut.off.1[20]),c("red","white","black"))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,17],shades.4)
title("Raw resids. from LM: High/-ve(red) & High/+ve(black) (5% tails)")
map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")

# MLR...
# First- & second-order polynomial fits of the coordinate data...
mlr.1 <- lm(EVOGDP_2000_2005_2006 ~ X+Y)
mlr.2 <- lm(EVOGDP_2000_2005_2006 ~ X+Y+I(X^2)+I(Y^2)+I(X*Y))
summary(mlr.1)
summary(mlr.2)

# Choosing a second-order MLR fit...

# Using raw residuals as in LM fit...
raw.resids.mlr <- EVOGDP_2000_2005_2006-mlr.2$fitted # Actual minus prediction
summary(raw.resids.mlr)

# Identifying & updating outlier information in one file
cut.off.2 <- quantile(raw.resids.mlr, probs = seq(0, 1, 0.05), na.rm=T) # for 5% tails - alter accordingly...
indicator.5 <-ifelse(raw.resids.mlr>=cut.off.2[2] & raw.resids.mlr<=cut.off.2[20], 0, 1)
data1@data <- cbind(data1@data, raw.resids.mlr, indicator.5)
data1@data <- as.data.frame(data1@data)
attach(data1@data)
data.1 <- data1@data
#fix(data.1)

# Raw residual map for MLR...
shades.5 = shading(c(cut.off.2[2],cut.off.2[20]),c("red","white","black"))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,19],shades.5)
title("Raw resids. from MLR: High/-ve(red) & High/+ve(black) (5% tails)")
map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")

# LR...
# With coordinate data as explanatory variables (i.e. first-order polynomial).

# Using locfit...
require(locfit)
# Ignore warning message...

# Finding the bandwidth for a non-robust LR (i.e. not a lowess fit)
# using generalised cross-validation (GCV) approach.
summary(gcvplot(EVOGDP_2000_2005_2006~X+Y,data=data.1, scale=F,alpha=seq(0.005,0.01,by=0.001),
deg=1,kern="tricube",lfproc=locfit.raw))

```

```

# Choosing a LR fit with bandwidth chosen from above...
bwd.2 <- 0.008
lr <- locfit(EVOGDP_2000_2005_2006~X+Y,data=data.1, scale=F, alpha=bwd.2,
deg=1,kern="tricube",lfproc=locfit.raw)

# Raw residuals...
lr.p <- fitted.locfit(lr)
raw.resids.lr <- EVOGDP_2000_2005_2006-lr.p # Actual minus prediction
summary(raw.resids.lr)

# Identifying & updating outlier information in one file
cut.off.3 <- quantile(raw.resids.lr, probs = seq(0, 1, 0.05), na.rm=T) # for 5% tails - alter accordingly...
indicator.6 <- ifelse(raw.resids.lr>=cut.off.3[2] & raw.resids.lr<=cut.off.3[20], 0, 1)
data1@data <- cbind(data1@data, raw.resids.lr, indicator.6)
data1@data <- as.data.frame(data1@data)
attach(data1@data)
data.1 <- data1@data
#fix(data.1)

# Raw residual map for LR...
shades.6 = shading(c(cut.off.3[2],cut.off.3[20]),c("red","white","black"))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,21],shades.6)
title("Raw resids. from LR: High/-ve(red) & High/+ve(black) (5% tails)")
map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")

# GWR...
# With coordinate data as explanatory variables (i.e. first-order polynomial).
# Using spgwr...

# Defining the coordinates...
coords.1<-cbind(data.1[,9],data.1[,10])

# Finding the bandwidth for GWR using Akaike Information Criterion (AIC) approach.
#gwr.aic.bwd <-gwr.sel(EVOGDP_2000_2005_2006~X+Y,data=data.1,coords=coords.1,adapt=TRUE,
#gweight=gwr.bisquare, method="aic")
#gwr.aic.bwd[1] # the optimum bandwidth

# Or finding the bandwidth for GWR using cross-validation approach.
#gwr.cv.bwd <-gwr.sel(EVOGDP_2000_2005_2006~X+Y,data=data.1,coords=coords.1,adapt=TRUE,
#gweight=gwr.bisquare, method="cv")
#gwr.cv.bwd[1] # the optimum bandwidth

# Above optimisation can take a long time...
# So choosing a GWR fit with user-specified bandwidth of 0.03...
bwd.3 <- 0.03
gwr.p <-gwr(EVOGDP_2000_2005_2006~X+Y,data=data.1,coords=coords.1,adapt=bwd.3,
gweight=gwr.bisquare,predictions=T)
#gwr.p$$SDF

# GWR raw residuals...
raw.resids.gwr <- EVOGDP_2000_2005_2006-gwr.p$$SDF$pred
summary(raw.resids.gwr)

# Identifying & updating outlier information in one file
cut.off.4 <- quantile(raw.resids.gwr, probs = seq(0, 1, 0.05), na.rm=T) # for 5% tails - alter accordingly...
indicator.7 <- ifelse(raw.resids.gwr>=cut.off.4[2] & raw.resids.gwr<=cut.off.4[20], 0, 1)
data1@data <- cbind(data1@data, raw.resids.gwr, indicator.7)
data1@data <- as.data.frame(data1@data)

```

```

attach(data1 @data)
data.1 <- data1 @data
#fix(data.1)

# Raw residual map for GWR...
shades.7 = shading(c(cut.off.4[2],cut.off.4[20]),c("red","white","black"))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1 @data[,23],shades.7)
title("Raw resid. from GWR: High/-ve(red) & High/+ve(black) (5% tails)")
map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")

# NB - Future work should explore the identification of outliers using
# standardised residuals & corresponding robust regression models...

# 6. All identified outliers together #####

# Put all indicator data together...
indicator.8 <- indicator.1+indicator.2+indicator.3+indicator.4+indicator.5+indicator.6+indicator.7
summary(indicator.8)
# Histogram
X11(width=5.3,height=5.7)
hist(indicator.8,br=c(0,1,2,3,4,5,6,7))

# Thus a strong case for an outlier relates to an observation
# that has a indicator.8 value of 7...

data1 @data <- cbind(data1 @data, indicator.8)
data1 @data <- as.data.frame(data1 @data)
attach(data1 @data)
data.1 <- data1 @data
#fix(data.1)
#write.table(data.1,"Outliers_NUTS_level3.txt", col.names=T,row.names=F)

# A choropleth map of suspected outliers...
shades.7 = shading(c(1,3,5,7),c("white","yellow","orange","red","dark red"))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1 @data[,25],shades.7)
title("Suspected outliers - weak to strong (yellow to dark red) evidence")
choro.legend(-2400000,2200000,shades.7,
over="exactly", between="to under",
fmt="%4.0f",title="Indicator sum (max.: 7)",cex=0.8)
map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")
#text(1500000,2400000, "NUTS level 3", cex=2, col=3)

```

## Appendix 3 – R script for worked example 3

```
# 1. Preamble #####

# Worked example 3 - for technical report - challenge 10 - ESPON 2013 database
# NCG - P. Harris & M. Charlton
# 7/2/10

# Objective - to identify statistical outliers in some subset of this data:
# "X"
# "Y"
# "EVOGDP_2000_2005_2006"
# "SPAT_TYPE_1_1999_1999"
# "SPAT_TYPE_2_1999_1999"
# "UNEMP_R_2001_1999"
# "LU_AS_1_1996_1999"
# "LU_AS_2_1996_1999"
# "LU_AS_3_1996_1999"
# "LU_UF_1996_1999"
# "LU_AR_1996_1999"
# "LU_PC_1996_1999"
# "NAT_HAZ_2004_1999"
# "SF_CF_1999_1999"
# "SF_R_1999_1999"
# "SF_S_1999_1999"
# "SF_A_1999_1999"
# "CF_T_1999_1999"
# "CF_E_1999_1999"

# Methods: multivariate - aspatial only
# Only statistical methods:
# 1. Bagplots
# 2. Robust MD-squared analysis (RMD2-AQ-outlier)
# 3. Two techniques based on PCA for outlier detection (PCA-outlier-1 & PCA-outlier-2)

# R packages needed.....
# 1. GISTools (version 0.5-4) - depends on 2 to 11...
# 2. foreign (version 0.8-30)
# 3. gpclib (version 1.4-3)
# 4. maptools (version 0.7-16)
# 5. Matrix (version 0.999375-18)
# 6. RColorBrewer (version 1.0-2)
# 7. sp (version 0.9-28)
# 8. spam (version 0.15-2)
# 9. spdep (version 0.4-29)
# 10. spgwr (version 0.6-2)
# 11. tripack (version 1.2-11)
# 12. aplpack (version 1.2-1) - for bagplots
# 13. robustbase (version 0.4-5) - required for mvoutlier package
# 14. mvoutlier (version 1.4) - for robust MD-squared analysis and PCA outlier detection
```



```

# Base R system version 2.9.0
# N.B. Some of the above packages may still depend on other R packages - download these from R website...

# Relevant data files (see data & ArcGIS directories):

# Excel files...
# 1. ESPON_DATA_NCG_CHALLENGE_10_original.xls
# 2. ESPON_DATA_NCG_CHALLENGE_10_subsets.xls

# ArcGIS files...
# 3. Worked_example_345a_reduced.shp - ArcGIS shapefile of the data...

# The 27 variables...

# "NUTS3","NUTS23","NUTS2","NUTS1","NUTS0" - 5 different NUTS levels
# "New_ID" - relates to the regions name only & is purely numeric
# "NUTS3_2006" - the 2006 NUTS3 version
# "Region_2006" - name of 2006 NUTS3 version
# "X","Y" - centroids of regions
# "EVOGDP_2000_2005_2006" - Evolution of GDP
# and 16 likely contextual variables of "EVOGDP_2000_2005_2006" ...
# "SPAT_TYPE_1_1999_1999"
# "SPAT_TYPE_2_1999_1999"
# "UNEMP_R_2001_1999"
# "LU_AS_1_1996_1999"
# "LU_AS_2_1996_1999"
# "LU_AS_3_1996_1999"
# "LU_UF_1996_1999"
# "LU_AR_1996_1999"
# "LU_PC_1996_1999"
# "NAT_HAZ_2004_1999"
# "SF_CF_1999_1999"
# "SF_R_1999_1999"
# "SF_S_1999_1999"
# "SF_A_1999_1999"
# "CF_T_1999_1999"
# "CF_E_1999_1999"

# NOTE - Methods demonstrated in this worked example do not require
# a relationship between "EVOGDP_2000_2005_2006" and its likely
# contextual data - see worked examples 4 and 5 for this.

# NOTE - This example data set has been reduced to 731 values
# from an original 1351 values
# see readme in excel files on worked example data.

# 2. Importing data as a ArcGIS shapefile & using GISTools to do some maps #####

```

```

require(GISTools)
#help(GISTools)
# Ignore all warnings - this code is under development...

# Read in the shapefile...
data1 <- readShapePoly("Worked_example_345a_reduced.shp",
proj4string=CRS("+proj=Lambert_Azimuthal_Equal_Area+datum=D_ETRS_1989+ellps=GCS_ETRS_1989"))
colnames(data1@data)

# renaming each variable - as they have been truncated in ArcGIS...
colnames(data1@data) <- c("NUTS3","NUTS23","NUTS2","NUTS1","NUTS0",
"New_ID","NUTS3_2006","Region_2006",
"X","Y","EVOGDP_2000_2005_2006",
"SPAT_TYPE_1_1999_1999","SPAT_TYPE_2_1999_1999",
"UNEMP_R_2001_1999",
"LU_AS_1_1996_1999","LU_AS_2_1996_1999","LU_AS_3_1996_1999",
"LU_UF_1996_1999","LU_AR_1996_1999","LU_PC_1996_1999",
"NAT_HAZ_2004_1999",
"SF_CF_1999_1999",
"SF_R_1999_1999","SF_S_1999_1999","SF_A_1999_1999",
"CF_T_1999_1999","CF_E_1999_1999")

# Size of data set and adding an order ID...
n <- length(data1@data[,1])
Order_ID <- seq(1,n)
data1@data <- cbind(data1@data, Order_ID)
attach(data1@data)

# Coordinate data only...
coords <- cbind(data1@data[,9],data1@data[,10])

# Example multivariate data set one...
Mult.data.1 <- cbind(EVOGDP_2000_2005_2006,SPAT_TYPE_1_1999_1999,SPAT_TYPE_2_1999_1999,
UNEMP_R_2001_1999,LU_AS_1_1996_1999,LU_AS_2_1996_1999,LU_AS_3_1996_1999,LU_UF_1996_1999,
LU_AR_1996_1999,LU_PC_1996_1999,NAT_HAZ_2004_1999,SF_CF_1999_1999,SF_R_1999_1999,
SF_S_1999_1999,SF_A_1999_1999,CF_T_1999_1999,CF_E_1999_1999)
Mult.data.1 <- as.data.frame(Mult.data.1)
attach(Mult.data.1)

# Example multivariate data set two...
Mult.data.2 <- cbind(EVOGDP_2000_2005_2006,UNEMP_R_2001_1999,
NAT_HAZ_2004_1999,SF_CF_1999_1999)
Mult.data.2 <- as.data.frame(Mult.data.2)
attach(Mult.data.2)

# Example multivariate data set three (data set two with coordinates)...
Mult.data.3 <- cbind(X,Y,EVOGDP_2000_2005_2006,UNEMP_R_2001_1999,
NAT_HAZ_2004_1999,SF_CF_1999_1999)
Mult.data.3 <- as.data.frame(Mult.data.3)
attach(Mult.data.3)

# Creating a shading scheme and plotting a choropleth map of EVOGDP_2000_2005_2006...
shades.1 = auto.shading(EVOGDP_2000_2005_2006,5, cols=brewer.pal(5,'PuBuGn'))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,EVOGDP_2000_2005_2006,shades.1)
title("Evolution of GDP (2000 to 2005)")
choro.legend(-2300000,250000,shades.1,fmt="%4.0f",title='Evolution of GDP (%)',cex=0.8)
map.scale(100000,-1050000,500000,"x1000 km",4,0.5)
north.arrow(200000,750000,40000, col="blue")

# Creating a shading scheme and plotting a choropleth map of UNEMP_R_2001_1999...

```

```

shades.2 = auto.shading(UNEMP_R_2001_1999,5, cols=brewer.pal(5,'Greys'))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,UNEMP_R_2001_1999,shades.2)
title("Unemployment rate")
choro.legend(-2300000,250000,shades.2,fmt="%4.1f",title='Unemployment (%)',cex=0.8)
map.scale(100000,-1050000,500000,"x 1000 km",4,0.5)
north.arrow(200000,750000,40000, col="blue")

# Creating a shading scheme and plotting a choropleth map of NAT_HAZ_2004_1999...
shades.3 = auto.shading(NAT_HAZ_2004_1999,5, cols=brewer.pal(5,'Greens'))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,NAT_HAZ_2004_1999,shades.3)
title("Natural hazards")
choro.legend(-2300000,250000,shades.3,fmt="%4.0f",title='Indication',cex=0.8)
map.scale(100000,-1050000,500000,"x 1000 km",4,0.5)
north.arrow(200000,750000,40000, col="blue")

# Creating a shading scheme and plotting a choropleth map of SF_CF_1999_1999...
shades.4 = auto.shading(SF_CF_1999_1999,5, cols=brewer.pal(5,'Reds'))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,SF_CF_1999_1999,shades.4)
title("All Structural & Cohesion Fund expenditure")
choro.legend(-2350000,250000,shades.4,fmt="%4.0f",title='Str. & Coh. Fund',cex=0.8)
map.scale(100000,-1050000,500000,"x 1000 km",4,0.5)
north.arrow(200000,750000,40000, col="blue")

# 3. Bagplots #####

# Package for bagplots...
require(aplpack)

# For example, exploring these 2 variables...
z1 <- EVOGDP_2000_2005_2006
z2 <- UNEMP_R_2001_1999

# Summary statistics...
summary(z1)
summary(z2)

# Univariate boxplots...
X11(width=5.3,height=5.7)
boxplot(z1, main="Evolution of GDP", pch=19, cex=0.5)
X11(width=5.3,height=5.7)
boxplot(z2, main="Unemployment", pch=19, cex=0.5)

# The bagplot function
#help(bagplot)

# Example...
X11(width=5.3,height=5.7)
bagp.1 <- bagplot(z1,z2,xlab="Evolution of GDP",ylab="Unemployment",
main="Example bagplot: outliers in red (outside of bag)",cex=0.6)
bivariate.outliers.1 <- bagp.1$pxy.outlier

```

```

length(bivariate.outliers.1[,1])
bivariate.not.outliers.1 <- rbind(bagp.1$pxy.bag, bagp.1$pxy.outer)
length(bivariate.not.outliers.1[,1])

# Some data manipulations for mapping...
# Note can also use library(sqldf) to match datasets...
bivariate.outliers.1x <- merge(data1 @data, bivariate.outliers.1,
by.x=c("EVOGDP_2000_2005_2006", "UNEMP_R_2001_1999"), by.y=c("x", "y"))
indicator.1 <- c(rep(1, length(bivariate.outliers.1x[,1])))
bivariate.outliers.1x <- cbind(bivariate.outliers.1x, indicator.1)

bivariate.not.outliers.1x <- merge(data1 @data, bivariate.not.outliers.1,
by.x=c("EVOGDP_2000_2005_2006", "UNEMP_R_2001_1999"), by.y=c("x", "y"))
indicator.1 <- c(rep(0, length(bivariate.not.outliers.1x[,1])))
bivariate.not.outliers.1x <- cbind(bivariate.not.outliers.1x, indicator.1)

xx1 <- rbind(bivariate.not.outliers.1x, bivariate.outliers.1x)
data1 @data <- xx1[order(xx1[,28]),] # get data in correct order with Order_ID
attach(data1 @data)

# A choropleth map...
shades.5 = shading(c(0,1,2), c("white", "green", "red")) # i.e. green - no & red - yes
X11(width=8, height=7)
par(mar=c(0,0,2,0))
choropleth(data1, data1 @data[,29], shades.5)
title("Bivariate outliers (red): Bagplot of Evolution of GDP with Unemployment")
map.scale(100000, -1050000, 500000, "x 1000 km", 4, 0.5)
north.arrow(200000, 750000, 40000, col="blue")

# 4. Robust MD-squared analysis (RMD2-AQ-outlier) #####

# Following the paper of Filzmoser et al. (2005)...

# Load the necessary package...
require(mvoutlier)

# Note - multivariate data set one and similar data subsets can give rise to some technical problems
# with this technique, as it is designed for continuous multivariate normal data with outlying observations,
# whereas we have data sets that include categorical data.
# In this respect, we only explore example multivariate data sets two & three, which only have continuous
variables.
# This is not considered a problem as outliers are expected to be more hidden in continuous variables.
# Non-normality of continuous data may however still cause problems.
# Similar comments apply to the PCA methods of section 5...

# Therefore using example multivariate dataset two...

# The key function/plot for this identification technique is...
#help(aq.plot)
#X11(width=12, height=8)
#aq.plot(Mult.data.2)
#help(aq.plot)

# However, slightly adapting the aq.plot function to suit our needs...
aq.plot.1 <- function (x, delta = qchisq(0.975, df = ncol(x)), quan = 1/2,
alpha = 0.025)

```

```

{
  if (is.vector(x) == TRUE || ncol(x) == 1) {
    stop("x must be at least two-dimensional")
  }
  covr <- covMcd(x, alpha = quan)
  dist <- mahalanobis(x, center = covr$center, cov = covr$cov)
  s <- sort(dist, index = TRUE)
  z <- x
  if (ncol(x) > 2) {
    p <- princomp(x, covmat = covr)
    z <- p$scores[, 1:2]
    sdprop <- (p$sd[1] + p$sd[2])/sum(p$sd)
    cat("Projection to the first and second robust principal components.\n")
    cat("Proportion of total variation (explained variance): ")
    cat(sdprop)
    cat("\n")
  }
  par(mfrow = c(2, 2), mai = c(0.8, 0.6, 0.2, 0.2), mgp = c(2.4,
    1, 0))
  plot(z, col = 3, type = "n",
    main="(A) Data (by ID) projected on the first two RPCs",
    xlab = "First Robust Principal Component (RPC)", ylab = "Second Robust Principal Component (RPC)")
  text(z, dimnames(as.data.frame(z))[[1]], col = 3, cex = 0.8)
  plot(s$x, (1:length(dist))/length(dist), col = 3,
    main = paste("(B) Outlier detection: above ",
    100 * (1 - alpha), "% & adj. quantiles", sep = ""),
    xlab = "Ordered squared robust Mahalanobis distances",
    ylab = "Cumulative probability", type = "n")
  text(s$x, (1:length(dist))/length(dist), as.character(s$ix),
    col = 3, cex = 0.8)
  t <- seq(0, max(dist), by = 0.01)
  lines(t, pchisq(t, df = ncol(x)), col = 6)
  abline(v = delta, col = 5)
  xarw <- arw(x, covr$center, covr$cov, alpha = alpha)
  # note - arw() is the adaptive reweighted estimator for multivariate location and scatter...
  abline(v = xarw$cn, col = 4)
  legend(11000, 0.3, c("Chi-squared dist. func.", paste(100 * (1 - alpha), "% quantile", sep = ""),
    "Adjusted quantile"), col = c(6,5,4), lty = c(1,1,1), bty="n")
  plot(z, col = 3, type = "n", main = paste("(C) Outliers (in red) based on (user-specified) ",
    100 * (1 - alpha), "% quantile", sep = ""),
    xlab = "First RPC", ylab = "Second RPC")
  for (i in 1:nrow(x)) {
    if (dist[i] >= delta)
      text(z[i, 1], z[i, 2], dimnames(as.data.frame(x))[[1]][i],
        col = 2, cex = 0.8)
    if (dist[i] < delta)
      text(z[i, 1], z[i, 2], dimnames(as.data.frame(x))[[1]][i],
        col = 3, cex = 0.8)
  }
  plot(z, col = 3, type = "n", main = "(D) Outliers (in red) based on adjusted quantile",
    xlab = "First RPC", ylab = "Second RPC")
  for (i in 1:nrow(x)) {
    if (dist[i] >= xarw$cn)
      text(z[i, 1], z[i, 2], dimnames(as.data.frame(x))[[1]][i],
        col = 2, cex = 0.8)
    if (dist[i] < xarw$cn)
      text(z[i, 1], z[i, 2], dimnames(as.data.frame(x))[[1]][i],
        col = 3, cex = 0.8)
  }
  o <- (sqrt(dist) > min(sqrt(xarw$cn), sqrt(qchisq(0.975,
    dim(x)[2]))))
  l <- list(outliers = o)
  l
}

```

```

}

# Thus our take on the adjusted quantile plot...
X11(width=12,height=8)
mult.out.d2.m1 <- aq.plot.1(Mult.data.2)

# Identifying & updating outlier information in one file
indicator.2 <-ifelse(mult.out.d2.m1$outliers==F, 0, 1) # i.e. suspected outliers...
data1 @data <- cbind(data1 @data, indicator.2)
data1 @data <- as.data.frame(data1 @data)
attach(data1 @data)
data.1 <- data1 @data
#fix(data.1)

# A choropleth map of the multivariate outliers...
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1 @data[,30],shades.5)
title("Multivariate outliers: RMD2-AQ-outlier data set 2 (regions coloured red)")
map.scale(100000,-1050000,500000,"x 1000 km",4,0.5)
north.arrow(200000,750000,40000, col="blue")

# Univariate presentation of the same multivariate outliers...
# (i.e multivariate outliers in red as in the above choropleth map)
X11(width=12,height=8)
uni.plot(Mult.data.2)

# NB - see Filzmoser 2005 paper & mvoutlier reference manual for more options
# on the visualisation of multivariate outliers...

# And using example multivariate dataset three...

# The adjusted quantile plot...
X11(width=12,height=8)
mult.out.d3.m1 <- aq.plot.1(Mult.data.3)

# Identifying & updating outlier information in one file
indicator.3 <-ifelse(mult.out.d3.m1$outliers==F, 0, 1) # i.e. suspected outliers...
data1 @data <- cbind(data1 @data, indicator.3)
data1 @data <- as.data.frame(data1 @data)
attach(data1 @data)
data.1 <- data1 @data
#fix(data.1)

# A choropleth map of the multivariate outliers...
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1 @data[,31],shades.5)
title("Multivariate outliers: RMD2-AQ-outlier data set 3 (regions coloured red)")
map.scale(100000,-1050000,500000,"x 1000 km",4,0.5)
north.arrow(200000,750000,40000, col="blue")

# Univariate presentation of the same multivariate outliers...
# (i.e multivariate outliers in red as in the above choropleth map)
X11(width=16,height=8)
uni.plot(Mult.data.3)

```

```

# 5. PCA for outlier detection #####

# Following the paper of Filzmoser et al. (2008)...

# Again using the mvoutlier package
# And using only Mult.data.2 data set for simplicity...

# Sign Method for Outlier Identification in High Dimensions...
# i.e. PCA-outlier-1
# Simple version (sign1) & sophisticated (sign2) versions are possible...
# Using the simple version...
mult.out.d2.m2 <- sign1(Mult.data.2)

# Identifying & updating outlier information in one file
indicator.4 <- ifelse(mult.out.d2.m2$wfinal01==1, 0, 1) # i.e. suspected outliers are the wrong way around in this
case...
data1@data <- cbind(data1@data, indicator.4)
data1@data <- as.data.frame(data1@data)
attach(data1@data)
data.1 <- data1@data
#fix(data.1)

# A choropleth map of the multivariate outliers
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,32],shades.5)
title("Multivariate outliers: PCA-outlier-1 data set 2 (regions coloured red)")
map.scale(100000,-1050000,500000,"x 1000 km",4,0.5)
north.arrow(200000,750000,40000, col="blue")

# PCOut Method for Outlier Identification in High Dimensions
# i.e. PCA-outlier-2
mult.out.d2.m3 <- pcout(Mult.data.2)

# Identifying & updating outlier information in one file
indicator.5 <- ifelse(mult.out.d2.m3$wfinal01==1, 0, 1) # i.e. suspected outliers are the wrong way around in this
case...
data1@data <- cbind(data1@data, indicator.5)
data1@data <- as.data.frame(data1@data)
attach(data1@data)
data.1 <- data1@data
#fix(data.1)

# A choropleth map of the multivariate outliers
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,33],shades.5)
title("Multivariate outliers: PCA-outlier-2 data set 2 (regions coloured red)")
map.scale(100000,-1050000,500000,"x 1000 km",4,0.5)
north.arrow(200000,750000,40000, col="blue")

```

```

# 6. Summary #####

# Number of potential outliers

# Bagplot...
sort(-data1@data[,29])

#RMD2-AQ-outlier with multivariate data set 2
sort(-indicator.2)

#PCA-outlier-1 with multivariate data set 2
sort(-indicator.4)

#PCA-outlier-2 with multivariate data set 2
sort(-indicator.5)

```

## Appendix 4 – R script for worked example 4

```

# 1. Preamble #####

# Worked example 4 - for technical report - challenge 10 - ESPON 2013 database
# NCG - P. Harris & M. Charlton
# 6/2/10

# Objective - to identify statistical outliers in "EVOGDP_2000_2005_2006"
# in relation to some subset of the following explanatory/contextual data:
# "X"
# "Y"
# "SPAT_TYPE_1_1999_1999"
# "SPAT_TYPE_2_1999_1999"
# "UNEMP_R_2001_1999"
# "LU_AS_1_1996_1999"
# "LU_AS_2_1996_1999"
# "LU_AS_3_1996_1999"
# "LU_UF_1996_1999"
# "LU_AR_1996_1999"
# "LU_PC_1996_1999"
# "NAT_HAZ_2004_1999"
# "SF_CF_1999_1999"
# "SF_R_1999_1999"
# "SF_S_1999_1999"
# "SF_A_1999_1999"
# "CF_T_1999_1999"
# "CF_E_1999_1999"

# Methods: multivariate - aspatial & spatial

```



```

# Only statistical methods:
# 1. MLR (multiple linear regression),
# 2. LR (local regression) &
# 3. GWR (geographically weighted regression)

# R packages needed.....
# 1. GISTools (version 0.5-4) - depends on 2 to 11...
# 2. foreign (version 0.8-30)
# 3. gpclib (version 1.4-3)
# 4. maptools (version 0.7-16)
# 5. Matrix (version 0.999375-18)
# 6. RColorBrewer (version 1.0-2)
# 7. sp (version 0.9-28)
# 8. spam (version 0.15-2)
# 9. spdep (version 0.4-29)
# 10. spgwr (version 0.6-2) - for GWSS & GWR
# 11. tripack (version 1.2-11)
# 12. car (version 1.2-12) - for MLR
# 13. locfit (version 1.5-4)- for LR

# Base R system version 2.9.0
# N.B. Some of the above packages may still depend on other R packages - download these from R website...

# Relevant data files (see data & ArcGIS directories):

# Excel files...
# 1. ESPON_DATA_NCG_CHALLENGE_10_original.xls
# 2. ESPON_DATA_NCG_CHALLENGE_10_subsets.xls

# ArcGIS files...
# 3. Worked_example_345a_reduced.shp - ArcGIS shapefile of the data...

# The 27 variables...

# "NUTS3","NUTS23","NUTS2","NUTS1","NUTS0" - 5 different NUTS levels
# "New_ID" - relates to the regions name only & is purely numeric
# "NUTS3_2006" - the 2006 NUTS3 version
# "Region_2006" - name of 2006 NUTS3 version
# "X","Y" - centroids of regions
# "EVOGDP_2000_2005_2006" - Evolution of GDP
# and 16 likely contextual variables of "EVOGDP_2000_2005_2006" ...
# "SPAT_TYPE_1_1999_1999"
# "SPAT_TYPE_2_1999_1999"
# "UNEMP_R_2001_1999"
# "LU_AS_1_1996_1999"
# "LU_AS_2_1996_1999"
# "LU_AS_3_1996_1999"
# "LU_UF_1996_1999"
# "LU_AR_1996_1999"
# "LU_PC_1996_1999"
# "NAT_HAZ_2004_1999"
# "SF_CF_1999_1999"
# "SF_R_1999_1999"
# "SF_S_1999_1999"
# "SF_A_1999_1999"
# "CF_T_1999_1999"
# "CF_E_1999_1999"

```

```
# NOTE - This example data set has been reduced to 731 values from an original 1351 values
# see readme in excel files on worked example data.
```

```
# 2. Importing data as a ArcGIS shapefile & using GISTools to do some maps #####
```

```
require(GISTools)
#help(GISTools)
# Ignore all warnings - this code is under development...

# Read in the shapefile...
data1 <- readShapePoly("Worked_example_345a_reduced.shp",
proj4string=CRS("+proj=Lambert_Azimuthal_Equal_Area+datum=D_ETRS_1989+ellps=GCS_ETRS_1989"))
colnames(data1@data)

# renaming each variable - as they have been truncated in ArcGIS...
colnames(data1@data) <- c("NUTS3","NUTS23","NUTS2","NUTS1","NUTS0",
"New_ID","NUTS3_2006","Region_2006",
"X","Y","EVOGDP_2000_2005_2006",
"SPAT_TYPE_1_1999_1999","SPAT_TYPE_2_1999_1999",
"UNEMP_R_2001_1999",
"LU_AS_1_1996_1999","LU_AS_2_1996_1999","LU_AS_3_1996_1999",
"LU_UF_1996_1999","LU_AR_1996_1999","LU_PC_1996_1999",
"NAT_HAZ_2004_1999",
"SF_CF_1999_1999",
"SF_R_1999_1999","SF_S_1999_1999","SF_A_1999_1999",
"CF_T_1999_1999","CF_E_1999_1999")

# Size of data set and adding an order ID...
n <- length(data1@data[,1])
Order_ID <- seq(1,n)
data1@data <- cbind(data1@data, Order_ID)
attach(data1@data)

# Coordinate data only...
coords <- cbind(data1@data[,9],data1@data[,10])

# Creating a shading scheme and plotting a choropleth map of EVOGDP_2000_2005_2006...
shades.1 = auto.shading(EVOGDP_2000_2005_2006,5, cols=brewer.pal(5,'PuBuGn'))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,EVOGDP_2000_2005_2006,shades.1)
title("Evolution of GDP (2000 to 2005)")
choro.legend(-2300000,250000,shades.1,fmt="%4.0f",title='Evolution of GDP (%)',cex=0.8)
map.scale(100000,-1050000,500000,"x1000 km",4,0.5)
north.arrow(200000,750000,40000, col="blue")
```

```

# 3. Exploratory analyses for EVOGDP_2000_2005_2006 relationships #####

data.1 <- cbind(EVOGDP_2000_2005_2006,X,Y)
data.2 <- cbind(EVOGDP_2000_2005_2006,UNEMP_R_2001_1999,NAT_HAZ_2004_1999)
data.3 <-
cbind(EVOGDP_2000_2005_2006,LU_AS_1_1996_1999,LU_AS_2_1996_1999,LU_AS_3_1996_1999)
data.4 <- cbind(EVOGDP_2000_2005_2006,LU_UF_1996_1999,LU_AR_1996_1999,LU_PC_1996_1999)
data.5 <- cbind(EVOGDP_2000_2005_2006,SF_R_1999_1999,SF_S_1999_1999,SF_A_1999_1999)
data.6 <- cbind(EVOGDP_2000_2005_2006,SF_CF_1999_1999,CF_T_1999_1999,CF_E_1999_1999)

cor(data.1,use="pairwise.complete.obs")
cor(data.2,use="pairwise.complete.obs")
cor(data.3,use="pairwise.complete.obs")
cor(data.4,use="pairwise.complete.obs")
cor(data.5,use="pairwise.complete.obs")
cor(data.6,use="pairwise.complete.obs")

X11(width=6,height=6)
pairs(data.1)
X11(width=6,height=6)
pairs(data.2)
X11(width=6,height=6)
pairs(data.3)
X11(width=6,height=6)
pairs(data.4)
X11(width=6,height=6)
pairs(data.5)
X11(width=6,height=6)
pairs(data.6)

X11(width=6,height=4)
boxplot(EVOGDP_2000_2005_2006~SPAT_TYPE_1_1999_1999,xlab="SPAT_TYPE_1_1999_1999",
ylab="EVOGDP_2000_2005_2006",cex=0.5, main="Evolution of GDP with Spatial typology 1")

X11(width=6,height=4)
boxplot(EVOGDP_2000_2005_2006~SPAT_TYPE_2_1999_1999,xlab="SPAT_TYPE_2_1999_1999",
ylab="EVOGDP_2000_2005_2006",cex=0.5, main="Evolution of GDP with Spatial typology 2")

# Exploratory investigations suggests that
# "X"
# "Y"
# "SF_CF_1999_1999"
# "SPAT_TYPE_2_1999_1999"
# have moderate relationships with EVOGDP_2000_2005_2006

# Coding for a categorical variable in a regression model using factor()...
SPAT_TYPE_2_1999_1999.f <- factor(SPAT_TYPE_2_1999_1999)

# For basic MLR analysis...
require(car)

# Full MLR model
mlr.1 <- lm(EVOGDP_2000_2005_2006 ~ X+Y+SF_CF_1999_1999+SPAT_TYPE_2_1999_1999.f)
summary(mlr.1)
vif(mlr.1) # Variance inflation factor (for collinearity)
AIC(mlr.1) # note R gives n*AIC

# AIC stepwise MLR model
mlr.2 <- stepAIC(mlr.1)
summary(mlr.2)
vif(mlr.2)
AIC(mlr.2)

```

```

# Results suggest that mlr.1 model is OK...

# We now assume (for section 4.) that the same explanatory variables
# are also important locally with LR and GWR...

# We can also investigate GW correlations using the spgwr function gw.cov

data.1 <- data1@data
coordinates(data.1) <- c("X", "Y")

# GW summary statistics at observation locations (i.e. region centroids)...
# Calculated using 10% of nearby data.
bwd.1 <- 0.1
gwss <- gw.cov(data.1, vars=c(11,22), adapt=bwd.1, cor = TRUE)
names(gwss$SDF) # The GW summary statistics calculated...

# GW correlations...
GW.corr <- gwss$SDF$cor.EVOGDP_2000_2005_2006.SF_CF_1999_1999.
summary(GW.corr) # some evidence of relationship nonstationarity...

# Updating information in one file
data1@data <- cbind(data1@data, GW.corr)
data1@data <- as.data.frame(data1@data)
attach(data1@data)
data.1 <- data1@data
#fix(data.1)

# A choropleth map of GW correlations ...
shades.2 = auto.shading(GW.corr,5, cols=brewer.pal(5,'Greys'))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,29],shades.2)
title("Relationship nonstationarity: Evolution of GDP with Str./Coh. Fund")
choro.legend(-2300000,250000,shades.2,fmt="%4.1f",title='Correlation',cex=0.8)
map.scale(100000,-1050000,500000,"x1000 km",4,0.5)
north.arrow(200000,750000,40000, col="blue")

# 4. Residual analysis with MLR, LR and GWR models #####

# Using raw residuals from mlr-1 fit...
raw.resids.mlr <- EVOGDP_2000_2005_2006-mlr.1$fitted # Actual minus prediction
summary(raw.resids.mlr)

# Identifying & updating outlier information in one file
cut.off.1 <- quantile(raw.resids.mlr, probs = seq(0, 1, 0.05), na.rm=T) # for 5% tails - alter accordingly...
indicator.1 <- ifelse(raw.resids.mlr>=cut.off.1[2] & raw.resids.mlr<=cut.off.1[20], 0, 1)
data1@data <- cbind(data1@data, raw.resids.mlr, indicator.1)
data1@data <- as.data.frame(data1@data)
attach(data1@data)
data.1 <- data1@data
#fix(data.1)

# Raw residual map for MLR...

```

```

shades.3 = shading(c(cut.off.1[2],cut.off.1[20]),c("red","white","black"))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,30],shades.3)
title("Raw resid. from MLR: High/-ve(red) & High/+ve(black) (5% tails)")
map.scale(100000,-1050000,500000,"x1000 km",4,0.5)
north.arrow(200000,750000,40000, col="blue")

# LR using locfit...
require(locfit)
# Ignore warning message...

# Finding the bandwidth for a non-robust LR (i.e. not a lowess fit)
# using generalised cross-validation (GCV) approach.
summary(gcvplot(EVOGDP_2000_2005_2006~X+Y+SF_CF_1999_1999+SPAT_TYPE_2_1999_1999.f,
data=data.1, scale=F,alpha=seq(0.1,1,by=0.1),
deg=1,kern="tricube",lproc=locfit.raw))

# Choosing a LR fit with bandwidth chosen from above...
bwd.2 <- 0.7
lr <- locfit(EVOGDP_2000_2005_2006~X+Y+SF_CF_1999_1999+SPAT_TYPE_2_1999_1999.f,
data=data.1, scale=F, alpha=bwd.2,deg=1,kern="tricube",lproc=locfit.raw)

# Raw residuals...
lr.p <- fitted.locfit(lr)
raw.resids.lr <- EVOGDP_2000_2005_2006-lr.p # Actual minus prediction
summary(raw.resids.lr)

# Identifying & updating outlier information in one file
cut.off.2 <- quantile(raw.resids.lr, probs = seq(0, 1, 0.05), na.rm=T) # for 5% tails - alter accordingly...
indicator.2 <- ifelse(raw.resids.lr>=cut.off.2[2] & raw.resids.lr<=cut.off.2[20], 0, 1)
data1@data <- cbind(data1@data, raw.resids.lr, indicator.2)
data1@data <- as.data.frame(data1@data)
attach(data1@data)
data.1 <- data1@data
#fix(data.1)

# Raw residual map for LR...
shades.4 = shading(c(cut.off.2[2],cut.off.2[20]),c("red","white","black"))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,32],shades.4)
title("Raw resid. from LR: High/-ve(red) & High/+ve(black) (5% tails)")
map.scale(100000,-1050000,500000,"x1000 km",4,0.5)
north.arrow(200000,750000,40000, col="blue")

# GWR using spgwr...
# Defining the coordinates...
coords.1<-cbind(data.1[,9],data.1[,10])

# Finding the bandwidth for GWR using Akaike Information Criterion (AIC) approach.
#gwr.aic.bwd <-gwr.sel(EVOGDP_2000_2005_2006~X+Y+SF_CF_1999_1999+SPAT_TYPE_2_1999_1999.f,
#data=data.1,coords=coords.1,adapt=TRUE,
#gweight=gwr.bisquare, method="aic")
#gwr.aic.bwd[1] # the optimum bandwidth

# Or finding the bandwidth for GWR using cross-validation (CV) approach.
gwr.cv.bwd <-gwr.sel(EVOGDP_2000_2005_2006~X+Y+SF_CF_1999_1999+SPAT_TYPE_2_1999_1999.f,
data=data.1,coords=coords.1,adapt=TRUE,

```

```

gweight=gwr.bisquare, method="cv")
gwr.cv.bwd[1] # the optimum bandwidth

# Using CV bandwidth...
bwd.3 <- gwr.cv.bwd[1]
gwr.p <-gwr(EVOGDP_2000_2005_2006~X+Y+SF_CF_1999_1999+SPAT_TYPE_2_1999_1999.f,
data=data.1,coords=coords.1,adapt=bwd.3,gweight=gwr.bisquare,predictions=T)
#gwr.p$SDF

# GWR raw residuals...
raw.resids.gwr <- EVOGDP_2000_2005_2006-gwr.p$SDF$pred
summary(raw.resids.gwr)

# Identifying & updating outlier information in one file
cut.off.3 <- quantile(raw.resids.gwr, probs = seq(0, 1, 0.05), na.rm=T) # for 5% tails - alter accordingly...
indicator.3 <-ifelse(raw.resids.gwr>=cut.off.3[2] & raw.resids.gwr<=cut.off.3[20], 0, 1)
data1 @data <- cbind(data1 @data, raw.resids.gwr, indicator.3)
data1 @data <- as.data.frame(data1 @data)
attach(data1 @data)
data.1 <- data1 @data
#fix(data.1)

# Raw residual map for GWR...
shades.5 = shading(c(cut.off.3[2],cut.off.3[20]),c("red","white","black"))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1 @data[,34],shades.5)
title("Raw resids. from GWR: High/-ve(red) & High/+ve(black) (5% tails)")
map.scale(100000,-1050000,500000,"x1000 km",4,0.5)
north.arrow(200000,750000,40000, col="blue")

# NB - Future work should explore the identification of outliers using
# standardised residuals & corresponding robust regression models...

# 5. All identified outliers together #####

# Put all indicator data together...
indicator.4 <- indicator.1+indicator.2+indicator.3
summary(indicator.4)
# Histogram
X11(width=5.3,height=5.7)
hist(indicator.4,br=c(0,1,2,3))

# Thus a strong case for an outlier relates to an observation
# that has a indicator.4 value of 3...

data1 @data <- cbind(data1 @data, indicator.4)
data1 @data <- as.data.frame(data1 @data)
attach(data1 @data)
data.1 <- data1 @data
#fix(data.1)

# A choropleth map of suspected outliers...
shades.6 = shading(c(1,2,3),c("white","yellow","orange","red"))

```

```

X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,36],shades.6)
title("Suspected outliers - weak to strong (yellow to red) evidence")
choro.legend(-2300000,250000,shades.6,over="exactly", between="to under",
fmt="%4.0f",title='Indicator sum (max.: 3)',cex=0.8)
map.scale(100000,-1050000,500000,"x1000 km",4,0.5)
north.arrow(200000,750000,40000, col="blue")

```

## Appendix 5 – R script for worked example 5

```

# 1. Preamble #####

# Worked example 5 - for technical report - challenge 10 - ESPON 2013 database
# NCG - P. Harris & M. Charlton
# 7/2/10

# Objective - to identify statistical clusters in "EVOGDP_2000_2005_2006" with
# respect to key spatial moments...
# GW means/standard deviations; Relationships - GW correlations/regressions;
# Spatial autocorrelation (Moran's I)
# For relationships -
# EVOGDP_2000_2005_2006 is related to some subset of the following explanatory/contextual data:
# "X"
# "Y"
# "SPAT_TYPE_1_1999_1999"
# "SPAT_TYPE_2_1999_1999"
# "UNEMP_R_2001_1999"
# "LU_AS_1_1996_1999"
# "LU_AS_2_1996_1999"
# "LU_AS_3_1996_1999"
# "LU_UF_1996_1999"
# "LU_AR_1996_1999"
# "LU_PC_1996_1999"
# "NAT_HAZ_2004_1999"
# "SF_CF_1999_1999"
# "SF_R_1999_1999"
# "SF_S_1999_1999"
# "SF_A_1999_1999"
# "CF_T_1999_1999"
# "CF_E_1999_1999"

# Methods: univariate & multivariate - all spatial
# Only statistical methods:
# 1. GWSS (geographically weighted summary statistics)
# 2. MLR (multiple linear regression) & GWR (geographically weighted regression)

```

```

# 3. Global and local Moran's I

# R packages needed.....
# 1. GISTools (version 0.5-4) - depends on 2 to 11...
# 2. foreign (version 0.8-30)
# 3. gpclib (version 1.4-3)
# 4. maptools (version 0.7-16)
# 5. Matrix (version 0.999375-18)
# 6. RColorBrewer (version 1.0-2)
# 7. sp (version 0.9-28)
# 8. spam (version 0.15-2)
# 9. spdep (version 0.4-29) - for global and local Moran's I
# 10. spgwr (version 0.6-2) - for GWSS and GWR
# 11. tripack (version 1.2-11)
# 12. car (version 1.2-12) - for MLR

# Base R system version 2.9.0
# N.B. Some of the above packages may still depend on other R packages - download these from R website...

# Relevant data files (see data & ArcGIS directories):

# Excel files...
# 1. ESPON_DATA_NCG_CHALLENGE_10_original.xls
# 2. ESPON_DATA_NCG_CHALLENGE_10_subsets.xls

# ArcGIS files...
# 3. Worked_example_345a_reduced.shp - ArcGIS shapefile of the data...

# The 27 variables...

# "NUTS3","NUTS23","NUTS2","NUTS1","NUTS0" - 5 different NUTS levels
# "New_ID" - relates to the regions name only & is purely numeric
# "NUTS3_2006" - the 2006 NUTS3 version
# "Region_2006" - name of 2006 NUTS3 version
# "X","Y" - centroids of regions
# "EVOGDP_2000_2005_2006" - Evolution of GDP
# and 16 likely contextual variables of "EVOGDP_2000_2005_2006" ...
# "SPAT_TYPE_1_1999_1999"
# "SPAT_TYPE_2_1999_1999"
# "UNEMP_R_2001_1999"
# "LU_AS_1_1996_1999"
# "LU_AS_2_1996_1999"
# "LU_AS_3_1996_1999"
# "LU_UF_1996_1999"
# "LU_AR_1996_1999"
# "LU_PC_1996_1999"
# "NAT_HAZ_2004_1999"
# "SF_CF_1999_1999"
# "SF_R_1999_1999"
# "SF_S_1999_1999"
# "SF_A_1999_1999"
# "CF_T_1999_1999"
# "CF_E_1999_1999"

# NOTE - This example data set has been reduced to 731 values from an original 1351 values

```



```

# see readme in excel files on worked example data.

# 2. Importing data as a ArcGIS shapefile & using GISTools to do some maps #####

require(GISTools)
#help(GISTools)
# Ignore all warnings - this code is under development...

# Read in the shapefile...
data1 <- readShapePoly("Worked_example_345a_reduced.shp",
proj4string=CRS("+proj=Lambert_Azimuthal_Equal_Area+datum=D_ETRS_1989+ellps=GCS_ETRS_1989"))
colnames(data1@data)

# renaming each variable - as they have been truncated in ArcGIS...
colnames(data1@data) <- c("NUTS3","NUTS23","NUTS2","NUTS1","NUTS0",
"New_ID","NUTS3_2006","Region_2006",
"X","Y","EVOGDP_2000_2005_2006",
"SPAT_TYPE_1_1999_1999","SPAT_TYPE_2_1999_1999",
"UNEMP_R_2001_1999",
"LU_AS_1_1996_1999","LU_AS_2_1996_1999","LU_AS_3_1996_1999",
"LU_UF_1996_1999","LU_AR_1996_1999","LU_PC_1996_1999",
"NAT_HAZ_2004_1999",
"SF_CF_1999_1999",
"SF_R_1999_1999","SF_S_1999_1999","SF_A_1999_1999",
"CF_T_1999_1999","CF_E_1999_1999")

# Size of data set and adding an order ID...
n <- length(data1@data[,1])
Order_ID <- seq(1,n)
data1@data <- cbind(data1@data, Order_ID)
attach(data1@data)

# Coordinate data only...
coords <- cbind(data1@data[,9],data1@data[,10])

# Creating a shading scheme and plotting a choropleth map of EVOGDP_2000_2005_2006...
shades.1 = auto.shading(EVOGDP_2000_2005_2006,5, cols=brewer.pal(5,'YlOrBr'))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,EVOGDP_2000_2005_2006,shades.1)
title("Evolution of GDP (2000 to 2005)")
choro.legend(-2300000,250000,shades.1,fmt="%4.0f",title='Evolution of GDP (%)',cex=0.8)
map.scale(100000,-1050000,500000,"x1000 km",4,0.5)
north.arrow(200000,750000,40000, col="blue")

# 3. Exploratory analyses for EVOGDP_2000_2005_2006 spatial moments #####

# Global summary statistics...
summary(EVOGDP_2000_2005_2006)

```

```

# standard deviation
sd <- (var(EVOGDP_2000_2005_2006))^0.5
sd

# As in worked example 4 we can investigate GW summary statistics
# using the spgwr function gw.cov

data.1 <- data1@data
coordinates(data.1) <- c("X", "Y")

# GW summary statistics at observation locations (i.e. region centroids)...
# Calculated using 10% of nearby data.
bwd.1 <- 0.1
gwss <- gw.cov(data.1, vars=c(11,22), adapt=bwd.1, cor = TRUE)
names(gwss$SDF) # The GW summary statistics calculated...

# GW means...
GW.mean <- gwss$SDF$mean.EVOGDP_2000_2005_2006
summary(GW.mean) # some evidence of mean nonstationarity...

# Updating information in one file
data1@data <- cbind(data1@data, GW.mean)
data1@data <- as.data.frame(data1@data)
attach(data1@data)
data.1 <- data1@data
#fix(data.1)

# A choropleth map of GW means ...
shades.2 = auto.shading(GW.mean,5, cols=brewer.pal(5,'YlOrBr'))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,29],shades.2)
title("Mean nonstationarity: Evolution of GDP")
choro.legend(-2300000,2500000,shades.2,fmt="%4.1f",title='Mean',cex=0.8)
map.scale(100000,-1050000,500000,"x1000 km",4,0.5)
north.arrow(200000,750000,40000, col="blue")

# GW standard deviations...
GW.sd <- gwss$SDF$sd.EVOGDP_2000_2005_2006
summary(GW.sd) # some evidence of SD nonstationarity...

# Updating information in one file
data1@data <- cbind(data1@data, GW.sd)
data1@data <- as.data.frame(data1@data)
attach(data1@data)
data.1 <- data1@data
#fix(data.1)

# A choropleth map of GW SDs ...
shades.3 = auto.shading(GW.sd,5, cols=brewer.pal(5,'RdPu'))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,30],shades.3)
title("Standard Deviation nonstationarity: Evolution of GDP")
choro.legend(-2300000,2500000,shades.3,fmt="%4.1f",title='SD',cex=0.8)
map.scale(100000,-1050000,500000,"x1000 km",4,0.5)
north.arrow(200000,750000,40000, col="blue")

```

```

# For all GW summary statistics randomisation tests can be used to identify
# locations of unusually different local statistics -
# see Fotheringham et al. (2002); Harris and Brunsdon (2010)

# As an example: randomisation test for the standard deviation...
n.r1 <- 99 # Number of randomisations (the more the better)
out.x <- matrix(nrow=n,ncol=n.r1)
for(i in 1:n.r1)
{
  rand.dat <- sample(data.1[,11])
  data.2 <- cbind(data1@data, rand.dat)
  data.2 <- as.data.frame(data.2)
  attach(data.2)
  coordinates(data.2) <- c("X", "Y")
  gwss.rand <- gw.cov(data.2, vars=c(31), adapt=bwd.1)
  out.x[,i] <- gwss.rand$SDF$sd.V1
}
# combining the randomisation results with the actual result...
out.x1 <- cbind(GW.sd, out.x)
out.x2 <- t(apply(out.x1,1,rank))
Random.sd <- out.x2[,1]

# Updating information in one file
data1@data <- cbind(data1@data, Random.sd)
data1@data <- as.data.frame(data1@data)
attach(data1@data)
data.1 <- data1@data
#fix(data.1)

# A choropleth map of randomisation test result ...
shades.4 = shading(c(2.5,97.5),c("blue","white","green")) # test at 95% level...
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,31],shades.4)
title("Areas of unusually high (green) and low (blue) standard deviation")
map.scale(100000,-750000,500000,"x1000 km",4,0.5)
north.arrow(200000,750000,40000, col="red")
text(-320000,-1150000, "Randomisation test at 95% level for Evolution of GDP")

# 4. Exploratory analyses for EVOGDP_2000_2005_2006 w.r.t global relationships #

# From exploratory investigations of worked example 4:
# "SF_CF_1999_1999"
# "SPAT_TYPE_2_1999_1999"
# have moderate relationships with EVOGDP_2000_2005_2006
# i.e.

# Correlation & scatterplot...
cor(EVOGDP_2000_2005_2006,SF_CF_1999_1999,use="pairwise.complete.obs")
X11(width=6,height=6)
plot(EVOGDP_2000_2005_2006,SF_CF_1999_1999, main="Evolution of GDP with Str/Coh Fund",
pch=19, cex=0.5)

```

```

# Boxplot for categorical variable...
X11(width=6,height=4)
boxplot(EVOGDP_2000_2005_2006~SPAT_TYPE_2_1999_1999,xlab="SPAT_TYPE_2_1999_1999",
ylab="EVOGDP_2000_2005_2006",cex=0.5, main="Evolution of GDP with Spatial typology 2")

# Coding for a categorical variable in a regression model using factor()...
SPAT_TYPE_2_1999_1999.f <- factor(SPAT_TYPE_2_1999_1999)

# For useful basic MLR analysis...
require(car)

# Full MLR model
mlr.1 <- lm(EVOGDP_2000_2005_2006 ~ SF_CF_1999_1999+SPAT_TYPE_2_1999_1999.f)
summary(mlr.1)
vif(mlr.1) # Variance inflation factor (for collinearity)
AIC(mlr.1) # note R gives n*AIC

# AIC stepwise MLR model
mlr.2 <- step(mlr.1)
summary(mlr.2)
vif(mlr.2)
AIC(mlr.2)

# Results suggest that mlr.1 model is OK...

# We also assume that the same explanatory variables
# are also important locally with GWR...

# 5. Exploratory analyses for EVOGDP_2000_2005_2006 w.r.t local relationships ##

# We can also investigate GW correlations from the spgwr function gw.cov output in section 3.

# GW correlations...
GW.corr <- gwss$SDF$cor.EVOGDP_2000_2005_2006.SF_CF_1999_1999.
summary(GW.corr) # some evidence of relationship nonstationarity...

# Updating information in one file
data1 @data <- cbind(data1 @data, GW.corr)
data1 @data <- as.data.frame(data1 @data)
attach(data1 @data)
data.1 <- data1 @data
#fix(data.1)

# A choropleth map of GW correlations ...
shades.5 = auto.shading(GW.corr,5, cols=brewer.pal(5,'Greys'))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1 @data[,32],shades.5)
title("Relationship nonstationarity: Evolution of GDP with Str./Coh. Fund")
choro.legend(-2300000,2500000,shades.5,fmt="%4.1f",title='Correlation',cex=0.8)
map.scale(100000,-1050000,500000,"x1000 km",4,0.5)
north.arrow(200000,750000,40000, col="blue")

# Now using GWR to explore local relationships (using spgwr)...

```

```

# Defining the coordinates...
coords.1<-cbind(data.1[,9],data.1[,10])

# Finding the bandwidth for GWR using cross-validation (CV) approach.
gwr.cv.bwd <-gwr.sel(EVOGDP_2000_2005_2006~SF_CF_1999_1999+SPAT_TYPE_2_1999_1999.f,
data=data.1,coords=coords.1,adapt=TRUE,
gweight=gwr.bisquare, method="cv")
gwr.cv.bwd[1] # the optimum bandwidth

# GWR using CV bandwidth...
bwd.2 <- gwr.cv.bwd[1]
gwr.1 <-gwr(EVOGDP_2000_2005_2006~SF_CF_1999_1999+SPAT_TYPE_2_1999_1999.f,
data=data.1,coords=coords.1,adapt=bwd.2,gweight=gwr.bisquare)
#gwr.1$$SDF

# As an example, only investigating coefficients (or parameters) for SPAT_TYPE_2_1999_1999 class 2
gwr.coeff.1 <- gwr.1$$SDF$SPAT_TYPE_2_1999_1999.f2

# Updating information in one file
data1 @data <- cbind(data1 @data, gwr.coeff.1)
data1 @data <- as.data.frame(data1 @data)
attach(data1 @data)
data.1 <- data1 @data
#fix(data.1)

# A choropleth map of this particular part of the GWR output ...
shades.6 = auto.shading(gwr.coeff.1, 5, cols=brewer.pal(5,'Blues'))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1 @data[,33],shades.6)
title("Relationship nonstationarity: coefficient for spatial typology 2 (class 2)")
choro.legend(-2300000,250000,shades.6,fmt="%1.2f",title='GWR coefficient',cex=0.8)
map.scale(100000,-1050000,500000,"x1000 km",4,0.5)
north.arrow(200000,750000,40000, col="green")

# Randomisation test -
# To gauge whether or not any observed variation in the GWR coefficients is unusual...

# First get the actual variance of each local coefficient estimate...
output.x1 <- as(gwr.1$$SDF, "data.frame")[,2:8]
output.1 <- as.vector(apply(output.x1,2,var))

# Now find the same variances based on 99 randomisations of the data...
output.2 <- matrix(nrow=n.r1,ncol=7) # i.e. 7 columns (intercept plus number of indep. variables)
for(i in 1:n.r1)
{
print(i) # counter
c1<-t(coords.1)
c1 <- as.data.frame(c1)
c1.s <-sample(c1,length(data.1[,1]))
c1.s <- as.data.frame(c1.s)
c1.s <- t(c1.s)
coords.2 <- as.matrix(c1.s)
gwr.2 <-gwr(EVOGDP_2000_2005_2006~SF_CF_1999_1999+SPAT_TYPE_2_1999_1999.f,
data=data.1,coords=coords.2,adapt=bwd.2,gweight=gwr.bisquare)
output.x2 <- as(gwr.2$$SDF, "data.frame")[,2:8]
output.2[i,] <- as.vector(apply(output.x2,2,var))
}

# p-values for each coefficient estimate

```

```

output.3 <- rbind(output.1,output.2)
r.1 <- rank(output.3[,1])
r.11 <- ((n.r1+2)-r.1[1])/(n.r1+1)
r.2 <- rank(output.3[,2])
r.22 <- ((n.r1+2)-r.2[1])/(n.r1+1)
r.3 <- rank(output.3[,3])
r.33 <- ((n.r1+2)-r.3[1])/(n.r1+1)
r.4 <- rank(output.3[,4])
r.44 <- ((n.r1+2)-r.4[1])/(n.r1+1)
r.5 <- rank(output.3[,5])
r.55 <- ((n.r1+2)-r.5[1])/(n.r1+1)
r.6 <- rank(output.3[,6])
r.66 <- ((n.r1+2)-r.6[1])/(n.r1+1)
r.7 <- rank(output.3[,7])
r.77 <- ((n.r1+2)-r.7[1])/(n.r1+1)

# Thus in this case, all Monte Carlo tests are based on 99 randomisations of the data.
# The larger the p-value,the more support is given to the null hypothesis
# of a stationary regression coefficient estimate.
rand.test.1 <- cbind(r.11,r.22,r.33,r.44,r.55,r.66,r.77)
rand.test.1

# 6. Global and local autocorrelation #####

# Global Moran's I
# Local Moran's I (a Local Indicator of Spatial Association LISA)

require(spdep) # for global and local Moran's I

# Firstly, two different examples to define spatial distances or spatial topology.
# A measure of distance is needed to calculate local and global Moran's I statistics.

data1.labs = poly.labels(data1)

# 1. Queen's case spatial topology (from chess)
data1.nb1 = poly2nb(data1)
X11(width=8,height=7)
par(mar=c(0,0,2,0))
plot(data1,col='grey')
plot(data1.nb1,coordinates(data1.labs),col='red',add=TRUE)
title("Queen's case spatial topology")
map.scale(100000,-1050000,500000,"x1000 km",4,0.5)
north.arrow(200000,750000,40000, col="blue")

# 2. Alternatively, base topology on nearness of polygons rather than contiguity.
# Here we define two polygons as neighbours if they are within some distance d of one another.
# Thus letting d = 100000m, for example...
data1.nb2 = dnearneigh(poly.labels(data1),0,100000)
X11(width=8,height=7)
par(mar=c(0,0,2,0))
plot(data1,col='grey')
plot(data1.nb2,coordinates(data1.labs),col='red',add=TRUE)
title("Regions whose centroids are within 100km of each other")
map.scale(100000,-1050000,500000,"x1000 km",4,0.5)
north.arrow(200000,750000,40000, col="blue")

```

```

# Tests for global Moran's I statistic using 'Queens case topology' contiguity matrix
data1.lw = nb2listw(data1.nb1,zero.policy=T)
attach(data.frame(data1))

# Global Moran's I:
moran(EVOGDP_2000_2005_2006, data1.lw, n, Szero(data1.lw), zero.policy=T)

# The assumptions underlying the following test are sensitive to the form
# of the graph of neighbour relationships and other factors.
# Results may be checked against those of moran.mc permutations.
moran.test(EVOGDP_2000_2005_2006,data1.lw, zero.policy=T)

# A permutation test for Moran's I statistic calculated by using nsim random permutations of x for
# the given spatial weighting scheme, to establish the rank of the observed statistic in relation to the
# nsim simulated values.
moran.mc(EVOGDP_2000_2005_2006,data1.lw, zero.policy=T,nsim=10000)

# Local Moran's I also using 'Queens case topology' contiguity matrix
Local.moran <- localmoran(EVOGDP_2000_2005_2006,data1.lw, zero.policy=T)

# Summary of local Moran's I
summary(Local.moran)

# Updating information in one file
data1@data <- cbind(data1@data, Local.moran[,1])
data1@data <- as.data.frame(data1@data)
attach(data1@data)
data.1 <- data1@data
#fix(data.1)

# A choropleth map of local Moran's I for EVOGDP_2000_2005_2006...
shades.7 = shading(c(0,0.572),c("red","grey","blue")) # shading relates to global value...
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,34],shades.7)
title("Autocorrelation nonstationarity: Local Moran's I for Evolution of GDP")
map.scale(100000,-750000,500000,"x1000 km",4,0.5)
north.arrow(200000,750000,40000, col="black")
text(-360000,-1150000, "-ve (red); +ve/below global (grey); +ve/above global (blue)")

```

## Appendix 6 – R script for worked example 6

```

# 1. Preamble #####

# Worked example 6 - for technical report - challenge 10 - ESPON 2013 database

```

```

# NCG - P. Harris & M. Charlton
# 8/2/10

# Objective - to identify statistical outliers in:
# "EVOGDP_2000_2005_2006" at three different NUTS levels (2, 1 & 0)
# - i.e. investigate the effects of MAUP on outlier identification
# Results are compared with those of worked example 2 for NUTS3.

# Methods: univariate - aspatial & spatial (i.e. the same as that presented in worked example 2)
# Only statistical methods:
# 1. Standard and Adjusted boxplots,
# 2. Hawkins' test (includes the use of GWSS -
#     geographically weighted summary statistics - GW means and variances),
# 3. LM (local mean, i.e. a GW mean)
# 4. MLR (multiple linear regression),
# 5. LR (local regression) &
# 6. GWR (geographically weighted regression)

# R packages needed.....
# 1. GISTools (version 0.5-4) - depends on 2 to 11...
# 2. foreign (version 0.8-30)
# 3. gpclib (version 1.4-3)
# 4. maptools (version 0.7-16)
# 5. Matrix (version 0.999375-18)
# 6. RColorBrewer (version 1.0-2)
# 7. sp (version 0.9-28)
# 8. spam (version 0.15-2)
# 9. spdep (version 0.4-29)
# 10. spgwr (version 0.6-2) - for GWSS & GWR
# 11. tripack (version 1.2-11)
# 12. moments (version 0.11) - for skewness
# 13. robustbase (version 0.4-5) - for adjusted boxplots
# 14. locfit (version 1.5-4)- for LR

# Base R system version 2.9.0
# N.B. Some of the above packages may still depend on other R packages -
# download these from R website...

# Relevant data files (see data & ArcGIS directories):

# Excel files...
# 1. ESPON_DATA_NCG_CHALLENGE_10_original.xls
# 2. ESPON_DATA_NCG_CHALLENGE_10_subsets.xls

# ArcGIS files...
# 3. Worked_example_6c_Dissolve_nuts2a.shp - ArcGIS shapefile of the NUTS2 data (278 values)
# 4. Worked_example_6c_Dissolve_nuts1a.shp - ArcGIS shapefile of the NUTS1 data (95 values)
# 5. Worked_example_6c_Dissolve_nuts0a.shp - ArcGIS shapefile of the NUTS0 data (30 values)

# The 10 variables...

# "NUTSx", - the NUTS level - 2, 1 or 0
# "NUTS3_outlier_mean" - mean of outlier indicator.8 from worked example 2

```



```

#           when going from NUTS 3 to larger scale
# "NUTS3_outlier_max" - maximum of outlier indicator.8 from worked example 2
#           when going from NUTS 3 to larger scale
# "GDP_2000_2006" - mean of NUTS3 data when going from NUTS 3 to larger scale
# "GDP_2005_2006" - mean of NUTS3 data when going from NUTS 3 to larger scale
# "POP_T_2000_2006" - mean of NUTS3 data when going from NUTS 3 to larger scale
# "POP_T_2005_2006" - mean of NUTS3 data when going from NUTS 3 to larger scale
# "X","Y" - centroids of NUTS regions
# "EVOGDP_2000_2005_2006" - the variable of interest re-calculated from
#           the relevant variables above

# Change the following script in six places to go through each NUTS level...

# 2. Importing data as a ArcGIS shapefile & using GISTools to do a map... #####

require(GISTools)
#help(GISTools)
# Ignore all warnings - this code is under development...

# Read in the shapefile...
#data1 <- readShapePoly("Worked_example_6c_Dissolve_nuts2a.shp",
data1 <- readShapePoly("Worked_example_6c_Dissolve_nuts1a.shp",
#data1 <- readShapePoly("Worked_example_6c_Dissolve_nuts0a.shp",
proj4string=CRS("+proj=Lambert_Azimuthal_Equal_Area+datum=D_ETRS_1989+ellps=GCS_ETRS_1989"))
colnames(data1@data)

# Renaming each variable - as they have been altered by ArcGIS commands...
colnames(data1@data) <- c("NUTS2", "NUTS3_outlier_mean", "NUTS3_outlier_max",
"GDP_2000_2006", "GDP_2005_2006", "POP_T_2000_2006", "POP_T_2005_2006",
"X","Y")

# Size of data set and adding an order ID...
n <- length(data1@data[,1])
Order_ID <- seq(1,n)
data1@data <- cbind(data1@data, Order_ID)
attach(data1@data)

# Calculating the new EVOGDP_2000_2005_2006 values...
EVOGDP_2000_2005_2006 <-
((data1@data[,5]/data1@data[,7])*1000)/((data1@data[,4]/data1@data[,6])*1000)*100
data1@data <- cbind(data1@data, EVOGDP_2000_2005_2006)
data1@data <- as.data.frame(data1@data)
attach(data1@data)
data.1 <- data1@data
#fix(data.1)

# Creating a shading scheme and plotting a choropleth map of EVOGDP_2000_2005_2006...
shades.1 = auto.shading(EVOGDP_2000_2005_2006,5, cols=brewer.pal(5,'Greys'))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,EVOGDP_2000_2005_2006,shades.1)
title("Evolution of GDP (2000 to 2005)")
choro.legend(-2400000,2200000,shades.1,fmt="%4.0f",title='Evolution of GDP (%)',cex=0.8)

```

```

map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")
#text(1500000,2400000, "NUTS level 2", cex=2, col=3)
text(1500000,2400000, "NUTS level 1", cex=2, col=3)
#text(1500000,2400000, "NUTS level 0", cex=2, col=3)

# 3. Boxplots #####

# Let EVOGDP_2000_2005_2006 be z1...
z1 <- EVOGDP_2000_2005_2006

# Exploring this data...
summary(z1) # summary statistics
sort(z1) # ordered data

# Histogram
X11(width=5.3,height=5.7)
hist(z1, main="Histogram: Evolution of GDP (2000 to 2005)",xlab="Evolution of GDP")

# Standard boxplot with defaults
X11(width=5.3,height=5.7)
boxplot(z1, main="Std. boxplot: Evolution of GDP (2000 to 2005)", pch=19, cex=0.5)

# Standard Boxplot statistics...
# Change 'coef' accordingly...
# Default 'coef' is 1.5...
# The higher the 'coef' value the stricter the limits/cut-offs & vice versa...
bp <- boxplot.stats(z1, coef=1.5)
bp$stats
bp$stats[1] # the lower limit/cut-off - i.e. values below are deemed outlying...
bp$stats[5] # the upper limit/cut-off - i.e. values above are deemed outlying...
bp$conf
sort(bp$out)
length(bp$out) # number of potential outliers...
# help(boxplot.stats) # for details...

# Identifying & updating outlier information in one file
indicator.1 <-ifelse(z1>bp$stats[1]& z1<bp$stats[5], 0, 1) # i.e. suspected outliers...
data1 @data <- cbind(data1 @data, indicator.1)
data1 @data <- as.data.frame(data1 @data)
attach(data1 @data)
data.1 <- data1 @data
#fix(data.1)

# A choropleth map of standard boxplot outliers
shades.2 = shading(c(0,1,2),c("blue","white","red")) # i.e. white - no & red - yes
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1 @data[,12],shades.2)
title("Std. boxplot outliers (regions coloured red)")
map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")

# Need moments package to assess skewness (before adjusted boxplots)
require(moments)
# Ignore warning message...

```

```

skewness(z1)

# Package for adjusted boxplots...
require(robustbase)

# Adjusted boxplot with defaults
X11(width=5.3,height=5.7)
adjbox(z1, main="Adj. boxplot: Evolution of GDP (2000 to 2005)", pch=19, cex=0.5)

# Adjusted Boxplot statistics...
# Change 'coef' accordingly...
# Default 'coef' is 1.5...
# The higher the 'coef' value the stricter the limits/cut-offs & vice versa...
abp <- adjboxStats(z1, coef=1.5)
abp$stats
abp$stats[1] # the lower limit/cut-off - i.e. values below are deemed outlying...
abp$stats[5] # the upper limit/cut-off - i.e. values above are deemed outlying...
abp$conf
sort(abp$out)
length(abp$out) # number of potential outliers...
#help(adjboxStats) # for details...

# Identifying & updating outlier information in one file
indicator.2 <- ifelse(z1>abp$stats[1]& z1<abp$stats[5], 0, 1) # i.e. suspected outliers...
data1@data <- cbind(data1@data, indicator.2)
data1@data <- as.data.frame(data1@data)
attach(data1@data)
data.1 <- data1@data
#fix(data.1)

# A choropleth map of adjusted boxplot outliers
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,13],shades.2)
title("Adj. boxplot outliers (regions coloured red)")
map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")

# 4. GW summary statistics and Hawkins' Spatial Outlier Test #####

# First need to find GW means (i.e. LMs) and GW variances for Hawkin's test...
# In this case, using the gw.cov function in spgwr to find the GW means/variances...

# To re-cap...
colnames(data.1)

# Defining coordinates....
coordinates(data.1) <- c("X", "Y")

# GW summary statistics at observation locations (i.e. region centroids)...
# Calculated using 10% of nearby EVOGDP_2000_2005_2006 data.
bwd.1 <- 0.1
gwss <- gw.cov(data.1, vars=11, adapt=bwd.1)
#help(gw.cov) # for details...
names(gwss$SDF) # The GW summary statistics calculated...

```

```

# GW means and variances...
GW.mean <- gwss$SDF$mean.V1
GW.variance <- (gwss$SDF$sd.V1)^2

# Hawkins' Test for Spatial Outliers...
Hawk.N <- bwd.1*length(X) # number of neighbouring data
Hawk.lm <- GW.mean # the local mean at observation points
Hawk.alv <- mean(GW.variance) # the average local variance with same bandwidth

Hawk.test <- (Hawk.N*(EVOGDP_2000_2005_2006-Hawk.lm)^2)/((Hawk.N+1)*Hawk.alv) # test statistic
summary(Hawk.test)

# Critical values of the chi-squared distribution
chi_10 <- 2.70554
chi_5 <- 3.84146
chi_2.5 <- 5.02389
chi_1 <- 6.63490
chi_0.5 <- 7.87944
chi_0.01 <- 10.828

# Updating outlier information in one file
indicator.3 <- ifelse(Hawk.test <= chi_5, 0, 1) # change critical level accordingly...
data1@data <- cbind(data1@data, Hawk.test, indicator.3)
data1@data <- as.data.frame(data1@data)
attach(data1@data)
data.1 <- data1@data
#fix(data.1)

# A choropleth map of spatial outliers ...
shades.3 = shading(c(chi_5,chi_1,chi_0.01),c("white","yellow","orange","red"))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,14],shades.3)
title("Spatial outliers: at 5/1/0.01 % (yellow/orange/red) critical levels")
map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")

# 5. Residual analysis with LM, MLR, LR and GWR models #####

# LM...
# Using GW.mean from before...
GW.mean.r <- EVOGDP_2000_2005_2006-GW.mean # Actual minus prediction
summary(GW.mean.r)

# Identifying & updating outlier information in one file
cut.off.1 <- quantile(GW.mean.r, probs = seq(0, 1, 0.05), na.rm=T) # for 5% tails
indicator.4 <- ifelse(GW.mean.r >= cut.off.1[2] & GW.mean.r <= cut.off.1[20], 0, 1)
data1@data <- cbind(data1@data, GW.mean.r, indicator.4)
data1@data <- as.data.frame(data1@data)
attach(data1@data)
data.1 <- data1@data
#fix(data.1)

# Raw residual map for LM...

```

```

shades.4 = shading(c(cut.off.1[2],cut.off.1[20]),c("red","white","black"))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,16],shades.4)
title("Raw resids. from LM: High/-ve(red) & High/+ve(black) (5% tails)")
map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")

# MLR...
# First- & second-order polynomial fits of the coordinate data...
mlr.1 <- lm(EVOGDP_2000_2005_2006 ~ X+Y)
mlr.2 <- lm(EVOGDP_2000_2005_2006 ~ X+Y+I(X^2)+I(Y^2)+I(X*Y))
summary(mlr.1)
summary(mlr.2)

# Choosing a second-order MLR fit...

# Using raw residuals as in LM fit...
raw.resids.mlr <- EVOGDP_2000_2005_2006-mlr.2$fitted # Actual minus prediction
summary(raw.resids.mlr)

# Identifying & updating outlier information in one file
cut.off.2 <- quantile(raw.resids.mlr, probs = seq(0, 1, 0.05), na.rm=T) # for 5% tails
indicator.5 <- ifelse(raw.resids.mlr>=cut.off.2[2] & raw.resids.mlr<=cut.off.2[20], 0, 1)
data1@data <- cbind(data1@data, raw.resids.mlr, indicator.5)
data1@data <- as.data.frame(data1@data)
attach(data1@data)
data.1 <- data1@data
#fix(data.1)

# Raw residual map for MLR...
shades.5 = shading(c(cut.off.2[2],cut.off.2[20]),c("red","white","black"))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,18],shades.5)
title("Raw resids. from MLR: High/-ve(red) & High/+ve(black) (5% tails)")
map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")

# LR...
# With coordinate data as explanatory variables (i.e. first-order polynomial).

# Using locfit...
require(locfit)
# Ignore warning message...

# Finding the bandwidth for a non-robust LR (i.e. not a lowest fit)
# using generalised cross-validation (GCV) approach.
summary(gcvplot(EVOGDP_2000_2005_2006~X+Y,data=data.1, scale=F,
#alpha=seq(0.02,0.1,by=0.01), # for NUTS2
alpha=seq(0.1,0.2,by=0.01), # for NUTS1
#alpha=seq(0.2,1,by=0.1), # for NUTS0
deg=1,kern="tricube",lproc=locfit.raw))

# Choosing a LR fit with bandwidth chosen from above...
#bwd.2 <- 0.03 # for NUTS2
bwd.2 <- 0.2 # for NUTS1
#bwd.2 <- 0.6 # for NUTS0
lr <- locfit(EVOGDP_2000_2005_2006~X+Y,data=data.1, scale=F, alpha=bwd.2,

```

```

deg=1,kern="tricube",lproc=locfit.raw)

# Raw residuals...
lr.p <- fitted.locfit(lr)
raw.resids.lr <- EVOGDP_2000_2005_2006-lr.p # Actual minus prediction
summary(raw.resids.lr)

# Identifying & updating outlier information in one file
cut.off.3 <- quantile(raw.resids.lr, probs = seq(0, 1, 0.05), na.rm=T) # for 5% tails
indicator.6 <- ifelse(raw.resids.lr >= cut.off.3[2] & raw.resids.lr <= cut.off.3[20], 0, 1)
data1 @data <- cbind(data1 @data, raw.resids.lr, indicator.6)
data1 @data <- as.data.frame(data1 @data)
attach(data1 @data)
data.1 <- data1 @data
#fix(data.1)

# Raw residual map for LR...
shades.6 = shading(c(cut.off.3[2],cut.off.3[20]),c("red","white","black"))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1 @data[,20],shades.6)
title("Raw resids. from LR: High/-ve(red) & High/+ve(black) (5% tails)")
map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")

# GWR...
# With coordinate data as explanatory variables (i.e. first-order polynomial).
# Using spgwr...

# Defining the coordinates...
coords.1 <- cbind(data.1[,8],data.1[,9])

# Finding the bandwidth for GWR using Akaike Information Criterion (AIC) approach.
#gwr.aic.bwd <- gwr.sel(EVOGDP_2000_2005_2006~X+Y,data=data.1,coords=coords.1,adapt=TRUE,
#gweight=gwr.bisquare, method="aic")
#gwr.aic.bwd[1] # the optimum bandwidth

# Or finding the bandwidth for GWR using cross-validation approach.
gwr.cv.bwd <- gwr.sel(EVOGDP_2000_2005_2006~X+Y,data=data.1,coords=coords.1,adapt=TRUE,
gweight=gwr.bisquare, method="cv")
gwr.cv.bwd[1] # the optimum bandwidth

bwd.3 <- gwr.cv.bwd[1]
gwr.p <- gwr(EVOGDP_2000_2005_2006~X+Y,data=data.1,coords=coords.1,adapt=bwd.3,
gweight=gwr.bisquare,predictions=T)
#gwr.p$SDF

# GWR raw residuals...
raw.resids.gwr <- EVOGDP_2000_2005_2006-gwr.p$SDF$pred
summary(raw.resids.gwr)

# Identifying & updating outlier information in one file
cut.off.4 <- quantile(raw.resids.gwr, probs = seq(0, 1, 0.05), na.rm=T) # for 5% tails
indicator.7 <- ifelse(raw.resids.gwr >= cut.off.4[2] & raw.resids.gwr <= cut.off.4[20], 0, 1)
data1 @data <- cbind(data1 @data, raw.resids.gwr, indicator.7)
data1 @data <- as.data.frame(data1 @data)
attach(data1 @data)
data.1 <- data1 @data
#fix(data.1)

# Raw residual map for GWR...

```

```

shades.7 = shading(c(cut.off.4[2],cut.off.4[20]),c("red","white","black"))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,22],shades.7)
title("Raw resid. from GWR: High/-ve(red) & High/+ve(black) (5% tails)")
map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")

# 6. All identified outliers together #####

# Put all indicator data together...
indicator.8 <- indicator.1+indicator.2+indicator.3+indicator.4+indicator.5+indicator.6+indicator.7
summary(indicator.8)
# Histogram
X11(width=5.3,height=5.7)
hist(indicator.8,br=c(0,1,2,3,4,5,6,7))

# Thus a strong case for an outlier relates to an observation
# that has a indicator.8 value of 7...

data1@data <- cbind(data1@data, indicator.8)
data1@data <- as.data.frame(data1@data)
attach(data1@data)
data.1 <- data1@data
#fix(data.1)

# A choropleth map of suspected outliers...
shades.7 = shading(c(1,3,5,7),c("white","yellow","orange","red","dark red"))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,data1@data[,24],shades.7)
title("Suspected outliers - weak to strong (yellow to dark red) evidence")
choro.legend(-2400000,2200000,shades.7,
over="exactly", between="to under",
fmt="%4.0f",title="Indicator sum (max.: 7)",cex=0.8)
map.scale(1800000,650000,1000000,"x1000 km",4,0.5)
north.arrow(1800000,1400000,80000, col="blue")
#text(1500000,2400000, "NUTS level 2", cex=2, col=3)
text(1500000,2400000, "NUTS level 1", cex=2, col=3)
#text(1500000,2400000, "NUTS level 0", cex=2, col=3)

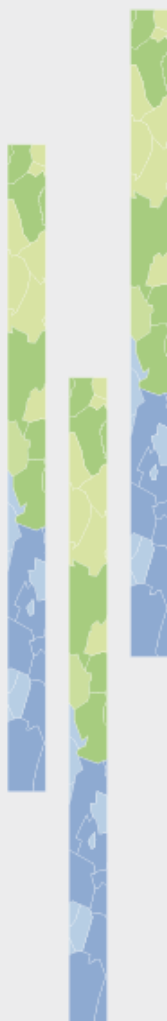
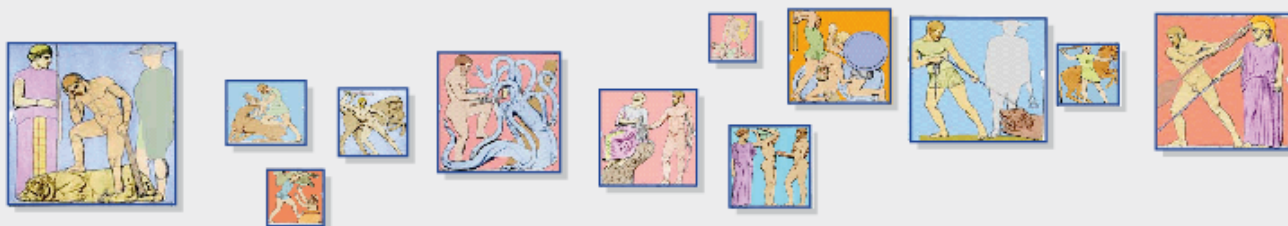
# 7. The effects of MAUP #####

# This relationship would be expected to weaken from NUTS2 to NUTS0
X11(width=5.3,height=5.7)
plot(jitter(NUTS3_outlier_max,factor=0.4), jitter(indicator.8,factor=0.4),
main="MAUP and its impact on outlier detection",
xlab="Strongest indication of an outlier for any constituent NUTS3 region",
#ylab="Indication of an outlier in a corres. aggreg. NUTS2 region", ylim=c(0,7),

```

```
ylab="Indication of an outlier in a corres. aggreg. NUTS1 region", ylim=c(0,7),  
#ylab="Indication of an outlier in a corres. aggreg. NUTS0 region", ylim=c(0,7),  
pch=19, cex=0.5)  
abline(0,1)  
cor(NUTS3_outlier_max, indicator.8)
```





## Spatial analysis for quality control

**Phase 1:**  
The identification of logical input errors  
and statistical outliers

**Draft: 2**

### MAIN RESULTS

- Exceptional values can arise from logical input errors and true outlying data.
- The accurate identification of an exceptional value is important as input errors should be treated differently to true outlying data.
- Input errors can usually be identified mathematically or sometimes, statistically. Outliers are identified statistically.
- A weighted evidence approach to statistically identify outliers is presented using a worked example that has been coded with R open source software.

**ESPON 2013 DATABASE**

**MARCH 2011**



## List of authors

Paul Harris, NCG - NUIM

Martin Charlton, NCG - NUIM

Stewart Fotheringham, NCG - NUIM

### Contact

[Paul.Harris@nuim.ie](mailto:Paul.Harris@nuim.ie)

[Martin.Charlton@nuim.ie](mailto:Martin.Charlton@nuim.ie)

[Stewart.Fotheringham@nuim.ie](mailto:Stewart.Fotheringham@nuim.ie)

National Centre for Geocomputation (NCG),

National University of Ireland (Maynooth)

Tel. (+353) 1 708 6455

# **1 Introduction and aims**

The ESPON 2013 Database should be as free from errors as possible. It follows from this that detecting errors is an important activity in both data entry and data checking. This technical report (which accompanies the final report) aims to clarify the results in the Second Interim Report (SIR) on how mathematical, statistical and spatial analysis tools can be applied to the ESPON 2013 Database in order to find 'logical input errors' (stage 1 detection) and 'statistical outliers' (stage 2 detection).

We first respond to queries or criticisms raised by the ESPON management on the SIR. Secondly (and in some respects, a response to criticisms), we propose a novel approach to the detection of statistical outliers (stage 2 detection), which we have developed since the SIR. As demonstration of this 'weight of evidence' approach, it is applied to a real (space-time) ESPON data set. The case study is presented with an R script so that the 'weight of evidence' approach is reproducible.

## 2 Responses to queries or criticisms

We shall simply respond to each point made in turn.

*"Please also deliver the R-Scripts as a file."* **Response** - This was already done, and the R script for this technical report has been similarly delivered.

*"Chapter 4, however, lacks a comparison between the various techniques. Which techniques have shown the best results for these worked examples and are therefore the most promising?"* **Response** - For statistical outliers, this is a very difficult task to do with any objectivity or certainty. Outliers in the ESPON database can arise for many different reasons and as such, require different approaches for their detection. Furthermore, the calibration of an outlier detection technique commonly involves many judged, user-specified model decisions, where in the end, it is common to refer to an unusual observation as simply *potentially* outlying (i.e. statistical detection cannot usually say that an observation is outlying with certainty). For these reasons and more, we have adopted a 'weight of evidence' approach to outlier detection, using many of the SIR techniques. The approach is still in its infancy and future work should consider various refinements (see [section 3](#) for details). For the detection of logical input errors (stage 1 detection), we recommend that their methods of detection are simply updated as they arise. It is impossible to foresee the infinite number of ways that a logical input error can occur. We have provided a start to this issue in the SIR and here, the metadata for each variable should provide a helpful basis for this detection.

*"For ESPON it is also important to know what exactly the benefit is for ESPON of the research done on improving these techniques and procedures. The benefit in general is clear: the detection of outliers and errors can be done easier, faster and better. However, the report does not show (yet) what has been detected in the datasets that are/will be incorporated in the ESPON 2013 Database. The usefulness of this research for ESPON should be made more clear."* **Response** - We demonstrate our 'weight of evidence' approach to a real case study dataset in [section 3](#), where observations are flagged as outlying together with the nature of their 'outlyingness' (i.e. spatial, temporal, etc.). The approach can easily be embedded within the database architecture. Many of the detection techniques from the SIR have already been incorporated in this respect, some of which are key components of the more coherent 'weight of evidence' approach.

*"Furthermore, detecting is one step, solving the problem is the next. Are the outliers and errors detected real errors? Can they be solved? And if so, how are they solved. It would be good if this could be shown in the next version of the report."* **Response** - Observations can only be flagged as *potentially* outlying, but using a 'weight of evidence' approach allows some observations to have a higher *potential* in this respect, than others. Observations that are flagged as outlying from many detection techniques should be viewed with the strongest suspicion. The decision of what to do with an outlier (e.g. remove, replace, etc.) should ultimately reside with some expert on the given data set (or its provider). Mechanisms can be put in place within the database architecture to do this in an efficient and effective manner.

### 3 Case study: the 'weight of evidence' approach

In this case study, we apply our 'weight of evidence' approach to detect outliers in an ESPON space-time, *univariate* data set (where an extension to *multivariate* case is simple and direct). This approach applies nine representative and complementary statistical outlier detection techniques, where observations are flagged as outlying according to the outcome of each technique. This then builds up a 'weight of evidence' for the likelihood of a given observation being outlying (i.e. statistical evidence is strongest for an observation that is flagged as outlying for all nine techniques).

According to our chosen techniques, we can indicate the nature of the outliers. Outliers can have any combination of aspatial, spatial, temporal or relational characteristics (see Fig. 1). Furthermore, this methodology can be specified such that a different weighting is attached to each technique. For example, we may want to attach more weight to observations that are outlying in a temporal sense. The detection techniques are all non-robust in form. However robust forms are preferred, as outliers can compromise the calibration of a given technique prior to its use as a method of detection. Extensions to robust forms, for each technique, should be the first set of refinements that are considered with this approach.

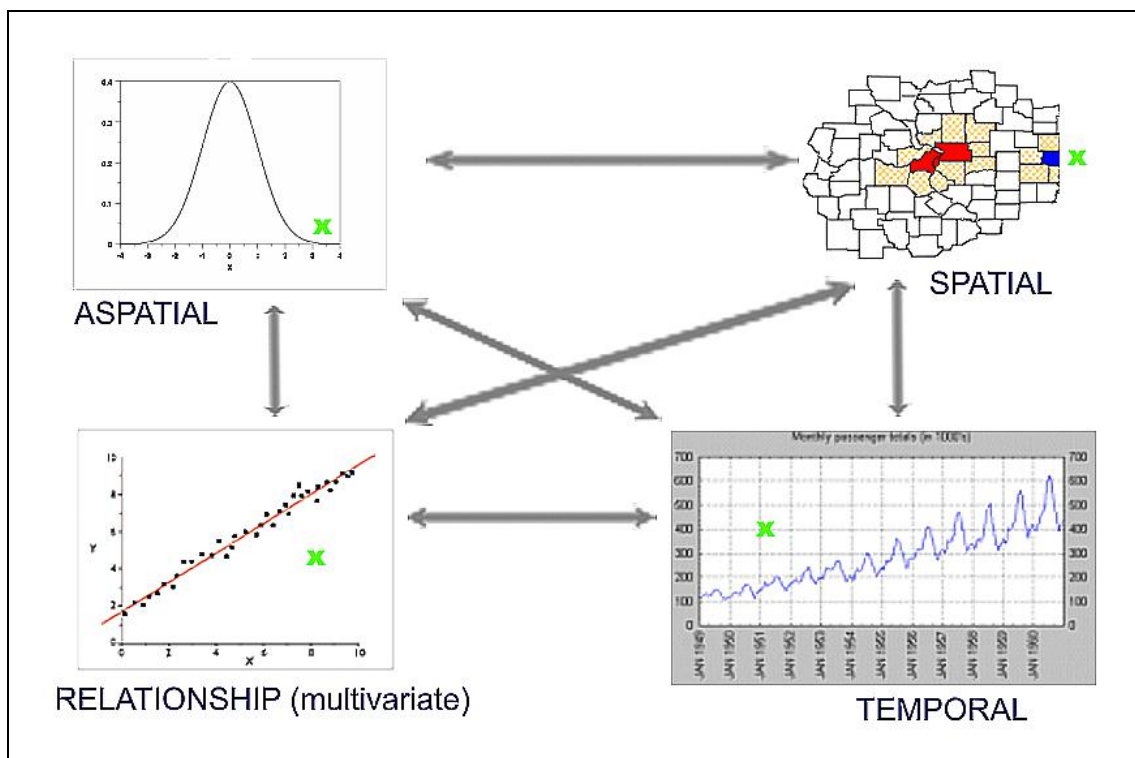


Figure 1 Types of outliers detected.

The nine techniques are: (i) boxplot statistics (for aspatial, univariate outliers); (ii) Hawkins' statistic (for spatial, univariate outliers); (iii) time series statistics (for temporal, univariate outliers); (iv) regression (for outliers from an aspatial, linear relationship); (v) locally weighted regression (for outliers from an aspatial, nonlinear relationship); (vi) geographically weighted regression (for outliers from a spatial, nonlinear relationship); (vii) principal component analysis (PCA, for

outliers from a 'model-free', aspatial, linear relationship); (viii) locally weighted PCA (for outliers from a 'model-free', aspatial, nonlinear relationship); and (ix) geographically weighted PCA (for outliers from a 'model-free', spatial, nonlinear relationship). Here the PCA methods do not require domains-specific knowledge (i.e. how to choose covariates in a regression) and as such, are termed 'model-free'.

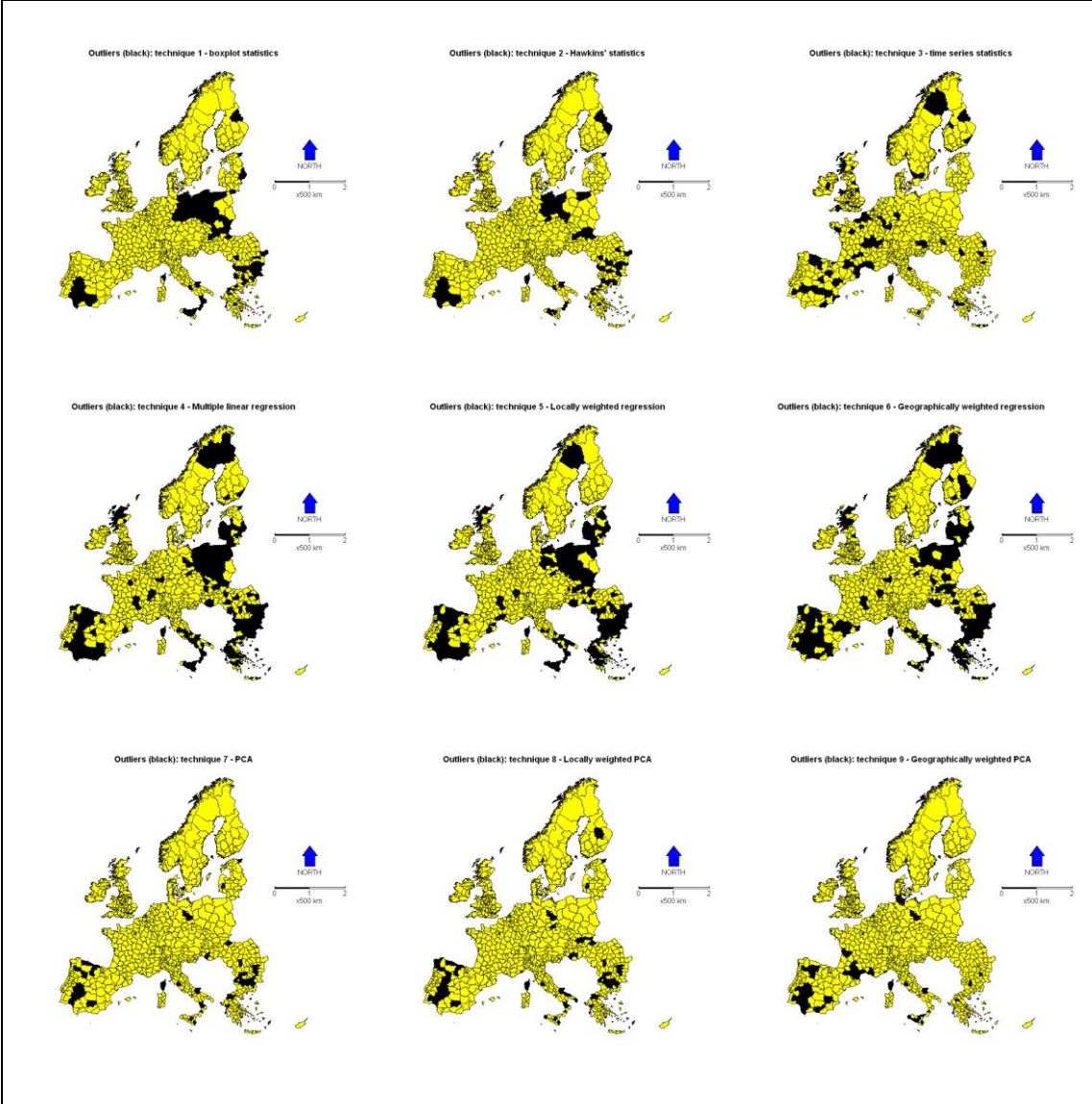


Figure 2 Outliers for each of nine 'weight of evidence' techniques.

We demonstrate our 'weight of evidence' outlier detection approach to ESPON unemployment rate data covering 709 NUTS23 regions for eight consecutive years 2000 to 2007. Here we show that outlier detection can not only be used as a data cleaning or screening exercise, but can also be used to uncover interesting or unusual relationships in the process that has not been considered before. Observe also that the regression methods can also be used for the prediction of missing data.

Full details of the outlier detection approach are given in the associated R script (see Appendix). Fig. 2 presents the results (maps) for each of the nine detection techniques and Fig.3 presents the summary of the results (i.e. the final 'weight of evidence' map). Thus as examples, there is strong evidence of at least one outlying unemployment rate (from the eight years) in a NUTS23 region in SW Spain and a NUTS23 region in N Corsica. Conversely, there is little evidence of at least one outlying unemployment rate (from the eight years) in a NUTS23 region in SW Ireland and all NUTS23 regions in Norway.

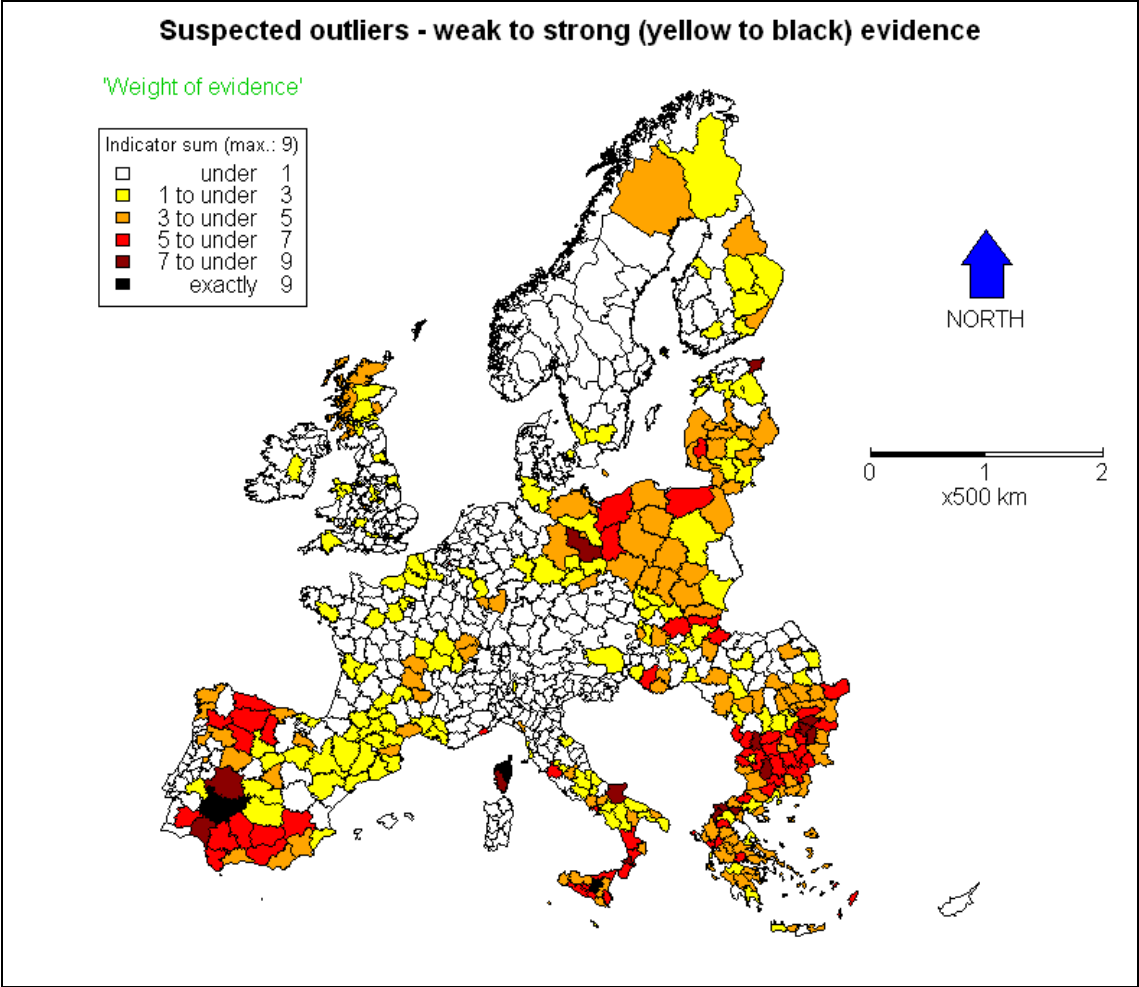


Figure 3 Final 'weight of evidence' map.

**Reading**

Fotheringham AS, Brunson C, Charlton ME (2002) Geographically Weighted Regression - the analysis of spatially varying relationships. Wiley, Chichester.

Ihaka R, Gentleman R (1996) R: A language for data analysis and graphics. Journal of Computational and Graphical Statistics 5, 299-314.

Jolliffe IT (2002) Principal Components Analysis, 2<sup>nd</sup> edition, New York: Springer-Verlag

Loader C (2004) Smoothing: Local Regression Techniques. In Gentle J, Härdle W, Mori Y (eds) Handbook of Computational Statistics. Springer-Verlag, Heidelberg.

## Appendix: Worked example of the 'weight of evidence' approach

```
# This program is free software: you can redistribute it and/or modify
# it under the terms of the GNU General Public License as published by
# the Free Software Foundation, either version 3 of the License, or
# (at your option) any later version.

# This program is distributed in the hope that it will be useful,
# but WITHOUT ANY WARRANTY; without even the implied warranty of
# MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the
# GNU General Public License for more details.

# You should have received a copy of the GNU General Public License
# along with this program. If not, see <http://www.gnu.org/licenses/>.

# Challenge 10 - ESPON 2013 database

# P. Harris (NCG), M. Charlton (NCG) & C. Brunsdon (University of Leicester, UK)

# 24/10/10

# 'Weight of evidence' approach for detecting statistical outliers.

# There is no single 'best' outlier detection technique, so our aim is to:
# a. Apply a representative selection of outlier detection techniques.
# b. Flag an observation if it is a likely outlier according to each technique.
# c. Build up a weight of evidence for the likelihood of a given observation being statistically outlying.
# d. Suggest what type of outlier it is likely to be
#     - aspatial, spatial, temporal, relationship, some mixture...
# e. Consult an expert on the data to decide on the appropriate course of action.

# In this case, we employ 9 complementary and representative outlier detection techniques:

# 1. Boxplots
# 2. Hawkins' spatial statistic
# 3. Time series statistics
# 4. MLR residuals
# 5. LWR residuals
# 6. GWR residuals
# 7. PCA 'residuals'
# 8. LWPCA 'residuals'
# 9. GWPCA 'residuals'

# Techniques 1,2,4,5,6 and 7 were all investigated in the SIR.
```



```
# This script is for the weight of evidence approach in its most basic form.
# All outlier detection techniques are non-robust in design.
# Further work should replace each technique with a robust version
# to give a robust weight of evidence approach.

# Currently robust forms are readily available in R for:

# 1. Boxplots - robust form - adjusted boxplot - robustbase package
# 2. Hawkins' spatial statistic - robust form needs development/coding
# 3. Time series statistics - robust form needs development/coding
# 4. MLR residuals - robust form - many choices - e.g. robustbase package
# 5. LWR residuals - robust form - many choices - locfit package
# 6. GWR residuals - robust form needs development/coding
# 7. PCA 'residuals' - robust form - many choices - e.g. mvoutlier package
# 8. LWPCA 'residuals' - robust form needs development/coding
# 9. GWPCA 'residuals' - robust form needs development/coding

# This particular script is also for the univariate space-time case but
# techniques are directly extendable to the multivariate space-time case.

# In 7 cases - outliers are chosen according to standard boxplot statistics.
# We could have specified adjusted boxplots statistics here instead
# but this can be part of any robust weight of evidence approach (future work, see above).

# Furthermore, in all cases - statistical tests could be used instead of boxplots.
# As it stands - this is only done in 2 cases.

# This weight of evidence approach can be performance tested using data that have been infected with outliers.
# However results can be difficult to interpret since:
# 1. It is difficult to guarantee that the infections actually produce outliers.
# 2. The data may already contain outliers.

# Further work should explore the use of simulated data sets
# or more attention should be paid to the theoretical properties of each outlier detection technique.

# Currently - this basic weight of evidence approach gives equal weights to each of the 9 detection techniques.
# There is no reason why these weights cannot vary...

# For example, these weights may work better:
# 1. Boxplots - 1
# 2. Hawkins - 1
```

```
# 3. Time series - 1
# 4. MLR - 1/3
# 5. LWR - 1/3
# 6. GWR - 1/3
# 7. PCA - 1/3
# 8. LWPCA - 1/3
# 9. GWPCA - 1/3
```

```
# The case study data set is unemployment rate (= unemployment population/active population)
# at the NUTS 23 level covering 8 years: 2000-2007.
# Thus there are 8 variables at each of 790 regions = 6320 observations in total.
```

```
# We assume that the data has been checked for logical input errors (i.e. stage 1 detection is done).
```

```
# For brevity, we also assume that we only need at least one of eight time-specific unemployment rates
# in a region to be outlying.
# However in all techniques, except PCA/LWPCA/GWPCA, we actually identify outliers by year.
```

```
# An earlier version of output from this basic script (with clean and infected data) can be found at:
# http://www.espon.eu/main/Menu\_Events/Menu\_OpenSeminars/openseminar100609After.html
# This is the presentation given at the Madrid open seminar 2010.
```

```
# Observe that outlier detection can reveal the most interesting results
# i.e. - those not expected.
```

```
# Observe also that the regression outlier detection methods can be used for the prediction of missing data too.
```

```
# R packages needed.....
# 1. GISTools (version 0.5-4) and all its dependencies (see SIR)
# 2. locfit (version 1.5-4)- for LWR
```

```
#####
# A. Functions needed for PCA, LWPCA & GWPCA #####
#####
```

```
# Author: Chris Brunson (plus some edits by Paul Harris)
```

```
# Details can be found in forthcoming articles
```

```
# Sub-functions....
```

```
wpca <- function(x,wt,...) {
```

```

local.center <- function(x,wt)
  sweep(x,2,colSums(sweep(x,1,wt,'*'))/sum(wt))
  svd(sweep(local.center(x,wt),1,sqrt(wt),'*'),...)}

bw.by.nn <- function(loc,nn,eloc=loc) {
  result <- numeric(nrow(eloc))
  int.part <- floor(nn)
  frac.part <- nn - int.part
  if (frac.part == 0) {
    for (i in 1:nrow(eloc)) {
      d.sqr <- rowSums(sweep(loc,2,eloc[i,])**2)
      result[i] <- sort(d.sqr,partial=nn)[nn]}
  } else {
    for (i in 1:nrow(eloc)) {
      d.sqr <- rowSums(sweep(loc,2,eloc[i,])**2)
      d.tmp <- sort(d.sqr,partial=c(nn,nn+1))[c(nn,nn+1)]
      result[i] <- d.tmp[1]*(1-frac.part) + d.tmp[2]*frac.part }
  }
  sqrt(result)}

bw.by.nn.1 <- function(x,nn,ex=x) {
  result <- numeric(nrow(ex))
  int.part <- floor(nn)
  frac.part <- nn - int.part
  if (frac.part == 0) {
    for (i in 1:nrow(ex)) {
      d.sqr <- rowSums(sweep(x,2,ex[i,])**2)
      result[i] <- sort(d.sqr,partial=nn)[nn]}
  } else {
    for (i in 1:nrow(ex)) {
      d.sqr <- rowSums(sweep(x,2,ex[i,])**2)
      d.tmp <- sort(d.sqr,partial=c(nn,nn+1))[c(nn,nn+1)]
      result[i] <- d.tmp[1]*(1-frac.part) + d.tmp[2]*frac.part }
  }
  sqrt(result)}

# Basic functions....

lwpca <- function(x,bw,k=2,ex=x,pcafun=wpca,...) {
  n <- nrow(ex)
  m <- ncol(x)
  w <- array(0,c(n,m,k))
  d <- matrix(0,n,k)
  bw <- bw*bw
  for (i in 1:n) {
    wt <- rowSums(sweep(x,2,ex[i,])**2)/bw
    use <- wt < 1
    if (sum(use) < 5)
      stop(paste('You will need a larger bandwidth: location ',i,
        ' has only ',sum(use),'neighbours'))
  }
}

```

```

        wt <- (1 - wt[use])**2
        temp <- pcafun(x[use,],wt,nu=0,nv=k)
        w[i,] <- temp$v
        d[i,] <- temp$d[1:k]}
if (!is.null(rownames(x))) dimnames(w)[[1]] <- rownames(x)
if (!is.null(colnames(x))) dimnames(w)[[2]] <- colnames(x)
dimnames(w)[[3]] <- paste("PC",1:k,sep=")
list(loadings=w,var=d,bw=sqrt(bw))}

gwpc <- function(x,loc,bw,k=2,eloc=loc,pcafun=wpca,...) {
  n <- nrow(eloc)
  m <- ncol(x)
  w <- array(0,c(n,m,k))
  d <- matrix(0,n,k)
  bw <- bw*bw
  for (i in 1:n) {
    wt <- rowSums(sweep(loc,2,eloc[i,])**2)/bw
    use <- wt < 1
    if (sum(use) < 5)
      stop(paste("You will need a larger bandwidth: location 'i,
        ' has only ',sum(use),'neighbours'))
    wt <- (1 - wt[use])**2
    temp <- pcafun(x[use,],wt,nu=0,nv=k)
    w[i,] <- temp$v
    d[i,] <- temp$d[1:k]}
if (!is.null(rownames(x))) dimnames(w)[[1]] <- rownames(x)
if (!is.null(colnames(x))) dimnames(w)[[2]] <- colnames(x)
dimnames(w)[[3]] <- paste("PC",1:k,sep=")
list(loadings=w,var=d,bw=sqrt(bw))}

# Automatic bandwidth functions...

lwpc.autobw <- function(x,bw.interval,k=2,ex=x,verbose=FALSE,...) {
  bw <- optimise(function(z) lwpc.cv(x,z,k=k,...),bw.interval)
  if (verbose) {
    cat("Optimal' bandwidth = ",bw$minimum)
    cat(" Score is ",bw$objective)}
  lwpc(x,bw$minimum,k=k,ex=ex,...)}

lwpc.autobw.by.nn <- function(x,nn.interval,k=2,ex=x,verbose=FALSE,...) {
  bw <- optimise(function(z) lwpc.cv(x,bw.by.nn.l(x,z),k=k,...),nn.interval)
  if (verbose) {
    cat("Optimal' nn = ",bw$minimum)
    cat(" Score is ",bw$objective)}
  lwpc(x,bw.by.nn(x,bw$minimum),k=k,ex=ex,...)}

gwpc.autobw <- function(x,loc,bw.interval,k=2,eloc=loc,verbose=FALSE,...) {
  bw <- optimise(function(z) gwpc.cv(x,loc,z,k=k,...),bw.interval)

```

```

if (verbose) {
  cat("'Optimal' bandwidth = ",bw$minimum)
  cat(" Score is ",bw$objective)}
gwpca(x,loc,bw$minimum,k=k,eloc=eloc,...)}

gwpca.autobw.by.nn <- function(x,loc,nn.interval,k=2,eloc=loc,verbose=FALSE,...) {
  bw <- optimise(function(z) gwpca.cv(x,loc,bw.by.nn(loc,z),k=k,...),nn.interval)
  if (verbose) {
    cat("'Optimal' nn = ",bw$minimum)
    cat(" Score is ",bw$objective)}
  gwpca(x,loc,bw.by.nn(loc,bw$minimum),k=k,eloc=eloc,...)}

# Cross-validation functions...

lwpca.cv <- function(x,bw,k=2,...) {
  n <- nrow(x)
  m <- ncol(x)
  w <- array(0,c(n,m,k))
  bw <- bw*bw
  score <- 0
  for (i in 1:n) {
    wt <- rowSums(sweep(x,2,x[i,])**2)/bw
    wt[i] <- Inf
    use <- wt < 1
    wt <- (1 - wt[use])**2
    v <- wpca(x[use,],wt,nu=0,nv=k)$v
    v <- v %>% t(v)
    score <- score + sum((x[i,] - x[i,] %>% v)**2)
  }
  score}

gwpca.cv <- function(x,loc,bw,k=2,...) {
  n <- nrow(loc)
  m <- ncol(x)
  w <- array(0,c(n,m,k))
  bw <- bw*bw
  score <- 0
  for (i in 1:n) {
    wt <- rowSums(sweep(loc,2,loc[i,])**2)/bw
    wt[i] <- Inf
    use <- wt < 1
    wt <- (1 - wt[use])**2
    v <- wpca(x[use,],wt,nu=0,nv=k)$v
    v <- v %>% t(v)
    score <- score + sum((x[i,] - x[i,] %>% v)**2)
  }
  score}

```

```
# Outlier/residual functions...
```

```
lwpca.cv.contrib <- function(x,bw,k=2,...) {  
  n <- nrow(x)  
  m <- ncol(x)  
  w <- array(0,c(n,m,k))  
  bw <- bw*bw  
  score <- numeric(n)  
  for (i in 1:n) {  
    wt <- rowSums(sweep(x,2,x[i,])**2)/bw  
    wt[i] <- Inf  
    use <- wt < 1  
    wt <- (1 - wt[use])**2  
    v <- wpca(x[use,],wt,nu=0,nv=k)$v  
    v <- v %>% t(v)  
    score[i] <- sum((x[i,] - x[i,] %>% v))**2}  
  score}  
}
```

```
gwpca.cv.contrib <- function(x,loc,bw,k=2,...) {  
  n <- nrow(loc)  
  m <- ncol(x)  
  w <- array(0,c(n,m,k))  
  bw <- bw*bw  
  score <- numeric(n)  
  for (i in 1:n) {  
    wt <- rowSums(sweep(loc,2,loc[i,])**2)/bw  
    wt[i] <- Inf  
    use <- wt < 1  
    wt <- (1 - wt[use])**2  
    v <- wpca(x[use,],wt,nu=0,nv=k)$v  
    v <- v %>% t(v)  
    score[i] <- sum((x[i,] - x[i,] %>% v))**2}  
  score}  
}
```

```
#####  
# B. Data input and exploratory data analysis #####  
#####
```

```
# Read in a data
```

```
data0 <- read.table("Worked example data.csv", ",", header=T)
```

```
colnames(data0) # Note the first 8 columns relate to an experiment with infected data - ignore this
```

```
attach(data0)
```

```
# Variables of interest are:
```

```
# UNEMP.R1.2000, UNEMP.R1.2001, UNEMP.R1.2002, UNEMP.R1.2003, UNEMP.R1.2004, UNEMP.R1.2005,
```

```
# UNEMP.R1.2006 and UNEMP.R1.2007
```

```
# Distributions...
```

```
X11(width=12,height=7)
par(mfrow=c(2,4))
hist(UNEMP.R1.2000, freq=F)
hist(UNEMP.R1.2001, freq=F)
hist(UNEMP.R1.2002, freq=F)
hist(UNEMP.R1.2003, freq=F)
hist(UNEMP.R1.2004, freq=F)
hist(UNEMP.R1.2005, freq=F)
hist(UNEMP.R1.2006, freq=F)
hist(UNEMP.R1.2007, freq=F)
```

```
X11(width=12,height=7)
par(mfrow=c(2,4))
hist(log(UNEMP.R1.2000), freq=F)
hist(log(UNEMP.R1.2001), freq=F)
hist(log(UNEMP.R1.2002), freq=F)
hist(log(UNEMP.R1.2003), freq=F)
hist(log(UNEMP.R1.2004), freq=F)
hist(log(UNEMP.R1.2005), freq=F)
hist(log(UNEMP.R1.2006), freq=F)
hist(log(UNEMP.R1.2007), freq=F)
```

```
# Correlations...
```

```
cor(data0[9:16])
X11(width=18,height=12)
pairs(data0[9:16],cex=0.5)
```

```
# Importing data as a ArcGIS shapefile & using GISTools to do some maps
require(GISTools)
```

```
# Ignore all warnings - this code is under development...
```

```
# Read in the shapefile...
```

```
data1 <- readShapePoly("FinalTS23.shp",
proj4string=CRS("+proj=laea+lat_0=50.0+lon_0=15.0+x_0=0+y_0=0"))
colnames(data1 @data)
```

```
# Renaming each variable - as they have been truncated in ArcGIS...
```

```
colnames(data1 @data) <- c("NUTS23x", "NUTS23Ax",
"ACTIVE_2000x", "ACTIVE_2001x", "ACTIVE_2002x", "ACTIVE_2003x", "ACTIVE_2004x", "ACTIVE_2005x", "ACTIVE_2006x",
"ACTIVE_2007x",
"UNEMP_2000x", "UNEMP_2001x", "UNEMP_2002x", "UNEMP_2003x", "UNEMP_2004x", "UNEMP_2005x", "UNEMP_2006x",
"UNEMP_2007x",
```

```

"Xx","Yx")

# Checking the two data sets match up OK...
length(data0[,1])
length(data1@data[,1])
data.x <- cbind(data1@data$NUTS23x, NUTS23, data1@data$Xx, X, data1@data$Yx, Y)

# Coordinates....
Nuts23Coords <- as.matrix(cbind(X/1000, Y/1000))
colnames(Nuts23Coords) <- c("Easting","Northing")

# Maximum distances of sampled area in main directions...
max(X/1000)-min(X/1000)
max(Y/1000)-min(Y/1000)

# Renaming data...
data1@data <- data0
attach(data1@data)

# Create another data matrix for the analysis
#Data.1.scaled <- scale(as.matrix(data1@data[9:16])) # scaled
Data.1.scaled <- as.matrix(data1@data[9:16]) # not scaled
rownames(Data.1.scaled) <- data1@data[,17]
n1 <- length(Data.1.scaled[,1])

# An example choropleth map...
shades.e = auto.shading(UNEMP.R1.2000,5, cols=brewer.pal(5,'PuBuGn'))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,UNEMP.R1.2000,shades.e)
title("Unemployment rate: year 2000")
choro.legend(1300000,400000,shades.e,fmt="%4.2f",title='Rate',cex=0.8)
map.scale(1800000,-650000,1100000,"x500 km",2,1)
north.arrow(1800000,1000000,80000, col="blue")

#####
# 1. Standard boxplot outliers #####
#####

# Need to do this for each unemployment variable (i.e. each year) ...

z1 <- data0[,9]
bp <- boxplot.stats(z1, coef=1.5)
indicator.1.1 <- ifelse(z1>bp$stats[1]& z1<bp$stats[5], 0, 1) # i.e. suspected outliers...

z1 <- data0[,10]
bp <- boxplot.stats(z1, coef=1.5)

```



```

indicator.1.2 <-ifelse(z1>bp$stats[1]& z1<bp$stats[5], 0, 1)

z1 <- data0[,11]
bp <- boxplot.stats(z1, coef=1.5)
indicator.1.3 <-ifelse(z1>bp$stats[1]& z1<bp$stats[5], 0, 1)

z1 <- data0[,12]
bp <- boxplot.stats(z1, coef=1.5)
indicator.1.4 <-ifelse(z1>bp$stats[1]& z1<bp$stats[5], 0, 1)

z1 <- data0[,13]
bp <- boxplot.stats(z1, coef=1.5)
indicator.1.5 <-ifelse(z1>bp$stats[1]& z1<bp$stats[5], 0, 1)

z1 <- data0[,14]
bp <- boxplot.stats(z1, coef=1.5)
indicator.1.6 <-ifelse(z1>bp$stats[1]& z1<bp$stats[5], 0, 1)

z1 <- data0[,15]
bp <- boxplot.stats(z1, coef=1.5)
indicator.1.7 <-ifelse(z1>bp$stats[1]& z1<bp$stats[5], 0, 1)

z1 <- data0[,16]
bp <- boxplot.stats(z1, coef=1.5)
indicator.1.8 <-ifelse(z1>bp$stats[1]& z1<bp$stats[5], 0, 1)

# This indicator is one value out of eight is an outlier at a given location
indicator.1 <- indicator.1.1+indicator.1.2+indicator.1.3+indicator.1.4+indicator.1.5+indicator.1.6+indicator.1.7+indicator.1.8

# A choropleth map of the boxplot outliers...
shades.1 = shading(c(0,1,9),c("white","yellow","black")) # i.e. yellow - no & black - yes
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,indicator.1,shades.1)
title("Outliers (black): technique 1 - boxplot statistics")
map.scale(1800000,650000,1100000,"x500 km",2,1)
north.arrow(1800000,1000000,80000, col="blue")

#####
# 2. Outliers via Hawkins' statistic - using chi-squared #####
#####

# GW means and GW variances for Hawkin's statistic found
# using the gw.cov function in spgwr with the default bi-square weighting scheme.

```

```

# Bandwidth is user-specified...
bwd.1 <- 0.1 # 10% of nearby data

# Coordinates....
coordinates(data0) <- c("X", "Y") # note do not need to divide by a 1000

# Critical values of the chi-squared distribution
chi_10 <- 2.70554
chi_5 <- 3.84146
chi_2.5 <- 5.02389
chi_1 <- 6.63490
chi_0.5 <- 7.87944
chi_0.01 <- 10.828

# Hawkins test for all 8 unemployment variables (or years)

gwss <- gw.cov(data0, vars=9, adapt=bwd.1)
GW.mean <- gwss$SDF$mean.V1
GW.variance <- (gwss$SDF$sd.V1)^2
Hawk.N <- bwd.1*length(X) # number of neighbouring data
Hawk.lm <- GW.mean # the local mean at observation points
Hawk.alv <- mean(GW.variance) # the average local variance with same bandwidth
Hawk.test <- (Hawk.N*(UNEMP.R1.2000-Hawk.lm)^2)/((Hawk.N+1)*Hawk.alv) # test statistic
indicator.2.1 <- ifelse(Hawk.test <=chi_5, 0, 1) # change critical level accordingly...

gwss <- gw.cov(data0, vars=10, adapt=bwd.1)
GW.mean <- gwss$SDF$mean.V1
GW.variance <- (gwss$SDF$sd.V1)^2
Hawk.N <- bwd.1*length(X) # number of neighbouring data
Hawk.lm <- GW.mean # the local mean at observation points
Hawk.alv <- mean(GW.variance) # the average local variance with same bandwidth
Hawk.test <- (Hawk.N*(UNEMP.R1.2001-Hawk.lm)^2)/((Hawk.N+1)*Hawk.alv) # test statistic
indicator.2.2 <- ifelse(Hawk.test <=chi_5, 0, 1) # change critical level accordingly...

gwss <- gw.cov(data0, vars=11, adapt=bwd.1)
GW.mean <- gwss$SDF$mean.V1
GW.variance <- (gwss$SDF$sd.V1)^2
Hawk.N <- bwd.1*length(X) # number of neighbouring data
Hawk.lm <- GW.mean # the local mean at observation points
Hawk.alv <- mean(GW.variance) # the average local variance with same bandwidth
Hawk.test <- (Hawk.N*(UNEMP.R1.2002-Hawk.lm)^2)/((Hawk.N+1)*Hawk.alv) # test statistic
indicator.2.3 <- ifelse(Hawk.test <=chi_5, 0, 1) # change critical level accordingly...

gwss <- gw.cov(data0, vars=12, adapt=bwd.1)
GW.mean <- gwss$SDF$mean.V1
GW.variance <- (gwss$SDF$sd.V1)^2
Hawk.N <- bwd.1*length(X) # number of neighbouring data
Hawk.lm <- GW.mean # the local mean at observation points
Hawk.alv <- mean(GW.variance) # the average local variance with same bandwidth
Hawk.test <- (Hawk.N*(UNEMP.R1.2003-Hawk.lm)^2)/((Hawk.N+1)*Hawk.alv) # test statistic

```

```

indicator.2.4 <-ifelse(Hawk.test <=chi_5, 0, 1) # change critical level accordingly...

gwss <- gw.cov(data0, vars=13, adapt=bwd.1)
GW.mean <- gwss$SDF$mean.V1
GW.variance <- (gwss$SDF$sd.V1)^2
Hawk.N <- bwd.1*length(X) # number of neighbouring data
Hawk.lm <- GW.mean # the local mean at observation points
Hawk.alv <- mean(GW.variance) # the average local variance with same bandwidth
Hawk.test <- (Hawk.N*(UNEMP.R1.2004-Hawk.lm)^2)/((Hawk.N+1)*Hawk.alv) # test statistic
indicator.2.5 <-ifelse(Hawk.test <=chi_5, 0, 1) # change critical level accordingly...

gwss <- gw.cov(data0, vars=14, adapt=bwd.1)
GW.mean <- gwss$SDF$mean.V1
GW.variance <- (gwss$SDF$sd.V1)^2
Hawk.N <- bwd.1*length(X) # number of neighbouring data
Hawk.lm <- GW.mean # the local mean at observation points
Hawk.alv <- mean(GW.variance) # the average local variance with same bandwidth
Hawk.test <- (Hawk.N*(UNEMP.R1.2005-Hawk.lm)^2)/((Hawk.N+1)*Hawk.alv) # test statistic
indicator.2.6 <-ifelse(Hawk.test <=chi_5, 0, 1) # change critical level accordingly...

gwss <- gw.cov(data0, vars=15, adapt=bwd.1)
GW.mean <- gwss$SDF$mean.V1
GW.variance <- (gwss$SDF$sd.V1)^2
Hawk.N <- bwd.1*length(X) # number of neighbouring data
Hawk.lm <- GW.mean # the local mean at observation points
Hawk.alv <- mean(GW.variance) # the average local variance with same bandwidth
Hawk.test <- (Hawk.N*(UNEMP.R1.2006-Hawk.lm)^2)/((Hawk.N+1)*Hawk.alv) # test statistic
indicator.2.7 <-ifelse(Hawk.test <=chi_5, 0, 1) # change critical level accordingly...

gwss <- gw.cov(data0, vars=16, adapt=bwd.1)
GW.mean <- gwss$SDF$mean.V1
GW.variance <- (gwss$SDF$sd.V1)^2
Hawk.N <- bwd.1*length(X) # number of neighbouring data
Hawk.lm <- GW.mean # the local mean at observation points
Hawk.alv <- mean(GW.variance) # the average local variance with same bandwidth
Hawk.test <- (Hawk.N*(UNEMP.R1.2007-Hawk.lm)^2)/((Hawk.N+1)*Hawk.alv) # test statistic
indicator.2.8 <-ifelse(Hawk.test <=chi_5, 0, 1) # change critical level accordingly...

# This indicator is one value out of eight is an outlier at a given location
indicator.2 <- indicator.2.1+indicator.2.2+indicator.2.3+indicator.2.4+indicator.2.5+indicator.2.6+indicator.2.7+indicator.2.8

# A choropleth map of Hawkins' test results...
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,indicator.2,shades.1)
title("Outliers (black): technique 2 - Hawkins' statistics")
map.scale(1800000,650000,1100000,"x500 km",2,1)
north.arrow(1800000,1000000,80000, col="blue")

```

```
#####  
# 3. Time series outliers - using chi-squared #####  
#####
```

```
# Simplified version of Claude Grasland's time series outlier identification  
# In this case for one, not multiple spatial units...
```

```
Data.1.ts <- data1@data[9:16]  
row.means <- apply(Data.1.ts,1,mean)  
row.vars <- apply(Data.1.ts,1,var)  
row.sds <- row.vars^0.5  
op.1 <- sweep(Data.1.ts,1,row.means)  
resids.ts <- (op.1)^2/row.vars  
# Note the use of variances - thus we do a chi-square test here instead of looking at boxplot statistics  
#sort(resids.3)  
quant.ts <- quantile(resids.ts, probs = seq(0, 1, 0.01))  
threshold.ts <- quant.ts[100] # in this case the 99%tile
```

```
# Assume Gaussianity and do a Chi-squared test...  
threshold.tsx <- 3.84146 # chi-square test at 95% ...  
indicator.tsx <- ifelse(resids.ts>0& resids.ts<threshold.tsx, 0, 1) # i.e. suspected outliers...  
indicator.3 <- apply(indicator.tsx,1,sum)
```

```
# A choropleth map of the time series outliers  
shades.2 = shading(c(0,1,2),c("white","yellow","black")) # i.e. yellow - no & black - yes  
X11(width=8,height=7)  
par(mar=c(0,0,2,0))  
choropleth(data1,indicator.3,shades.2)  
title("Outliers (black): technique 3 - time series statistics")  
map.scale(1800000,650000,1100000,"x500 km",2,1)  
north.arrow(1800000,1000000,80000, col="blue")
```

```
#####  
# 4. Outliers via MLR - via boxplots #####  
#####
```

```
# In all regressions, we use unemployment from the two closest years to explain unemployment for a given year of interest.
```

```
# Note using Boxplot statistics for residuals
```

```
# Need to do this eight times...
```

```
# one for each unemployment variable & changing the independent variables accordingly
```

```

mlr <- lm(UNEMP.R1.2000 ~ UNEMP.R1.2001+UNEMP.R1.2002)
raw.resids.mlr <- UNEMP.R1.2000-mlr$fitted # Actual minus prediction (raw residuals)
# Identifying & updating outlier information in one file using Boxplot statistics for residuals
bp <- boxplot.stats(raw.resids.mlr, coef=1.5)
indicator.4.1 <- ifelse(raw.resids.mlr>bp$stats[1]& raw.resids.mlr<bp$stats[5], 0, 1)

mlr <- lm(UNEMP.R1.2001 ~ UNEMP.R1.2000+UNEMP.R1.2002)
raw.resids.mlr <- UNEMP.R1.2001-mlr$fitted
bp <- boxplot.stats(raw.resids.mlr, coef=1.5)
indicator.4.2 <- ifelse(raw.resids.mlr>bp$stats[1]& raw.resids.mlr<bp$stats[5], 0, 1)

mlr <- lm(UNEMP.R1.2002 ~ UNEMP.R1.2001+UNEMP.R1.2003)
raw.resids.mlr <- UNEMP.R1.2002-mlr$fitted
bp <- boxplot.stats(raw.resids.mlr, coef=1.5)
indicator.4.3 <- ifelse(raw.resids.mlr>bp$stats[1]& raw.resids.mlr<bp$stats[5], 0, 1)

mlr <- lm(UNEMP.R1.2003 ~ UNEMP.R1.2002+UNEMP.R1.2004)
raw.resids.mlr <- UNEMP.R1.2003-mlr$fitted
bp <- boxplot.stats(raw.resids.mlr, coef=1.5)
indicator.4.4 <- ifelse(raw.resids.mlr>bp$stats[1]& raw.resids.mlr<bp$stats[5], 0, 1)

mlr <- lm(UNEMP.R1.2004 ~ UNEMP.R1.2003+UNEMP.R1.2005)
raw.resids.mlr <- UNEMP.R1.2004-mlr$fitted
bp <- boxplot.stats(raw.resids.mlr, coef=1.5)
indicator.4.5 <- ifelse(raw.resids.mlr>bp$stats[1]& raw.resids.mlr<bp$stats[5], 0, 1)

mlr <- lm(UNEMP.R1.2005 ~ UNEMP.R1.2004+UNEMP.R1.2006)
raw.resids.mlr <- UNEMP.R1.2005-mlr$fitted
bp <- boxplot.stats(raw.resids.mlr, coef=1.5)
indicator.4.6 <- ifelse(raw.resids.mlr>bp$stats[1]& raw.resids.mlr<bp$stats[5], 0, 1)

mlr <- lm(UNEMP.R1.2006 ~ UNEMP.R1.2005+UNEMP.R1.2007)
raw.resids.mlr <- UNEMP.R1.2006-mlr$fitted
bp <- boxplot.stats(raw.resids.mlr, coef=1.5)
indicator.4.7 <- ifelse(raw.resids.mlr>bp$stats[1]& raw.resids.mlr<bp$stats[5], 0, 1)

mlr <- lm(UNEMP.R1.2007 ~ UNEMP.R1.2005+UNEMP.R1.2006)
raw.resids.mlr <- UNEMP.R1.2007-mlr$fitted
bp <- boxplot.stats(raw.resids.mlr, coef=1.5)
indicator.4.8 <- ifelse(raw.resids.mlr>bp$stats[1]& raw.resids.mlr<bp$stats[5], 0, 1)

# This indicator is one value out of eight is an outlier at a given location
indicator.4 <- indicator.4.1+indicator.4.2+indicator.4.3+indicator.4.4+indicator.4.5+indicator.4.6+indicator.4.7+indicator.4.8

# A choropleth map of MLR outliers...
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,indicator.4,shades.1)
title("Outliers (black): technique 4 - Multiple linear regression")
map.scale(1800000,650000,1100000,"x500 km",2,1)

```

```

north.arrow(180000,100000,80000, col="blue")

#####
# 5. Outliers via LWR - via boxplots #####
#####

# Using locfit...
require(locfit)
# Ignore warning message...

# In this case - we only use the closest year to the dependent variable, as too many variables can create problems...
# Automatic bandwidth selection also was problematic in some cases...

# Automatic bandwidth for a non-robust LR (i.e. not a lowess fit) using generalised cross-validation approach.
#summary(gcvplot(UNEMP.R1.2000 ~ UNEMP.R1.2001+UNEMP.R1.2002,data=data0, scale=F,alpha=seq(0.1,1,by=0.1),
#deg=1,kern="tricube",lproc=locfit.raw))

# As such - the same user-specified bandwidth is used for all eight fits
bwd.2 <- 0.4

# The eight LWR model fits...

lwr <- locfit(UNEMP.R1.2000 ~ UNEMP.R1.2001,data=data0, scale=F, alpha=bwd.2,
deg=1,kern="tricube",lproc=locfit.raw)
lwr.p <- fitted.locfit(lwr)
raw.resids.lr <- UNEMP.R1.2000-lwr.p # Actual minus prediction
bp <- boxplot.stats(raw.resids.lr, coef=1.5)
indicator.5.1 <-ifelse(raw.resids.lr>bp$stats[1]& raw.resids.lr<bp$stats[5], 0, 1)

lwr <- locfit(UNEMP.R1.2001 ~ UNEMP.R1.2002,data=data0, scale=F, alpha=bwd.2,
deg=1,kern="tricube",lproc=locfit.raw)
lwr.p <- fitted.locfit(lwr)
raw.resids.lr <- UNEMP.R1.2001-lwr.p # Actual minus prediction
bp <- boxplot.stats(raw.resids.lr, coef=1.5)
indicator.5.2 <-ifelse(raw.resids.lr>bp$stats[1]& raw.resids.lr<bp$stats[5], 0, 1)

lwr <- locfit(UNEMP.R1.2002 ~ UNEMP.R1.2003,data=data0, scale=F, alpha=bwd.2,
deg=1,kern="tricube",lproc=locfit.raw)
lwr.p <- fitted.locfit(lwr)
raw.resids.lr <- UNEMP.R1.2002-lwr.p # Actual minus prediction
bp <- boxplot.stats(raw.resids.lr, coef=1.5)
indicator.5.3 <-ifelse(raw.resids.lr>bp$stats[1]& raw.resids.lr<bp$stats[5], 0, 1)

lwr <- locfit(UNEMP.R1.2003 ~ UNEMP.R1.2004,data=data0, scale=F, alpha=bwd.2,
deg=1,kern="tricube",lproc=locfit.raw)
lwr.p <- fitted.locfit(lwr)
raw.resids.lr <- UNEMP.R1.2003-lwr.p # Actual minus prediction

```

```

bp <- boxplot.stats(raw.resids.lr, coef=1.5)
indicator.5.4 <- ifelse(raw.resids.lr > bp$stats[1] & raw.resids.lr < bp$stats[5], 0, 1)

lwr <- locfit(UNEMP.R1.2004 ~ UNEMP.R1.2005, data=data0, scale=F, alpha=bwd.2,
deg=1, kern="tricube", Ifproc=locfit.raw)
lwr.p <- fitted.locfit(lwr)
raw.resids.lr <- UNEMP.R1.2004 - lwr.p # Actual minus prediction
bp <- boxplot.stats(raw.resids.lr, coef=1.5)
indicator.5.5 <- ifelse(raw.resids.lr > bp$stats[1] & raw.resids.lr < bp$stats[5], 0, 1)

lwr <- locfit(UNEMP.R1.2005 ~ UNEMP.R1.2006, data=data0, scale=F, alpha=bwd.2,
deg=1, kern="tricube", Ifproc=locfit.raw)
lwr.p <- fitted.locfit(lwr)
raw.resids.lr <- UNEMP.R1.2005 - lwr.p # Actual minus prediction
bp <- boxplot.stats(raw.resids.lr, coef=1.5)
indicator.5.6 <- ifelse(raw.resids.lr > bp$stats[1] & raw.resids.lr < bp$stats[5], 0, 1)

lwr <- locfit(UNEMP.R1.2006 ~ UNEMP.R1.2007, data=data0, scale=F, alpha=bwd.2,
deg=1, kern="tricube", Ifproc=locfit.raw)
lwr.p <- fitted.locfit(lwr)
raw.resids.lr <- UNEMP.R1.2006 - lwr.p # Actual minus prediction
bp <- boxplot.stats(raw.resids.lr, coef=1.5)
indicator.5.7 <- ifelse(raw.resids.lr > bp$stats[1] & raw.resids.lr < bp$stats[5], 0, 1)

lwr <- locfit(UNEMP.R1.2007 ~ UNEMP.R1.2006, data=data0, scale=F, alpha=bwd.2,
deg=1, kern="tricube", Ifproc=locfit.raw)
lwr.p <- fitted.locfit(lwr)
raw.resids.lr <- UNEMP.R1.2007 - lwr.p # Actual minus prediction
bp <- boxplot.stats(raw.resids.lr, coef=1.5)
indicator.5.8 <- ifelse(raw.resids.lr > bp$stats[1] & raw.resids.lr < bp$stats[5], 0, 1)

# This indicator is one value out of eight is an outlier at a given location
indicator.5 <- indicator.5.1 + indicator.5.2 + indicator.5.3 + indicator.5.4 + indicator.5.5 + indicator.5.6 + indicator.5.7 + indicator.5.8

# A choropleth map of the LWR outliers...
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1, indicator.5, shades.1)
title("Outliers (black): technique 5 - Locally weighted regression")
map.scale(1800000, 650000, 1100000, "x500 km", 2, 1)
north.arrow(1800000, 1000000, 80000, col="blue")

#####
# 6. Outliers via GWR - via boxplots #####
#####

# Using spgwr...

```

```
# As with MLR - we again use unemployment from the two closest years to explain unemployment for a given year of interest.  
# In this case - all eight GWR fits have an optimally found bandwidth (by cross-validation)
```

```
# The eight GWR model fits...
```

```
gwr.cv.bwd <- gwr.sel(UNEMP.R1.2000 ~ UNEMP.R1.2001+UNEMP.R1.2002,  
data=data0,adapt=TRUE,gweight=gwr.bisquare, method="cv")  
gwr.p <- gwr(UNEMP.R1.2000 ~ UNEMP.R1.2001+UNEMP.R1.2002,data=data0,adapt=gwr.cv.bwd[1],  
gweight=gwr.bisquare,predictions=T)  
raw.resids.gwr <- UNEMP.R1.2000-gwr.p$SDF$pred  
bp <- boxplot.stats(raw.resids.gwr, coef=1.5)  
indicator.6.1 <- ifelse(raw.resids.gwr>bp$stats[1]& raw.resids.gwr<bp$stats[5], 0, 1)
```

```
gwr.cv.bwd <- gwr.sel(UNEMP.R1.2001 ~ UNEMP.R1.2000+UNEMP.R1.2002,  
data=data0,adapt=TRUE,gweight=gwr.bisquare, method="cv")  
gwr.p <- gwr(UNEMP.R1.2001 ~ UNEMP.R1.2000+UNEMP.R1.2002,data=data0,adapt=gwr.cv.bwd[1],  
gweight=gwr.bisquare,predictions=T)  
raw.resids.gwr <- UNEMP.R1.2001-gwr.p$SDF$pred  
bp <- boxplot.stats(raw.resids.gwr, coef=1.5)  
indicator.6.2 <- ifelse(raw.resids.gwr>bp$stats[1]& raw.resids.gwr<bp$stats[5], 0, 1)
```

```
gwr.cv.bwd <- gwr.sel(UNEMP.R1.2002 ~ UNEMP.R1.2001+UNEMP.R1.2003,  
data=data0,adapt=TRUE,gweight=gwr.bisquare, method="cv")  
gwr.p <- gwr(UNEMP.R1.2002 ~ UNEMP.R1.2001+UNEMP.R1.2003,data=data0,adapt=gwr.cv.bwd[1],  
gweight=gwr.bisquare,predictions=T)  
raw.resids.gwr <- UNEMP.R1.2002-gwr.p$SDF$pred  
bp <- boxplot.stats(raw.resids.gwr, coef=1.5)  
indicator.6.3 <- ifelse(raw.resids.gwr>bp$stats[1]& raw.resids.gwr<bp$stats[5], 0, 1)
```

```
gwr.cv.bwd <- gwr.sel(UNEMP.R1.2003 ~ UNEMP.R1.2002+UNEMP.R1.2004,  
data=data0,adapt=TRUE,gweight=gwr.bisquare, method="cv")  
gwr.p <- gwr(UNEMP.R1.2003 ~ UNEMP.R1.2002+UNEMP.R1.2004,data=data0,adapt=gwr.cv.bwd[1],  
gweight=gwr.bisquare,predictions=T)  
raw.resids.gwr <- UNEMP.R1.2003-gwr.p$SDF$pred  
bp <- boxplot.stats(raw.resids.gwr, coef=1.5)  
indicator.6.4 <- ifelse(raw.resids.gwr>bp$stats[1]& raw.resids.gwr<bp$stats[5], 0, 1)
```

```
gwr.cv.bwd <- gwr.sel(UNEMP.R1.2004 ~ UNEMP.R1.2003+UNEMP.R1.2005,  
data=data0,adapt=TRUE,gweight=gwr.bisquare, method="cv")  
gwr.p <- gwr(UNEMP.R1.2004 ~ UNEMP.R1.2003+UNEMP.R1.2005,data=data0,adapt=gwr.cv.bwd[1],  
gweight=gwr.bisquare,predictions=T)  
raw.resids.gwr <- UNEMP.R1.2004-gwr.p$SDF$pred  
bp <- boxplot.stats(raw.resids.gwr, coef=1.5)  
indicator.6.5 <- ifelse(raw.resids.gwr>bp$stats[1]& raw.resids.gwr<bp$stats[5], 0, 1)
```

```
gwr.cv.bwd <- gwr.sel(UNEMP.R1.2005 ~ UNEMP.R1.2004+UNEMP.R1.2006,  
data=data0,adapt=TRUE,gweight=gwr.bisquare, method="cv")  
gwr.p <- gwr(UNEMP.R1.2005 ~ UNEMP.R1.2004+UNEMP.R1.2006,data=data0,adapt=gwr.cv.bwd[1],  
gweight=gwr.bisquare,predictions=T)  
raw.resids.gwr <- UNEMP.R1.2005-gwr.p$SDF$pred
```



```

bp <- boxplot.stats(raw.resids.gwr, coef=1.5)
indicator.6.6 <- ifelse(raw.resids.gwr>bp$stats[1]& raw.resids.gwr<bp$stats[5], 0, 1)

gwr.cv.bwd <- gwr.sel(UNEMP.R1.2006 ~ UNEMP.R1.2005+UNEMP.R1.2007,
data=data0,adapt=TRUE,gweight=gwr.bisquare, method="cv")
gwr.p <- gwr(UNEMP.R1.2006 ~ UNEMP.R1.2005+UNEMP.R1.2007,data=data0,adapt=gwr.cv.bwd[1],
gweight=gwr.bisquare,predictions=T)
raw.resids.gwr <- UNEMP.R1.2006-gwr.p$SDF$pred
bp <- boxplot.stats(raw.resids.gwr, coef=1.5)
indicator.6.7 <- ifelse(raw.resids.gwr>bp$stats[1]& raw.resids.gwr<bp$stats[5], 0, 1)

gwr.cv.bwd <- gwr.sel(UNEMP.R1.2007 ~ UNEMP.R1.2005+UNEMP.R1.2006,
data=data0,adapt=TRUE,gweight=gwr.bisquare, method="cv")
gwr.p <- gwr(UNEMP.R1.2007 ~ UNEMP.R1.2005+UNEMP.R1.2006,data=data0,adapt=gwr.cv.bwd[1],
gweight=gwr.bisquare,predictions=T)
raw.resids.gwr <- UNEMP.R1.2007-gwr.p$SDF$pred
bp <- boxplot.stats(raw.resids.gwr, coef=1.5)
indicator.6.8 <- ifelse(raw.resids.gwr>bp$stats[1]& raw.resids.gwr<bp$stats[5], 0, 1)

# This indicator is one value out of eight is an outlier at a given location
indicator.6 <- indicator.6.1+indicator.6.2+indicator.6.3+indicator.6.4+indicator.6.5+indicator.6.6+indicator.6.7+indicator.6.8

# A choropleth map of GWR outliers...
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,indicator.6,shades.1)
title("Outliers (black): technique 6 - Geographically weighted regression")
map.scale(1800000,650000,1100000,"x500 km",2,1)
north.arrow(1800000,1000000,80000, col="blue")

#####
# 7. PCA outliers - via boxplots #####
#####

# For PCA, LWPCA and GWPCA we need to define k the number of components to 'keep'...
# In this case 2 or 3 seems ideal - can check this by cross-validation...
# This part of the PCA/LWPCA/GWPCA approach to outlier detection needs further work...

# Lets choose k=3...
comp.keep <- 3

# Global PCA result using the gwpcapca function with a very large fixed bandwidth...
pca.result <- gwpcapca(Data.1.scaled,Nuts23Coords,bw=10000000000000000,k=comp.keep,verbose=TRUE)

# Identify some outliers

```

```

resids.pca <- sqrt(gwpcv.cv.contrib(Data.1.scaled, Nuts23Coords, pca.result$bw,k=comp.keep))
#sort(resids.pca)
quant.pca <- quantile(resids.pca, probs = seq(0, 1, 0.01))
threshold.pca <- quant.pca[96] # in this case the 95%tile

# Standard Boxplot statistics for resids.pca
bp <- boxplot.stats(resids.pca, coef=1.5)
bp$stats
bp$stats[1]
bp$stats[5]
bp$conf
sort(bp$out)
length(bp$out)

# Identifying & updating outlier information in one file
indicator.7 <- ifelse(resids.pca>bp$stats[1]& resids.pca<bp$stats[5], 0, 1) # i.e. suspected outliers...

# A choropleth map of the PCA outliers...
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,indicator.7,shades.2)
title("Outliers (black): technique 7 - PCA")
map.scale(1800000,650000,1100000,"x500 km",2,1)
north.arrow(1800000,1000000,80000, col="blue")

#####
# 8. LWPCA outliers - via boxplots #####
#####

# Automatically finding the optimal adaptive bandwidth by cross-validation...
lwpcv.result <- lwpcv.cv.by.nn(Data.1.scaled,c(100,790),k=comp.keep,verbose=TRUE)

# Some output
#lwpcv.result$bw
#lwpcv.result$var
#lwpcv.result$loadings

# Identify some outliers
resids.lwpcv <- sqrt(lwpcv.cv.contrib(Data.1.scaled, lwpcv.result$bw,k=comp.keep))
quant.lwpcv <- quantile(resids.lwpcv, probs = seq(0, 1, 0.01))
threshold.lwpcv <- quant.lwpcv[96]

# Standard Boxplot statistics for resids.lwpcv
bp <- boxplot.stats(resids.lwpcv, coef=1.5)
bp$stats
bp$stats[1]
bp$stats[5]

```

```

bp$conf
sort(bp$out)
length(bp$out)

# Identifying & updating outlier information in one file
indicator.8 <- ifelse(resids.lwpca > bp$stats[1] & resids.lwpca < bp$stats[5], 0, 1) # i.e. suspected outliers...

# A choropleth map of the LWPCA outliers...
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,indicator.8,shades.2)
title("Outliers (black): technique 8 - Locally weighted PCA")
map.scale(1800000,650000,1100000,"x500 km",2,1)
north.arrow(1800000,1000000,80000, col="blue")

#####
# 9. GWPCA outliers - via boxplots #####
#####

# Automatically finding the optimal adaptive bandwidth by cross-validation...
gwpca.result <- gwpca.autobw.by.nn(Data.1.scaled,Nuts23Coords,c(100,790),k=comp.keep,verbose=TRUE)

# Some output
#gwpca.result$bw
#gwpca.result$svar
#gwpca.result$loadings

# Identify some outliers
resids.gwpca <- sqrt(gwpca.cv.contrib(Data.1.scaled, Nuts23Coords, gwpca.result$bw,k=comp.keep))
quant.gwpca <- quantile(resids.gwpca, probs = seq(0, 1, 0.01))
threshold.gwpca <- quant.gwpca[96]

# Standard Boxplot statistics for resids.gwpca
bp <- boxplot.stats(resids.gwpca, coef=1.5)
bp$stats
bp$stats[1]
bp$stats[5]
bp$conf
sort(bp$out)
length(bp$out)

# Identifying & updating outlier information in one file
indicator.9 <- ifelse(resids.gwpca > bp$stats[1] & resids.gwpca < bp$stats[5], 0, 1) # i.e. suspected outliers...

# A choropleth map of the GWPCA outliers...

```

```

shades.4 = shading(c(0,1,2),c("white","yellow","black")) # i.e. yellow - no & black - yes
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,indicator.9,shades.2)
title("Outliers (black): technique 9 - Geographically weighted PCA")
map.scale(1800000,650000,1100000,"x500 km",2,1)
north.arrow(1800000,1000000,80000, col="blue")

#####
# C. Linking all data together for weight of evidence map #####
#####

data.out <- cbind(indicator.1, indicator.2, indicator.3,indicator.4, indicator.5, indicator.6,
indicator.7, indicator.8, indicator.9, data1@data)

# To a text file
#write.table(data.out,"WoE outliers 1.txt", col.names=T,row.names=F)

data.out.1 <- cbind(X, Y, Order_ID, indicator.1, indicator.2, indicator.3,
indicator.4, indicator.5, indicator.6,indicator.7, indicator.8, indicator.9)

# To a text file
#write.table(data.out.1,"WoE outliers indicators 1.txt", col.names=T,row.names=F)

max(indicator.1)
max(indicator.2)
max(indicator.3)
max(indicator.4)
max(indicator.5)
max(indicator.6)
max(indicator.7)
max(indicator.8)
max(indicator.9)

indicator.1a <- indicator.1 >= 1
indicator.2a <- indicator.2 >= 1
indicator.4a <- indicator.4 >= 1
indicator.5a <- indicator.5 >= 1
indicator.6a <- indicator.6 >= 1

max(indicator.1a)
max(indicator.2a)
max(indicator.3)
max(indicator.4a)
max(indicator.5a)
max(indicator.6a)
max(indicator.7)

```

```

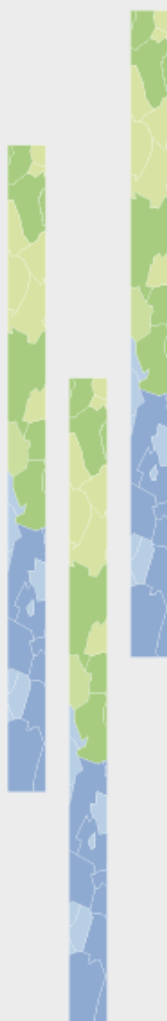
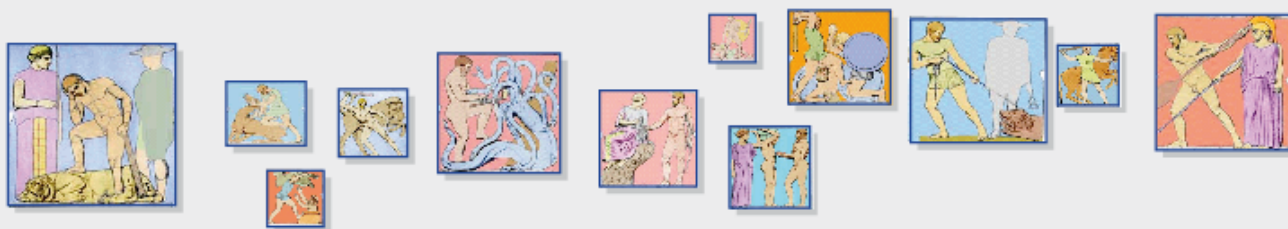
max(indicator.8)
max(indicator.9)

# Put all indicator data together...
indicator.10 <- indicator.1a+indicator.2a+indicator.3+indicator.4a+indicator.5a+indicator.6a+indicator.7+indicator.8+indicator.9

summary(indicator.10)
# Histogram
X11(width=5.3,height=5.7)
hist(indicator.10,br=c(0,1,2,3,4,5,6,7,8,9), main="Distribution of 'weight of evidence'",
xlab="Indicator sum")

# A choropleth map of suspected outliers...
shades.7 = shading(c(1,3,5,7,9),c("white","yellow","orange","red","dark red","black"))
X11(width=8,height=7)
par(mar=c(0,0,2,0))
choropleth(data1,indicator.10,shades.7)
title("Suspected outliers - weak to strong (yellow to black) evidence")
choro.legend(-2400000,2200000,shades.7,
over="exactly", between="to under",
fmt="%4.0f",title='Indicator sum (max.: 9)',cex=0.8)
map.scale(1800000,650000,1100000,"x500 km",2,1)
north.arrow(1800000,1400000,80000, col="blue")
text(-1900000,2400000, "Weight of evidence", cex=1, col=3)

```



## USING DOWNSCALED POPULATION IN LOCAL DATA GENERATION

### *A COUNTRY-LEVEL EXAMINATION*

#### CONTENT

- **Research Context and Approach.** This part outlines the background to and methodology of the examination of downscaled population data.
- **A Country-Level Examination.** This part presents the results of a Swedish examination of downscaled population, focusing on 1) population estimates in varying local settings, and 2) the estimation of overall population for UMZs of different sizes.
- **Summary and Discussion.** This part points out that while there are obvious limitations to downscaled population data, it is a quite reasonable tool for certain purposes. In particular, fairly good estimations of UMZ population can be obtained.

**ESPON 2013 DATABASE**  
**MARCH 2011**



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

16 pages

# LIST OF AUTHORS

Magnus Strömgren, Dept. of Social and Economic Geography, Umeå University

Einar Holm, Dept. of Social and Economic Geography, Umeå University

## **Contact**

[magnus.stromgren@geography.umu.se](mailto:magnus.stromgren@geography.umu.se)

[einar.holm@geography.umu.se](mailto:einar.holm@geography.umu.se)

tel. + 46 90 786 52 58

# TABLE OF CONTENTS

<b>Introduction .....</b>	<b>3</b>
<b>1 Research Context and Approach .....</b>	<b>4</b>
1.1 Methodology .....	4
<b>2 A Country-Level Examination.....</b>	<b>7</b>
2.1 Population Estimates in Varying Local Settings .....	7
2.1.1 Results by Municipality Group .....	9
2.1.2 The Residual Map .....	11
2.2 Estimations of UMZ Population.....	12
<b>3 Summary and Discussion .....</b>	<b>15</b>



# Introduction

In the ESPON db context, there is a need to utilize or present population data with a high degree of spatial resolution. For instance, in the disaggregation of socioeconomic data to grid level, detailed local population data is required for a proper downscaling of certain variables. Similarly, in reporting population figures for geographical subdivisions such as Urban Morphological Zones (UMZs), NUTS or even Local Administrative Units (LAU), level 2 population figures won't suffice. The approach that has been taken is to make use of downscaled population data—a population grid produced by the Joint Research Centre (JRC). This dataset, "Population density disaggregated with CORINE land cover 2000", distributes LAU, level 2 population data to a grid, mainly using CORINE land cover data.

However, there are a limited number of tests of the suitability of using the population grid for different purposes, as well as of its reliability in different settings. This technical report presents the results of a country-level examination of the population grid, using Swedish register population data.

# 1 Research Context and Approach

In addition to exploring the role of survey data, an important ESPON db activity for the Department of Social and Economic Geography, Umeå University is to carry out comparisons between Swedish data and data with EU coverage. The department has access to Swedish register data, which not only covers the entire population of Sweden for a substantial time period, but also has a high degree of spatial resolution. This resource makes possible a broad range of exploratory studies and evaluations.

This technical report presents the results of a country-level examination of downscaled population for the EU. In the study, Swedish register population data is used to examine the JRC population grid “Population density disaggregated with CORINE land cover 2000”. The population grid—which allocates LAU, level 2 2001 census population data to 100 m<sup>2</sup> squares, mainly using CORINE land cover data—is an important tool in the ESPON db project. First, it is part of the workflow to disaggregate socioeconomic data into a grid structure. This is presented in more detail in the technical report “Disaggregation of socioeconomic data into a regular grid: Results of the methodology testing phase”. Second, it is utilized in order to assign population to Urban Morphological Zones (UMZs).

However, the suitability of using the population grid for different purposes, as well as its reliability in different settings, has not been subject to much scrutiny. Still, some validations of the population grid have been performed. For instance, a comparison with Austrian reference data at the km<sup>2</sup> level showed an overall reduction by 50 percent in the disagreement with reference data, when compared to a non-weighted distribution of the population.<sup>1</sup> Against this background, it is not without interest to examine how the population grid compares to Swedish register data.

## 1.1 Methodology

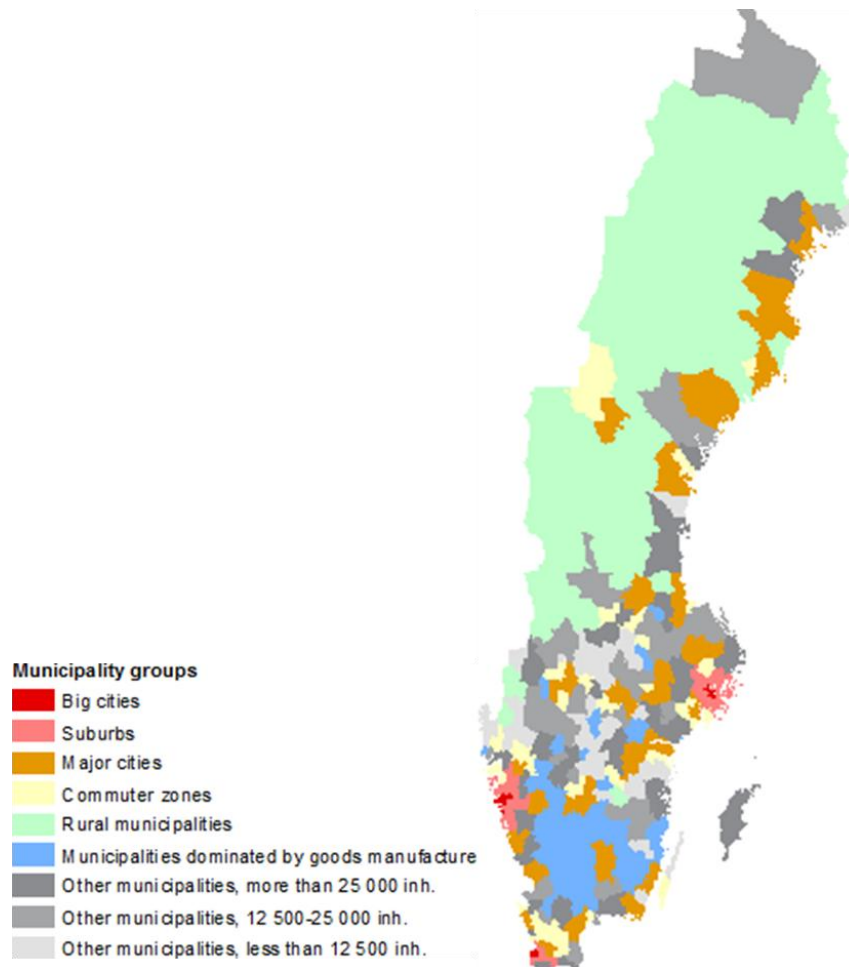
Since the population grid departs from population figures for LAU, level 2—in the Swedish case, municipalities—grid population summarized at that level can be expected to largely correspond to register data. However, at the local level, it may be more or less reliable. Similarly, the performance of the grid in estimating population figures for other geographical subdivisions (e.g., UMZs) is unclear. Taking into account how the population grid is employed in the ESPON db project, the examination focuses on 1) population estimates in varying local settings, and 2) the estimation of overall population for UMZs of different sizes.

The basis for the **first test**, concerning grid population estimates in varying local settings, is a calculation of residuals. This is carried out by comparing an aggregation of the population grid to square kilometers with corresponding Swedish register data. Absolute residuals are then summarized at the municipality level (n=290). In addition to looking at the results per municipality, results are also categorized by “municipality groups”—a classification of Sweden’s municipalities in nine different groups, created in 2005 by the Swedish Association of Local Authorities and Regions. The municipality group classification aims at defining homogenous regions, which share similar characteristics in terms of for instance population size, commuting patterns and

---

<sup>1</sup> Gallego J., Downscaling population density in the European Union with a land cover map and a point survey, JRC-Ispra.

employment profile. There are nine municipality groups, presented in Figure 1 and Table 2. This first test considers not only the absolute residual sum, but also the residual sum in relation to municipality area (expressed in km<sup>2</sup>) and population size. When relating residual sum to population size, initial figures are multiplied by 50 in order to get a more gini-style estimation of the overall discrepancy between grid and register data.



**Figure 1:** Municipality group map

ID	Category	Number of municipalities
1	Big cities	3
2	Suburbs	38
3	Major cities	27
4	Commuter zones	41
5	Rural municipalities	39
6	Municipalities dominated by goods manufacture	40
7	Other municipalities, more than 25,000 inhabitants	34
8	Other municipalities, 12,500–25,000 inhabitants	37
9	Other municipalities, less than 12,500 inhabitants	31

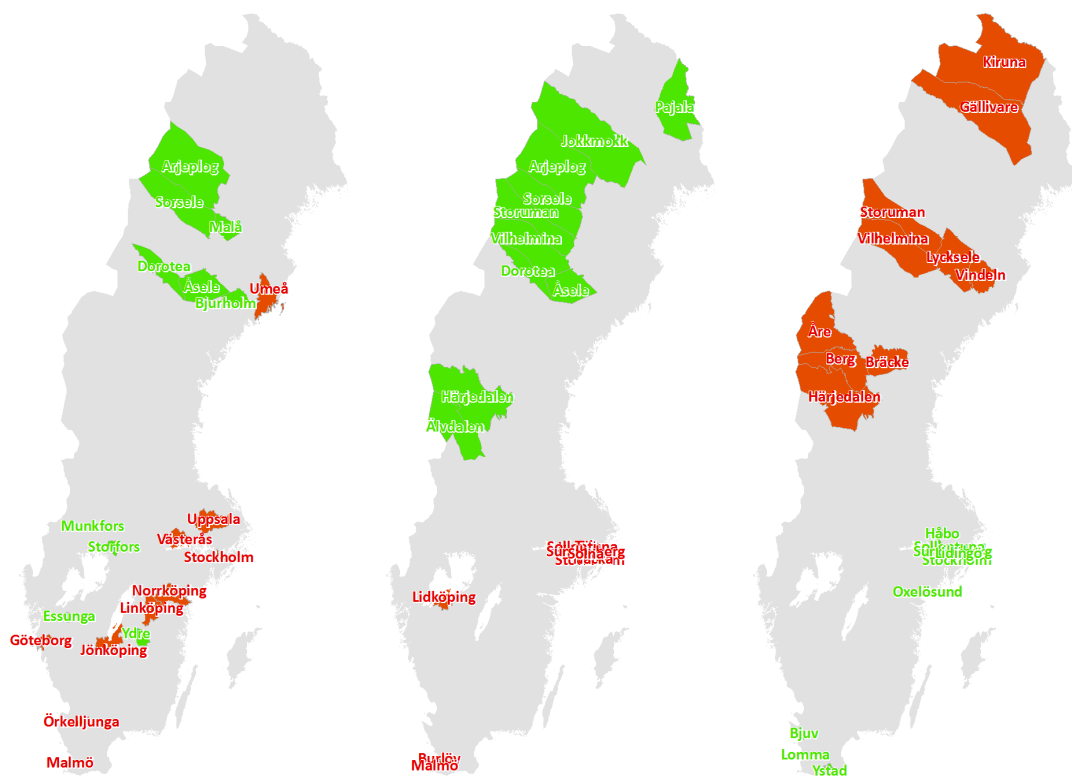
**Table 1:** Municipality groups: IDs and frequencies

The **second test** concerns using the population grid to estimate the population of Urban Morphological Zones (UMZs)—a delimitation of urban areas with EU coverage. In this test, overall population figures for each UMZ are calculated using both the original population grid and register data, which then are used for calculation of per-UMZ residuals. Thus, in contrast to the first test—which is based on the sum of absolute square residuals—this test focuses is the overall predictive capabilities of the grid, when it comes to UMZs of different sizes. Grid residuals within each UMZ have also been produced, primarily for purposes of trying to clarify patterns of over- and underestimation.

## 2 A Country-Level Examination

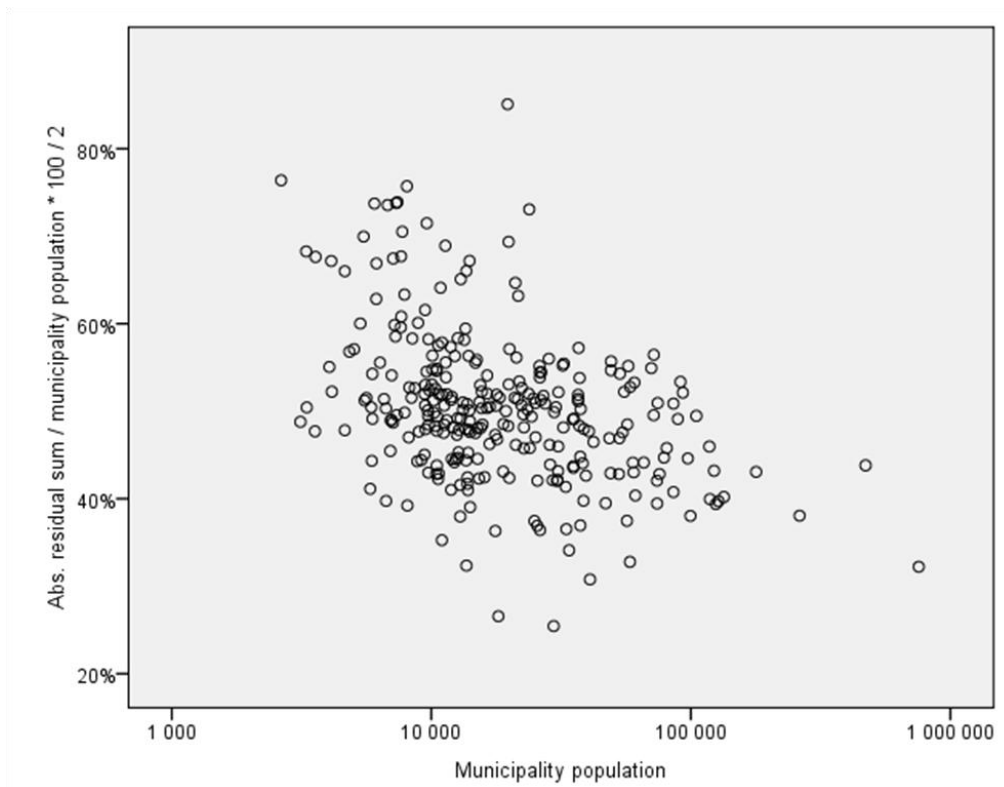
### 2.1 Population Estimates in Varying Local Settings

The first test of the population grid concerns population estimates in varying local settings. Clearly, the way discrepancies between grid and register data is associated to the local context depends on whether absolute residuals are just summarized or related to area or population. In Figure 2, the ten municipalities with highest (red) and lowest (green) absolute residuals are displayed, using three different measures: residual sum (left) as well as residual sum related to municipality area (middle) and population size (right). Municipalities with a comparatively large population (e.g., Malmö) tend to fare quite bad regarding residual sum and residual sum related to area, but pretty good when residual sum is related to population size. For municipalities with a comparatively small population (e.g., several municipalities in the inland of Northern Sweden), the situation is the opposite.



**Figure 2:** The ten municipalities with highest (red) and lowest (green) absolute residuals summarized (left) and in relation to area (middle) and population size (right)

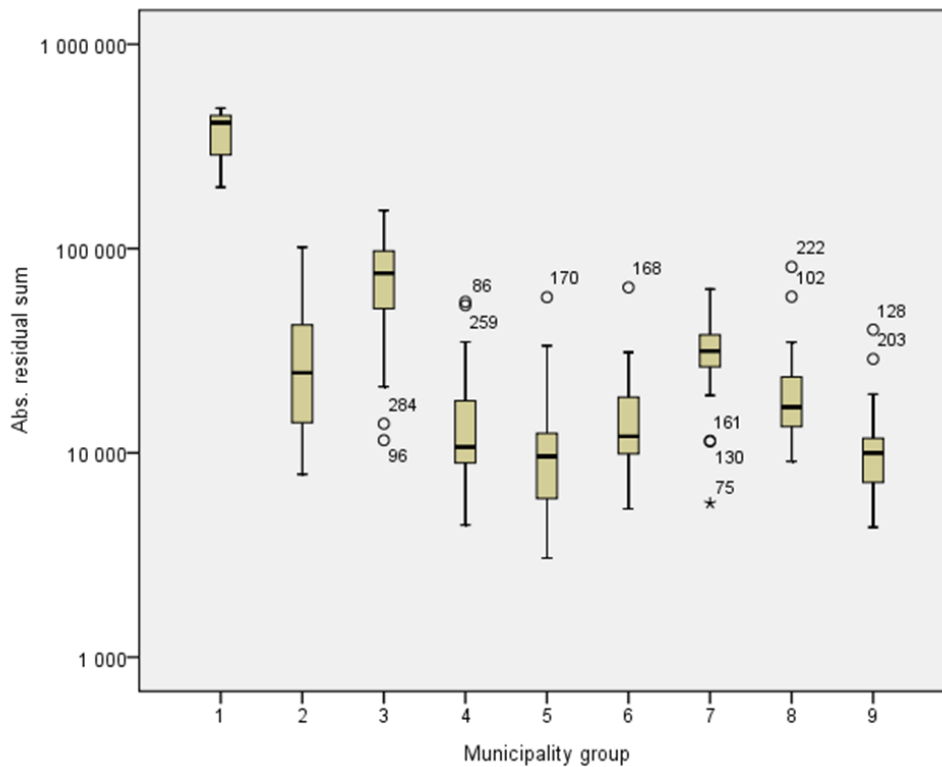
In Figure 3, all 290 municipalities are displayed in a scatterplot, with population size on the x-axis and absolute residuals by population size on the y-axis. The scale on the x-axis is logarithmical. There is a clear relationship between the two dimensions. As municipality population size increases, residual sum relative population size tends to decrease. However, this overall relationship is not without exceptions. In particular, for municipalities with a population of about 10,000 inhabitants, there are considerable variations in the level of overall discrepancy between grid and register data.



**Figure 3:** Municipality population size compared to absolute residual sum in relation to municipality population size

### 2.1.1 Results by Municipality Group

In order to get a better understanding of how the three residual measures are related to the local context, municipalities are categorized by the municipality group to which they belong (see Figure 1 and Table 1). By the use of boxplots to graphically present the results, variations within and between these varying kinds of local settings becomes apparent. The first boxplot (Figure 4) displays absolute residual sum. The by far highest median error can be found in group 1, "big cities". Municipality groups 3 ("major cities") and 7 ("other municipalities, more than 25,000 inhabitants") also exhibit comparatively large median errors (cf. Figure 2, left). It should be noted that the scale on the y-axis is logarithmical.

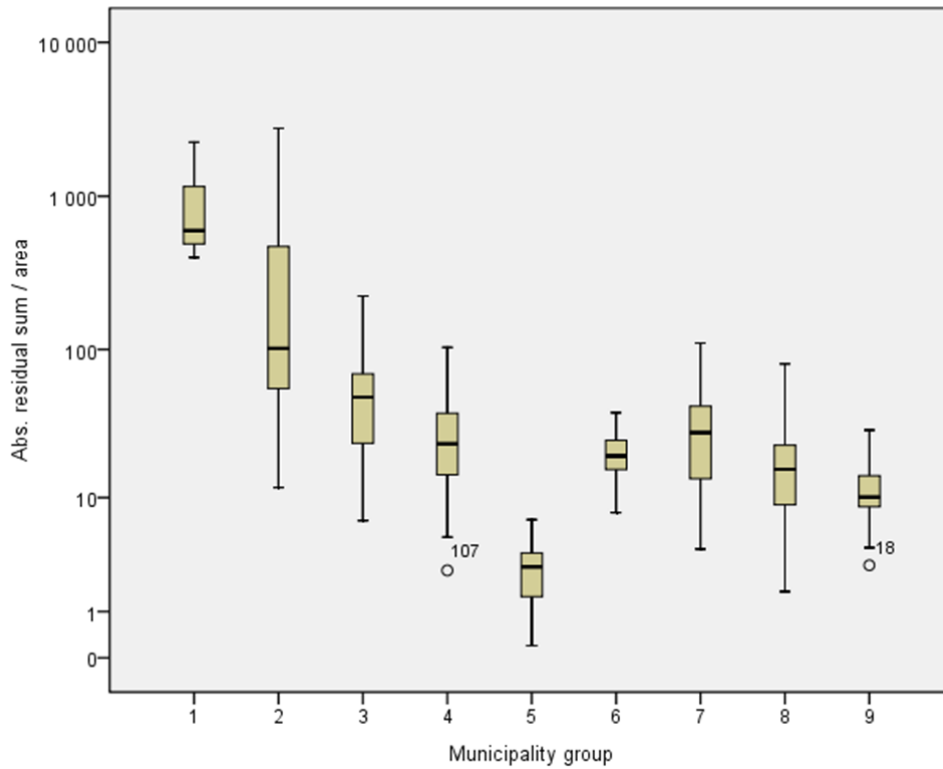


**Figure 4:** Absolute municipality residual sum subdivided by municipality group

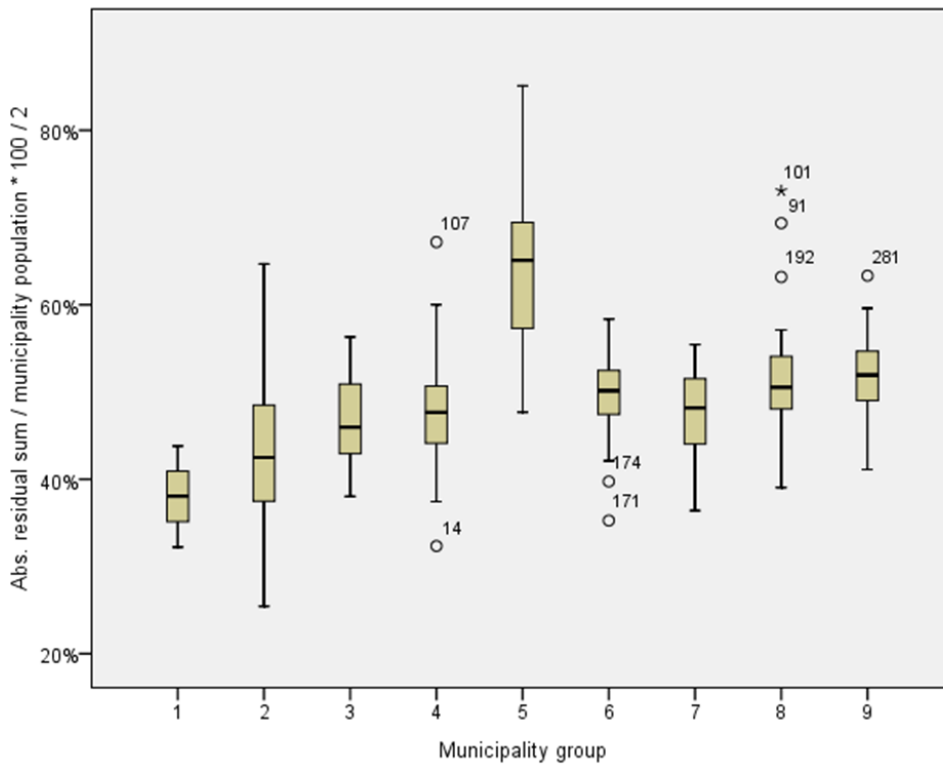
In relation to area (Figure 5), a somewhat similar pattern of differences between municipality groups emerges. The big cities category (1) exhibits the largest median error; rural municipalities (5) clearly the smallest (cf. Figure 2, middle). When it comes to municipality group 2, which represents suburban municipalities, there is a substantial internal variation. It should be noted that the scale on the y-axis is logarithmic. For the gini-style measure of residual sum related to municipality size, the pattern is quite different (Figure 6). Rural municipalities exhibit the largest median error; big cities by far the smallest (cf. Figure 2, right). There are substantial variations within not only suburban, but also rural municipalities.

Like the results presented in Figure 2 and Figure 3, the municipality group comparison indicates that there is a relationship between the three residual measures and population size. In addition, it reveals substantial variations within certain municipality groups. In the case of suburban municipalities, it is easy to see why such diversity may arise. The settlement structure in suburban areas varies considerably, ranging

from spacious residential area to crowded housing estates. Presumably, the mix of suburban housing in certain municipalities produces a population distribution more in line with the figures of the population grid.



**Figure 5 :** Absolute municipality residual sum in relation to area, subdivided by municipality group

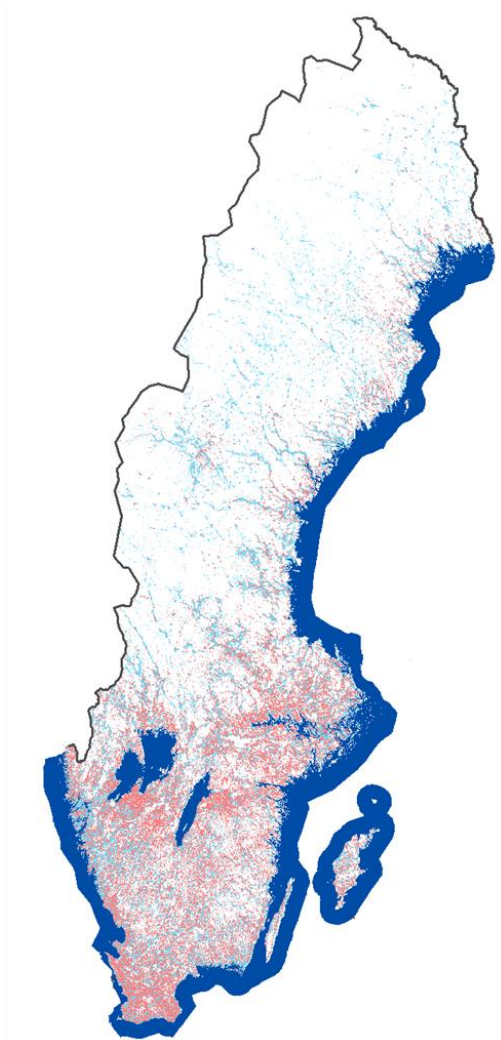


**Figure 6:** Absolute municipality residual sum in relation to population size, subdivided by municipality group

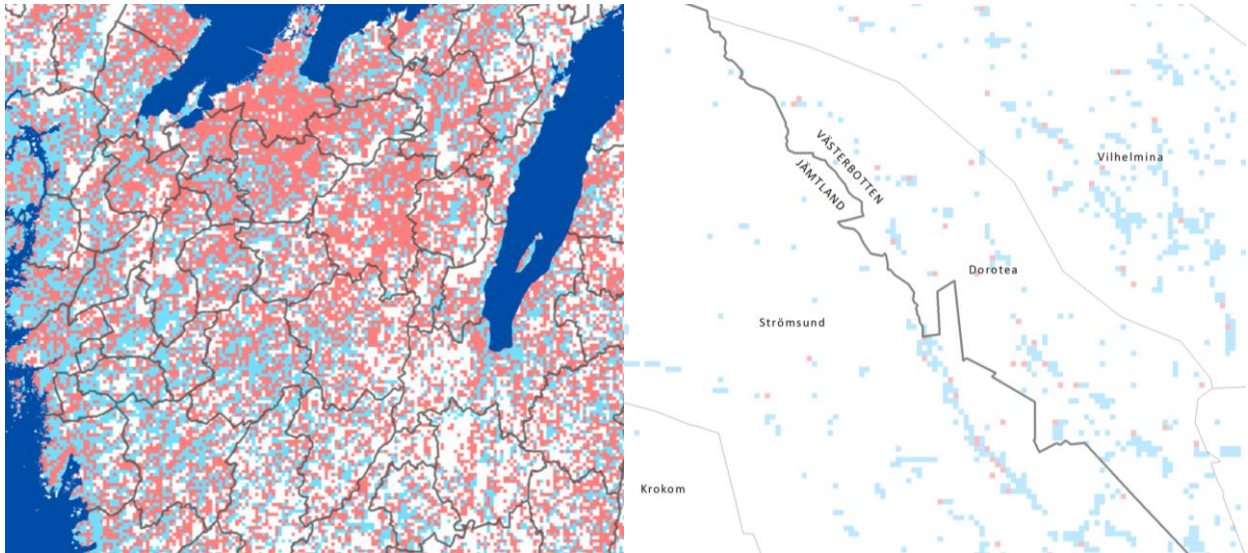


## 2.1.2 The Residual Map

In the interpretation of the results, the underlying km<sup>2</sup> residual map may yield some clues. Figure 7 shows a residual map for Sweden as a whole; Figure 8 two close-ups of the residuals, representing a part of Southern (left) and Northern (right) Sweden. In these maps, red color means that the grid population is larger than the register population. Conversely, blue color indicates that the grid population is smaller than the register population. It should be noted that these maps only show squares that are inhabited in register data. As can be seen in Figure 7 and—even more clearly—Figure 8, there is a tendency for the grid to underestimate the population size of inhabited squares in the inland of Northern Sweden. Primarily, this is due to the assignment of population figures to many actually uninhabited squares. Naturally, this is a phenomenon that is likely to be more pronounced in sparsely populated areas, such as the rural municipalities in the inland of Northern Sweden (cf. Figure 1).



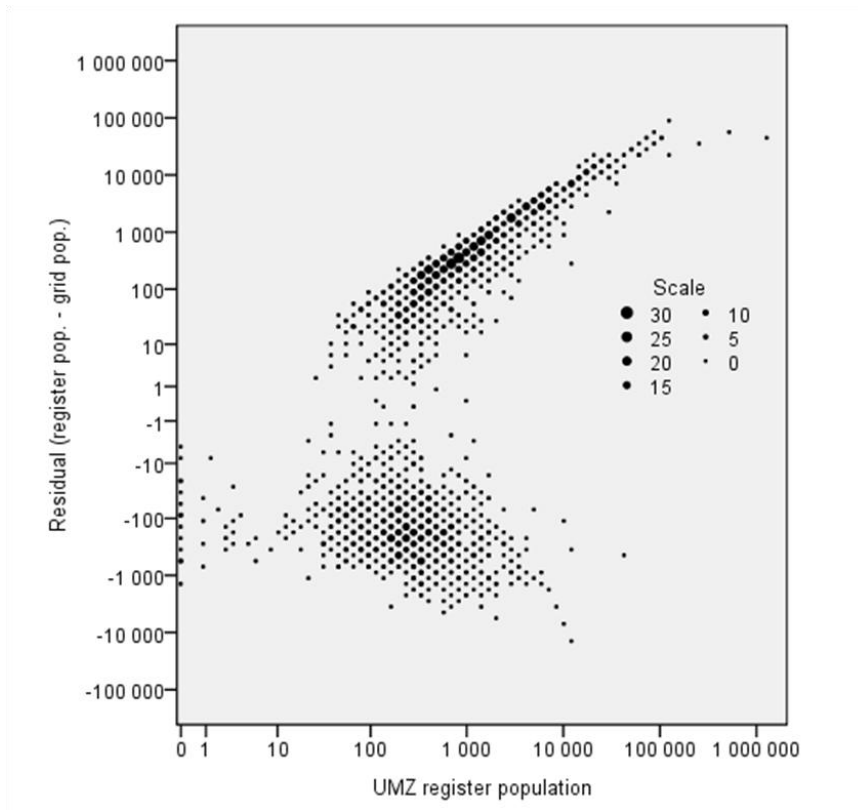
**Figure 7:** Absolute km<sup>2</sup> square residuals



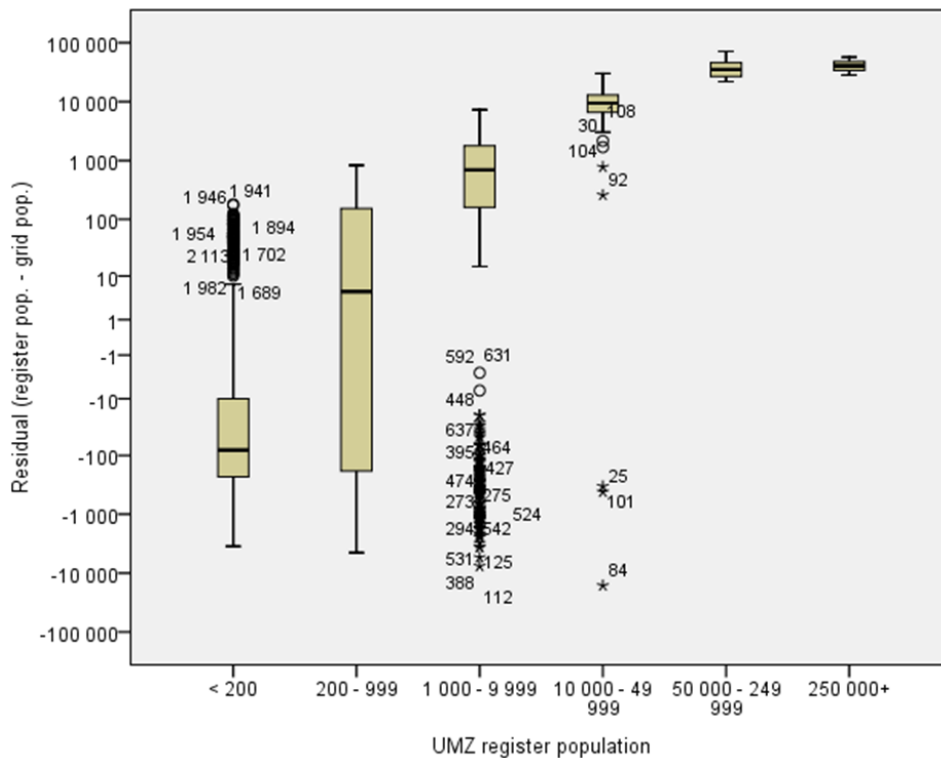
**Figure 8:** Close-up of  $\text{km}^2$  residuals in Southern (left) and Northern (right) Sweden

## 2.2 Estimations of UMZ Population

The second test of the population grid, which concerns estimations of overall UMZ population, reveals an intriguing pattern of varying degrees over- and underestimation depending on UMZ size. Figure 9 displays, in scatterplot form, UMZ register population size on the x-axis, and overall residuals (register population – grid population) on the y-axis. For both the x- and the y-axis, the scale is logarithmical. In addition, the observations are binned: the larger the dots, the more UMZs are located in and around that point in the scatterplot. All in all, the number of over- and underestimated UMZs are about equal. Two separate clusters—one of which exhibits a linear trend—are clearly evident. First, for most UMZs with about 1,000 inhabitants and more according to register data, population is underestimated, and the underestimation increases with UMZ size. Second, the population of many UMZs with a register population below 1,000 inhabitants is—more or less—overestimated. In the boxplot that makes up Figure 10, this phenomenon is evident by the large range and many outliers in the UMZ size classes “200-999” and “1000-9,999”, respectively. Generally, the amount of over- and underestimation is quite modest, especially when viewed in the light of actual population size.



**Figure 9:** UMZ residuals by UMZ register population



**Figure 10:** UMZ residuals by UMZ register population classes

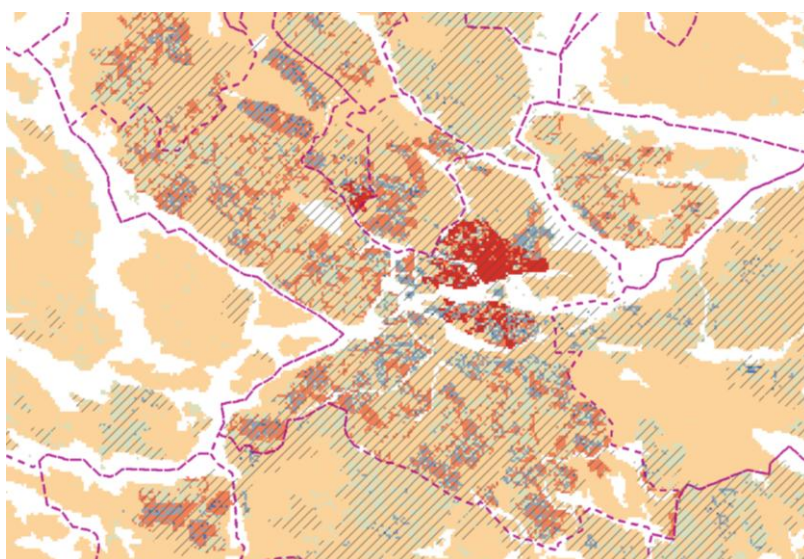
When it comes to the overestimation of many small UMZs, a possible explanation could be that the population grid overestimates areas with many buildings but small resident population. Spatial agglomerations of second homes, which are quite common in Sweden, are obvious examples of this kind of area. In Table 2, the UMZ

boundaries are related to on the one hand the Swedish definition of urban localities (*tätort*)—basically, any agglomeration of 200 or more inhabitants—and on the other hand a delimitation of “other concentrated settlement”. For UMZs that intersect any urban locality, 40 percent are overestimated. By contrast, for UMZs intersecting either only “other concentrated settlement” or neither category, the overestimation figure rises to about 80 percent. In practice, the “other concentrated settlement” category is largely made up by second home areas. Clearly, then, this finding lends some support to the notion of second home areas being responsible for the cluster of overestimated UMZs. Still, 80 percent of UMZs actually overlap urban localities, and a substantial proportion of those UMZs are also overestimated. In other words, the pattern of overestimation may also be a question of UMZ size.

<b>Relation to Swedish delimitations</b>	<b>% of UMZs</b>	<b>% of UMZs overestimated</b>
UMZ intersects neither urban locality ( <i>tätort</i> ) nor “other concentrated settlement”	10	82
UMZ intersects only “other concentrated settlement”	10	79
UMZ intersects urban locality ( <i>tätort</i> )	80	40

**Table 2:** UMZs in relation to Swedish definitions of settlements

Concerning the general and increasing underestimation of large UMZs, there is harder to find an explanation for the phenomenon. Figure 11 shows grid residuals for UMZs in the Stockholm area, including residuals for Stockholm UMZ—the largest UMZ in Sweden in terms of population size, and also among the most underestimated using the population grid. In this map, red and—in particular—dark red color means that the grid population is larger than the register population. Conversely, the two shades of blue indicate that the grid population is more or less smaller than the register population. In the city center, there is—not surprisingly—a clear tendency for the grid to overestimate the population, while suburban areas generally exhibit a mixed pattern of over- and estimation. While this overall residual pattern is likely to occur in many other larger UMZs, it gives no obvious clue as to the reasons for the overall trend towards increased population underestimation with increased UMZ size.



**Figure 11:** Close-up of 100 meter<sup>2</sup> residuals in UMZs in the Stockholm area

### 3 Summary and Discussion

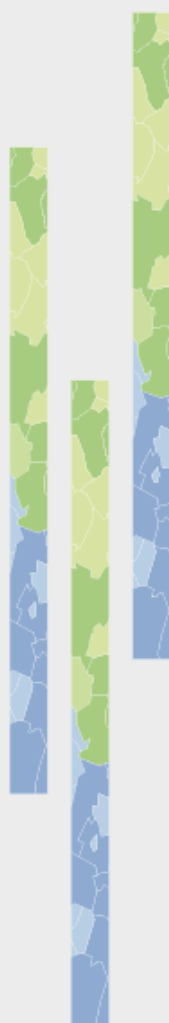
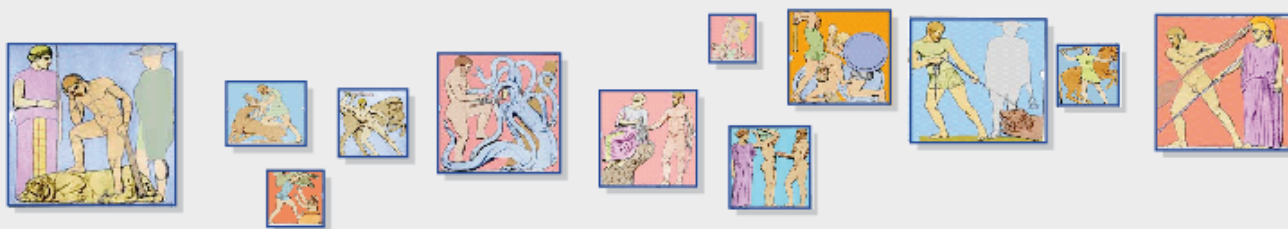
In the ESPON db context, there is a need for population data with a high degree of spatial resolution. For instance, when it comes to disaggregating certain socioeconomic variables, or reporting population figures for geographical subdivisions such as UMZs, available data with good spatial coverage (e.g. for NUTS 2 or LAU, level 2 regions) have obvious limitations. Therefore, downscaled population data—specifically, a JRC population grid covering the entire EU—has been employed for these purposes. Against this background, the dataset has been subject to a country-level examination using Swedish register population data. Taking into account how the population grid is employed in the ESPON db project, the examination focuses on 1) population estimates in varying local settings, and 2) the estimation of overall population for UMZs of different sizes.

The first test summarizes absolute local residuals for Sweden's 290 municipalities (i.e., LAU, level 2 subdivisions), using three different measures: residual sum as well as residual sum related to municipality area and population size. Results are also presented categorized by municipality group—a classification of municipalities in nine different groups according to their characteristics. The results indicate that there is a relationship between municipality population size on the one hand, and the three residual measures on the other. Municipalities with a large population are associated with low discrepancies between grid and register data when residual sum is related to population size, but high discrepancies in terms of residual sum and residual sum related to area. For small municipalities, such as many rural municipalities in Northern Sweden, the situation is the opposite. A map of the actual local residuals reveals that there is a tendency for the population grid to underestimate the population size of inhabited squares in such settings. Primarily, this is due to the assignment of population figures to many actually uninhabited squares. In an EU perspective, this is likely to be less of an issue. The municipality group comparison reveals substantial variations within certain municipality groups, particularly regarding the category representing suburban municipalities. Presumably, this reflects the considerable diversity in settlement structure and population distribution that exist in suburban areas.

The second test concerns using the population grid to estimate the population of Urban Morphological Zones (UMZs). When overall residuals are related to UMZ population size, the about equal number of over- and underestimated UMZs form two separate clusters. For large UMZs the number of inhabitants tends to be underestimated, and the underestimation increases with UMZ size, while the population of many small UMZs is—more or less—overestimated. A plausible explanation for the latter phenomenon is that the population grid overestimates areas with many buildings but small resident population, such as second home areas. Generally, the amount of overall over- and underestimation is quite modest, especially when actual UMZ population size is taken into account.

In this country-level examination of downscaled population data, the discrepancies between downscaled and register data varies depending on local setting, and is also highly influenced by the way residuals are expressed. In any case, it is hardly a stretch to conclude that local grid population estimates often are quite unreliable. Still, using the population grid to downscale socioeconomic data is a likely to enhance to quality of data, and—and least for now—no better alternative exists. In the estimation of overall UMZ population size, the population grid works quite well—at least in the Swedish context. Consequently, while there are obvious limitations to downscaled

population data, the JRC population grid is a quite reasonable tool for the enhancement of certain ESPON datasets.



## SPATIAL DATA CREATION FOR EUROPE

### *A TEST CASE*

#### CONTENT

- Research problem: The report investigates how to enhance the availability of spatially fine-grained socio-economic data required for addressing urgent societal problems in Europe.
- Methodological development  
The report proposes a method for merging available tables from surveys and censuses with each other — and with remote-sense based population grid data — in order to gain a richer set of spatially fine-grained socio-economic information.
- Tentative outputs and future possibilities  
The report gives examples of the outcomes of a test version of the method for France and Sweden and discusses lessons learned for applying such methods for creating synthetic localized individuals in the entire Europe.

**ESPON 2013 DATABASE**  
**MARCH 2011**



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

**15 PAGES**

# LIST OF AUTHORS

Einar Holm, Dept. of Social and Economic Geography, Umeå University

Magnus Strömgren, Dept. of Social and Economic Geography, Umeå University

## **Contact**

[einar.holm@geography.umu.se](mailto:einar.holm@geography.umu.se)

[magnus.stromgren@geography.umu.se](mailto:magnus.stromgren@geography.umu.se)

tel. + 46 90 786 52 58



# TABLE OF CONTENTS

<b>1</b>	<b>Introduction .....</b>	<b>3</b>
1.1	The Problem .....	3
1.2	An Intermediate Solution .....	3
1.3	On the Target Dataset .....	5
1.4	Previous Experiences .....	5
<b>2</b>	<b>The Test Case: France and Sweden .....</b>	<b>6</b>
<b>3</b>	<b>Results .....</b>	<b>9</b>
<b>4</b>	<b>Discussion .....</b>	<b>14</b>
	<b>References .....</b>	<b>15</b>

# 1 Introduction

## 1.1 The Problem

For the foreseeable future, most available indicators of the state of the European population in terms of basic demography, family status, origin, education, work and income, etc. are highly aggregated. A large amount of information is compiled by Eurostat as national totals and averages. A few items are presented at the NUTS 2 and 3 levels. There is, however, the JRC (Joint Research Centre) population grid, which — despite some shortcomings — is a quite reasonable tool for certain purposes (Strömberg and Holm, 2010). This situation is prevailing despite the fact that substantial parts of concurrent social and economic problems of Europe are related to huge differences in the detailed spatial distribution of resources and conditions for local populations, as well as to the selective location of individuals with different socio-economic properties within the national and regional populations of Europe.

Conditions for integration of immigrants in a country are completely different if the newcomers are distributed as the general population or if they are concentrated to a few ghettos. The main development problem for remote areas in Europe is not lack of jobs but lack of qualified local specialists for the emerging jobs replacing the outdated ones — as well as for remaining qualified tasks required to be performed locally also in the future. This adds up into a national mismatch decreasing efficient use of human resources and hindering economic growth. The majority of the population and new economic activities are located in city regions covering a tiny fraction of space whereas a small share of the (aging) population is distributed over substantial parts of the countryside; containing economic activities receiving a large part of European subsidies and development support. Information for analysis and counteraction related to such urgent problems are effectively hidden by current aggregation levels in available data.

## 1.2 An Intermediate Solution

Since radical changes in availability of spatially fine-grained socio-economic data are far beyond the horizon, it is important to use current data as efficiently as possible. In this report, we propose and test a methodology based on merging available national and regional data tables from surveys and censuses with each other, as well as with remote-sense based population grid data. The idea is that the tables together contain more information about the joint distribution over space and attributes than do the single tables side by side.

One possible route is to start by creating an artificial synthetic individual database of the European population. In a first assessment, the attributes of the individuals are assigned more or less randomly, i.e. only conditioned by totals for nations and regions. It is advisable to align the distribution of the start population based on at least two attributes, like age and region. After that, in the main step of the procedure, the artificial individuals successively exchange attribute values with each other. Each exchange is performed as long as it improves the total fit between the observed table cells and the corresponding cells in tables created by aggregating the current artificial

individuals. In the end, a synthetic individual database is created that is indistinguishable from the authentic (but unobserved) one when it comes to produce all the observed data tables together. The created synthetic individuals are jointly consistent with all information in the supplied tables (Figure 1).

After that, arbitrarily tables with dimensions corresponding to the attributes of the synthetic individuals can be computed. Many of those possible new tables were not contained in the set of used data tables, like population by age, sex, labor participation per km<sup>2</sup> square and so represent the new, created information. Of course, especially with just a few tables as the point of departure, the created set of attribute values for each individual is just an example out of many other distributions of sets of attributes that also, when aggregated, will be perfectly consistent with all supplied tables. The degrees of freedom for existence of such alternative populations however decreases with each additional supplied table used as constraint on the resulting population composition.

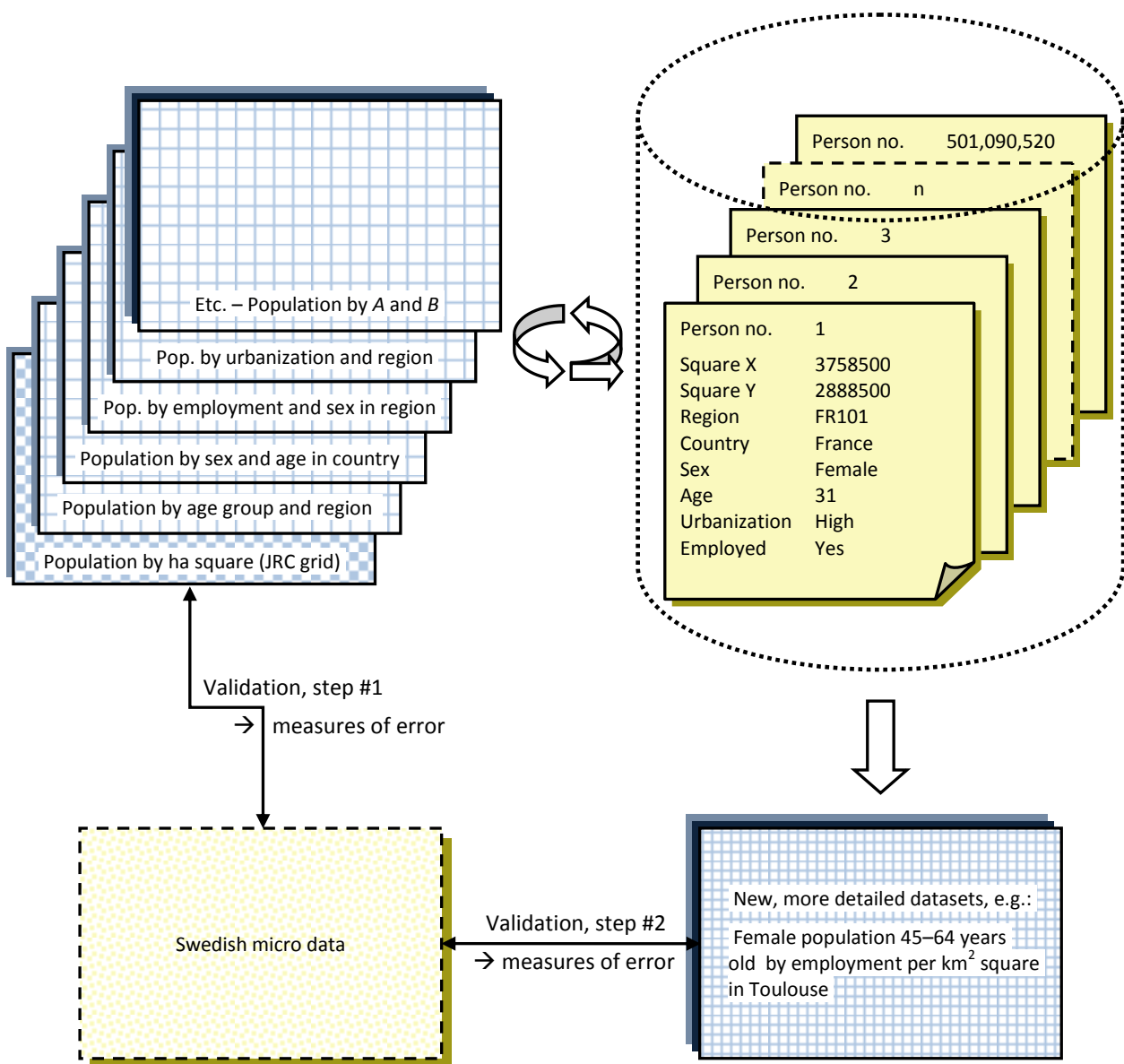


Figure 1. Using disaggregated population data together with different survey and census tables to produce a synthetic individual database — as well as new datasets.

### 1.3 On the Target Dataset

Most imaginable specific analyses requiring fine-grained spatial data not available today do not require the opposite extreme: micro data — information on individuals. For each of the problems exemplified above, the analysis would become greatly enhanced just by data at the km<sup>2</sup> level cross-classified with a few population attributes like age, sex, origin, etc. in addition to population size. As the intended applications for the created database increase, there will be additional requirements of attributes to cross-classify. Soon, the number of cells in the combined multi-dimensional table for each km<sup>2</sup> square will exceed the number of individuals in most km<sup>2</sup> squares; in the cross-classified version, most cells would be empty. So — based on Swedish experiences — as soon as the number of attributes/dimensions exceed seven or eight, the number of cells in the multidimensional representation exceed the number of individuals times attributes in the represented population. Thus, already from a resolution (and storage) point of view in all but the simplest applications, an individual representation of the target population is preferable even when no individual-level analysis is in the pipeline. In addition, many kinds of analyses would be enhanced by having access to the individual data, e.g. by being able to distinguish between individual-, group- and regional-level contributions (and interactions) in a multi-level analysis of variance components. Finally, if the quest were to feed an agent-based microsimulation with empirical content, individual data would be required. If not, the other supplied arguments for creating artificial individuals still stand.

### 1.4 Previous Experiences

The problem of scarce availability of spatially fine-grained socio-economic data is of course not a recently discovered one, but rather a long-standing severe constraint on country-based and European-wide spatial analysis faced by many researchers. As an example, U.K. geographers (e.g. Ballas *et al.*, 2007; Birkin and Clarke, 1989, 1998; Smith *et al.*, 2009; Voas and Williamson, 2000) have developed several examples of a methodology based on imputing the richer attribute information available in national surveys into the spatially more fine-grained census information on population by age and sex, etc. for local districts (wards), as constrained by the dimensions common for both the surveys and the census. Their efforts to create U.K. and Ireland local micropopulations containing more relevant information for current social research compared to what is available directly from published census tables target the same kind of problem as the one addressed for all countries of Europe in this report.

While U.K. geographers have pioneered the development of methods for generating synthetic populations, there are examples of such population generation from other countries as well. Using U.S. census data, Beckman *et al.* (1996) developed a method for creating synthetic populations for use in activity-based transportation models. In developing a microsimulation model for Southern France, a synthetic population dataset was produced (Aschan-Leygonie *et al.*, 2000; Holm *et al.*, 1999, 2000). In that case census tables for communes were available. Since communes constitute a quite small spatial unit, it was possible to construct a fairly detailed micropopulation with a high degree of spatial resolution. However, for the current endeavor targeting all of Europe it is not possible to presuppose the existence of such data sources for each country.

## 2 The Test Case: France and Sweden

In order to demonstrate the methodology, results from a test case with a somewhat constrained scope and methodology compared to what is required for a full-scale application for the entire Europe is presented in the following section. The test has been applied on data for two countries, France and Sweden, instead of for all European countries. The goal for the test is to fill a defined table with 26 combined age, sex and employment attributes for the population of each km<sup>2</sup> square. Data for the test is a small set of information contained in only five tables:

1. Total population per km<sup>2</sup>. The table contains one row for each of the ca 860,000 inhabited km<sup>2</sup> squares of the two countries. In addition to total population per km<sup>2</sup> in year 2001, the table rows contain coordinates as well as codes for the NUTS 0, NUTS 2 and NUTS 3 region they belong to. Source: The JRC population grid.
2. Population by age in three age groups (0–14, 15–64 and 65+ years) in each NUTS 3 region year 2001. The table contains the number of persons in the different age groups in each NUTS 3 region of (mainland) France and Sweden — 96 departments and 21 counties, respectively. Source: Eurostat regional demographic statistics.
3. Population by sex in each NUTS 3 region year 2001. Source: Eurostat regional demographic statistics.
4. Employment by age in six age groups (15–24, 25–34, 35–44, 45–54, 55–64, and 65+ years) in each NUTS 2 region year 2001 — corresponding to 22 regions in (mainland) France and 8 national areas in Sweden. Source: Labour force survey.
5. Employment by sex in each NUTS 2 region year 2001. Source: Labour force survey.

The goal for the test model was set to “expand” table 1 into a result table distributing the total population of each km<sup>2</sup> square by age (six groups), sex and employment as accurately as possible given the information supplied by the five used tables. Figure 2 shows the first twenty rows of the result out of the ca 860,000 rows for each inhabited km<sup>2</sup> square in France and Sweden. The program creating the table works as follows:

1. One km<sup>2</sup> row is selected randomly with probability proportional to the number of remaining, not yet allocated individuals (random draw without replacement).
2. One of the 26 columns with all non-empty combinations of age group, sex and employment status is selected randomly with a probability reflecting the relative remaining “distance” to the constraint maximum in each relevant cell of the empirical tables — e.g. how close the created dataset is to the observed number of 45–54 years old employed females in a certain NUTS 2 region.
3. It is tested whether or not one person can be added to the chosen row and column cell without violating any constraint given by the observed cell values of the five tables. Firstly, total observed population of the chosen km<sup>2</sup> square should not be exceeded. Secondly, each of the 26 combined columns corresponds to specific cells in two or four of the remaining four tables. If e.g. column 5 was selected, then it is checked that current number of females allocated into the NUTS 3 area of the selected square in table 3 contains less



2. The employed allocation scheme in the test model application never reallocates a person from the first hit into another combination of km<sup>2</sup> square and age/sex/employment group. By accident it might happen that a combination becomes filled up to the total for the km<sup>2</sup> square at an early stage of the allocation process. That excludes later arrivals from entering otherwise still vacant positions in other combinations. Inspection of the result table reveals some cases with just a few columns filled especially for squares with small population in total. This calls for an allocation scheme in line with the intentions described above and in the discussion.

Despite the shortcomings of this simplified test model, it seems to enhance the information content possible to get from generally available socio-economic tables for Europe as demonstrated in the results section.

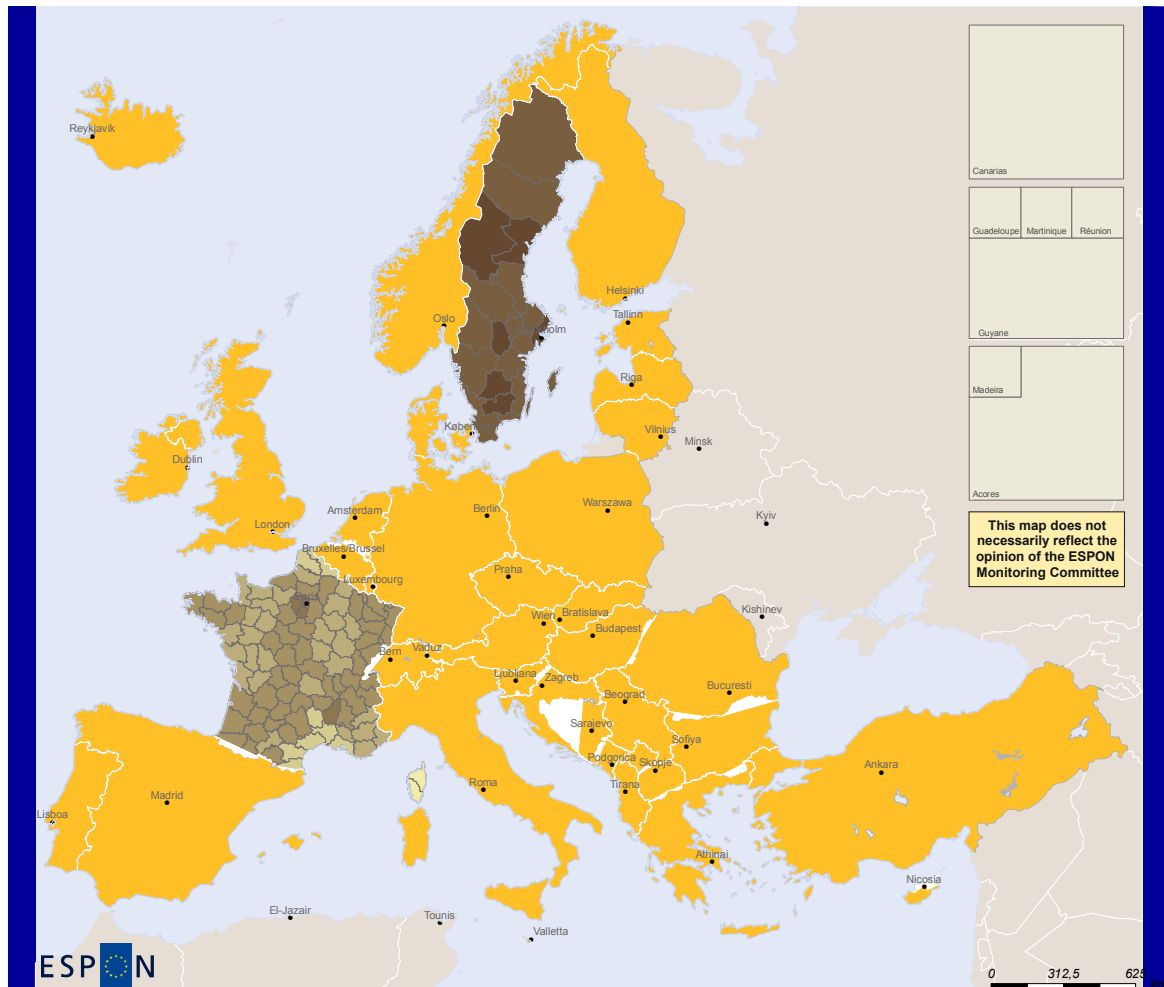
### 3 Results

The following maps illustrate some outcomes of the test. Map 1 shows work participation level for females aged 45–64 years per NUTS 3 region in France and Sweden as calculated by an earlier version of the test model. Map 2 demonstrates the same information for Sweden based on observed data. Comparing observed and calculated levels for Sweden reveals some errors in level and regional distribution. For instance, the differences between regions are smaller for calculated data compared to observed. This may reflect discrepancies between the used data sources — register data for the observed Swedish map; the Labour force survey and other Eurostat data for the calculated version — but also errors in the allocation algorithm.

The next two isarithm maps show work participation level for ages 15–64 years based on model-calculated values for km<sup>2</sup> squares (Map 3) and corresponding observed Swedish data (Map 4). Map 3 is created using the latest version of the test model. Comparing observed and calculated levels for Sweden demonstrates that the range of values correspond quite well with each other. However, the overall calculated employment levels are too low, due to a remaining error in the allocation algorithm. In this context, it should be mentioned that detailed maps would become more spatially diverse and interesting, should the information-creation model also use tables that create differences between km<sup>2</sup> squares within regions.



# Synthetic population

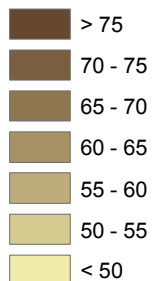


EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

© EuroGeographics Association for administrative boundaries

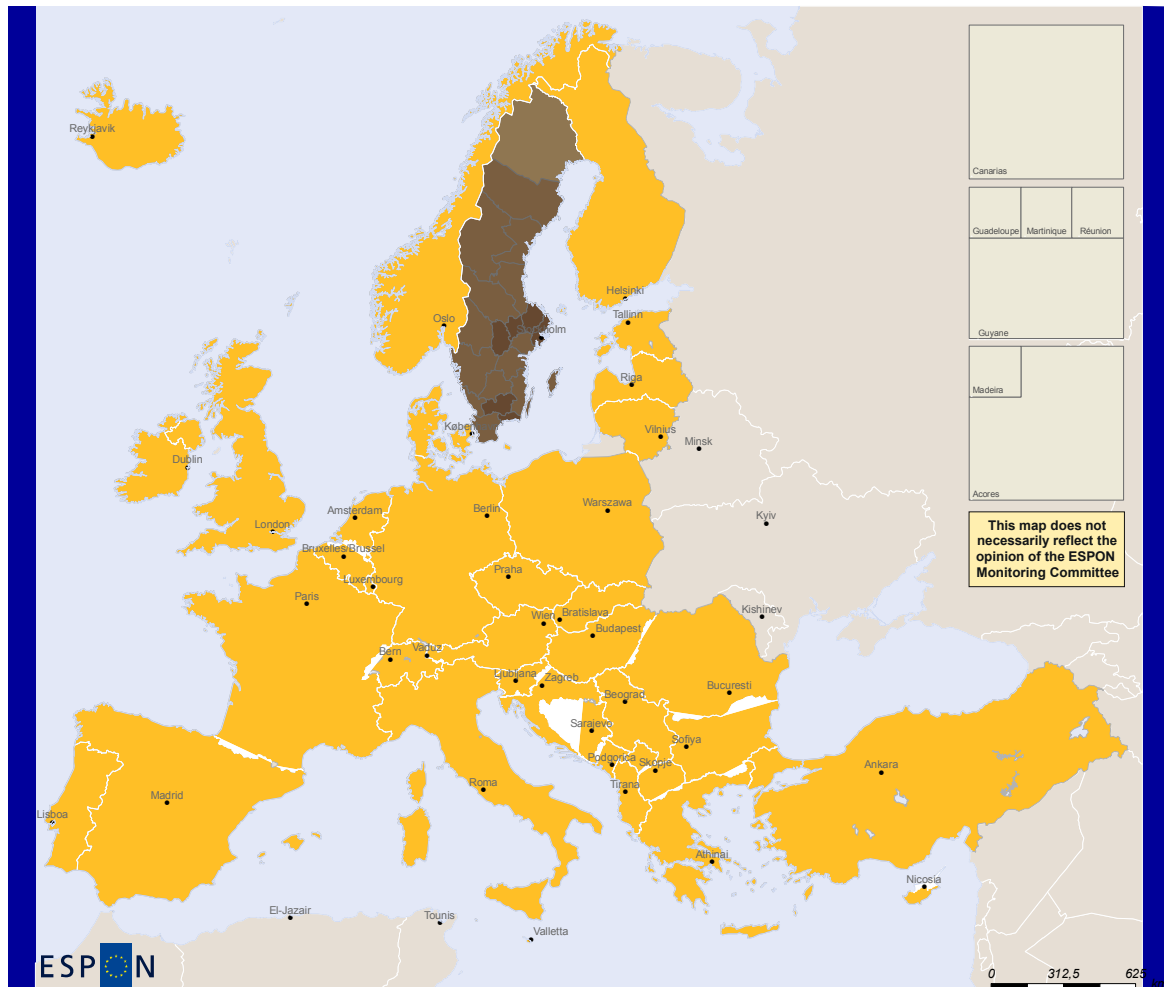
## Legend

### Work participation level 2001 (females 45–64 years)



Map 1. Work participation level for females aged 45–64 years per NUTS 3 region in France and Sweden as calculated by the test model (earlier version).

# Observed population



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

© EuroGeographics Association for administrative boundaries

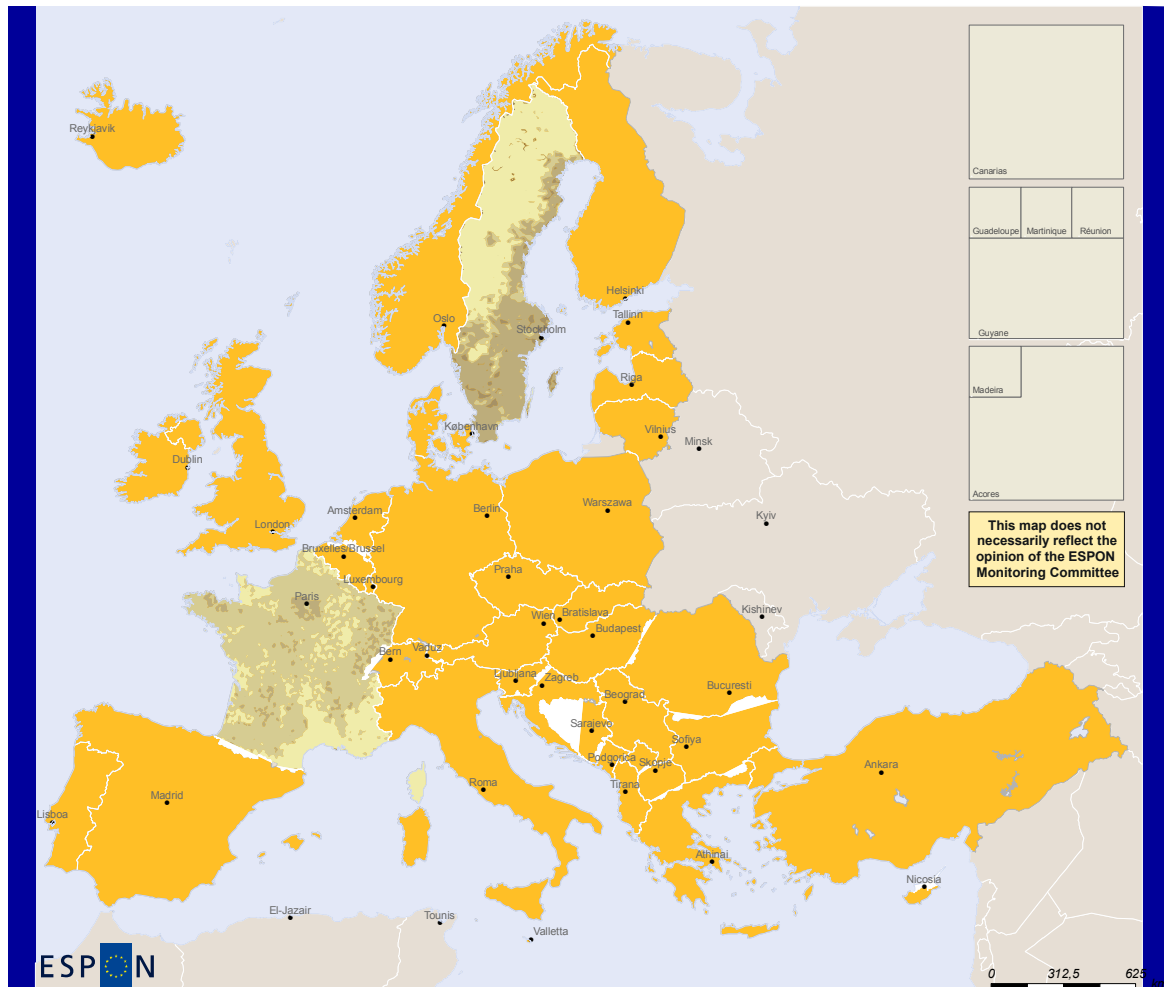
## Legend

### Work participation level 2001 (females 45–64 years)

- > 75
- 70 -75
- < 70

Map 2. Work participation level for females aged 45–64 years per NUTS 3 region in Sweden based on observed data.

# Synthetic population

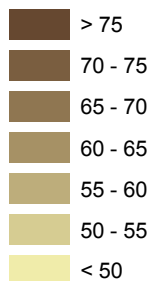



 EUROPEAN UNION  
 Part-financed by the European Regional Development Fund  
 INVESTING IN YOUR FUTURE

© EuroGeographics Association for administrative boundaries

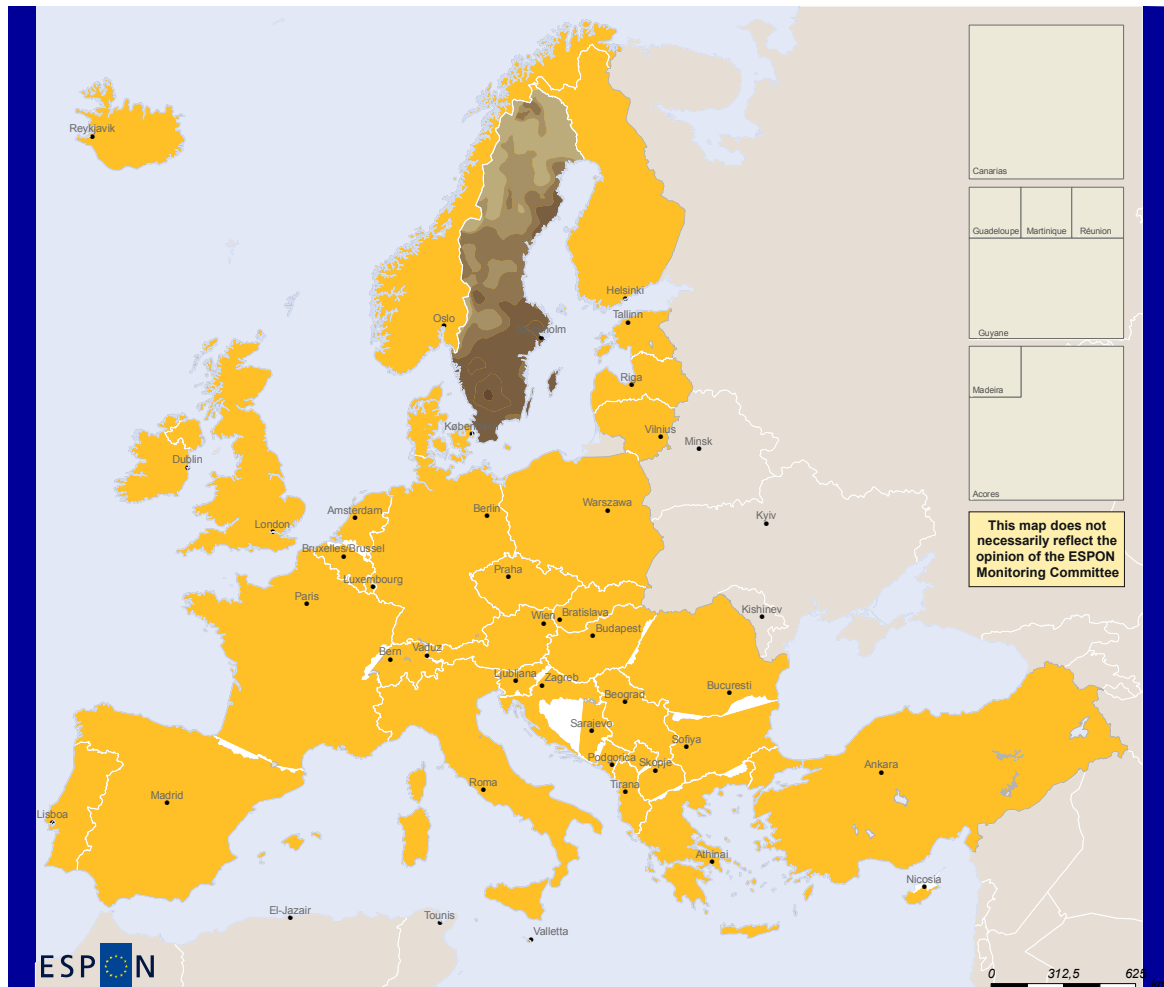
## Legend

### Work participation level 2001 (15–64 years)



Map 3. Work participation level for ages 15–64 years in France and Sweden as calculated by the test model (latest version).

# Observed population



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

© EuroGeographics Association for administrative boundaries

## Legend

### Work participation level 2001 (15–64 years)

- > 75
- 70 - 75
- 65 - 70
- 60 - 65
- < 60

Map 4. Work participation level for ages 15–64 years in Sweden based on observed data.

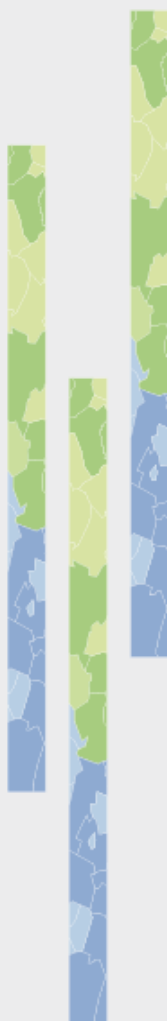
## 4 Discussion

Some lessons can be learned from this test application of a method for enhancing the spatial resolution of available socio-economic data for Europe:

1. The presented test application demonstrates that it is possible to enhance the spatial and socio-economic detail of available regional data for Europe by simultaneously merging them based on the discussed simulation methodology.
2. If increased spatial resolution is the main target then at least one of the used tables must contain a dimension differentiating areas on the finest targeted level of resolution. This was one of the most obvious shortcomings of the test model. None of the used tables differentiated between any properties of the km<sup>2</sup> squares within a NUTS 3 region. Therefore, the produced differences between such squares become tiny and random. The obvious remedy is to find tables relating e.g. population density classes to age, education level, employment or anything else relevant for a targeted application and thereafter connect each of the smallest spatial units to the density class they belong to.
3. In this application, the synthetic individuals only appeared for the nanoseconds necessary to populate the defined, relatively small result table. With more input tables and potential output detail, the two-step procedure suggested initially becomes the main candidate. There would be something like the procedure demonstrated in this application in order to create a reasonably consistent start population with attributes. After that, there would be a procedure to systematically exchange attributes between the synthetic individuals in order to further improve the fit with the input tables.
4. In a real application with many tables, it becomes necessary to introduce an explicit idea about how to handle inconsistencies between the different tables. It might be the case that some tables are judged more reliable than others *á priori* and/or it might be the case that some variables have a more consistent representation in several tables compared to other variables. One approach would be to use such judgments as prior probabilities for the different tables' information in the spirit of Bayesian estimation. Then, conflicting information is averaged away by means of differing weights. Another route would be to promote one table as the master table and force all other tables to adopt while maintaining their relative distributions in attributes not contained in the master table.

## References

- Aschan-Leygonie, C., Mathian, H., Sanders, L. and Mäkilä, K. (2000): "A spatial microsimulation of population dynamics in Southern France: a model integrating individual decisions and spatial constraints", in Ballot, G. and Weisbuch, G. (eds.) *Applications of Simulation to Social Sciences*, pp. 109–125. Oxford: Hermes Science Publishing.
- Ballas, D., Clarke, G., Dorling, D. and Rossiter, D. (2007): "Using SimBritain to model the geographical impact of national government policies", *Geographical Analysis*, vol. 39, pp. 44–77.
- Beckman, R., Baggerly, K. and McKay, M. (1996): "Creating Synthetic Baseline Populations", *Transport Research A*, vol. 30, pp. 415–429.
- Birkin, M. and Clarke, M. (1989): "The generation of individual and household incomes at the small area level using synthesis", *Regional Studies*, vol. 23, pp. 535–548.
- Birkin, M. and Clarke, M. (1998): "SYNTHESIS: a synthetic spatial information system for urban and regional analysis: methods and examples", *Environment and Planning A*, vol. 20, pp. 1645–1671.
- Holm, E., Lindgren, U., Mäkilä, K., Aschan-Leygonie, C., Baudet-Michel, S., Mathian, H., Sanders, L., Gautier (1999): "Micro-simulation of a population in a region with a strong urban growth", in A multiscale investigation into the dynamics of land abandonment in Southern France, vol. 5 (contract ENV4CT95-0159, DGX11 of EU).
- Holm, E., Lindgren, U. and Malmberg, G. (2000): "Dynamic microsimulation", in Fotheringham, S. and Wegener, M. (eds.) *Spatial Models and GIS: New Potential and New Models*, pp. 143–165. London: Taylor & Francis.
- Smith, D. M., Clarke, G. P. and Harland, K. (2009): "Improving the synthetic data generation process in spatial microsimulation models", *Environment and Planning A*, vol. 41, pp. 1251–1268.
- Strömberg, M. and Holm, E. (2010): *Using Downscaled Population in Local Data Generation: A Country-Level Examination*. ESPON Technical Report.
- Voas, D. and Williamson, P. (2000): "An evaluation of the combinatorial optimisation approach to the creation of synthetic microdata", *International Journal of Population Geography*, vol. 6, pp. 349–366.



## MAPPING GUIDE

### *CARTOGRAPHY FOR ESPON PROJECTS*

#### CONTENT

- **Maps and ESPON 2013.** This part presents the content of the map-kit tool ("European, local and global"). All the elements that have to be necessarily represented on each map are described.
- **Enhancing information.** This part explain how symbolize ESPON 2013 data with the good rules of graphic semiology.
- **Maps are tool for communication.** This part insists on the fact that a map has necessarily to deliver a clear message.

**ESPON 2013 DATABASE**

**JANUARY 2011**



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

**38 PAGES**

# LIST OF AUTHORS

Christine Zanin, University Paris 7, UMS 2414 RIATE

Nicolas Lambert, UMS 2414 RIATE

Ronan Ysebaert, UMS 2414 RIATE

## **Contact**

[christine.zanin@univ-paris-diderot.fr](mailto:christine.zanin@univ-paris-diderot.fr)

[nicolas.lambert@ums-riate.fr](mailto:nicolas.lambert@ums-riate.fr)

[ronan.ysebaert@ums-riate.fr](mailto:ronan.ysebaert@ums-riate.fr)

tel. + 33 1 57 27 65 32



# TABLE OF CONTENT

<b>Introduction.....</b>	<b>3</b>
<b>1. Maps and ESPON 2013 (Map-Kits) .....</b>	<b>4</b>
1.1 The "ESPON SPACE" Map-Kit .....	5
1.2 The "ESPON SPACE AND CANDIDATES" Map-Kit .....	8
1.3 The "ZOOM-IN" Map-Kit.....	10
1.4 The "ZOOM-OUT" Map-Kit .....	12
1.5 The "GLOBAL" Map-Kit.....	14
1.6 The "EUROMED" Map-Kit.....	16
<b>2 Enhancing information .....</b>	<b>17</b>
2.1 Differentiation of data type .....	17
2.1.1 Qualitative data .....	17
2.1.2 Quantitative data with absolute values.....	19
2.1.3 Quantitative data with interval or ratio values.....	20
2.1.4 Ordinal or ranked data.....	21
2.2 When using two variations of colour?.....	22
2.3 Choice of data ranges.....	22
2.3.1 Natural Break.....	23
2.3.2 Equal Count or quantile .....	23
2.3.3 Equal Ranges .....	23
2.3.4 Jenks method.....	24
2.3.5 Standard Deviation .....	24
2.3.6 Geometric progression.....	25
<b>3 Maps are tool for communication .....</b>	<b>27</b>
3.1 Bad choices in term of representation of the data .....	28
3.2 Improving the efficiency of the map .....	31
<b>ANNEXE 1 - Relation of graphical variables to perceptual characteristics .....</b>	<b>34</b>
<b>ANNEXE 2 - Numbers of categories that can be perceived at a glance .....</b>	<b>34</b>
<b>ANNEXE 3: Differences in value or lightness .....</b>	<b>35</b>
<b>References .....</b>	<b>38</b>

# Introduction

Maps are a great way of displaying statistical data. It allows summarizing a complex and important information into clear and compact presentation. They can bring a great help in spotting patterns within data.

Maps are accessible for many reasons. People understand maps (at least, think they do). People like maps because they attract attention and brighten up presentation. Nevertheless, and in a scientific versus, the interest of the representation of geographical information on maps can be summarized in three main points<sup>1</sup>.

**The localisation** is the most elementary subject related to geographic information. It allows answering to question "Where can we find this phenomenon?" The precision of the localisation depends on the quality of this kind of information such as statistical databases, statistical yearbook and so on. Locate a geographical object has generally a sense only if it is possible to compare it to other one "Why this object is located here and not there?". Answers can be read off directly from the map without any other help.

**The comparison:** Geographical objects analysis makes a concrete sense when it is possible to compare them. "What is the situation of this region as compare to the other one?"; "Can we observe geographical pattern, such as discontinuities, concentration?" Maps are useful tools for interpreting and pointing out specific geographical patterns, which are impossible to catch with an only statistical analysis.

**Planning:** Since the relations between European territories are very intensive, territorial planning on a special location must interfere with other territories and have to.

Despite many interests to use maps within ESPON, these kinds of documents have also their limits. Maps always generalise and simplify information. Mapping is more than just rendering; it also getting to know the phenomenon which is to be mapped. That's why mapping is not an easy action. Deliver the right message must remain the first objective of map design and mapping allows you to orchestrate the elements of the map to best convey its message to its audience. Thus, the design of maps is mainly concerned with making choices: the choice of mapping method (proportional symbol or choropleth map, isoline or grid map or even a cartogram), the choice of the aggregation level on which information as to be depicted, the choice on the level of statistic areas and the type of data (absolute or relative representation), the choice of graphic variables (such as differences in size, value, grain, colour, direction and shape) to be used. These choices are fundamental's one, they influence people's conception and visualisation of space.

This technical report is not a formal cartography book but allows everyone to understand easily how to produce an effective and operational map in the ESPON 2013 program. The report is organized in 3 parts: (i) Maps and ESPON 2013 (description and explanation of map-kit tool); (ii) Enhancing information (mapping methods and graphic semiology); (iii) Maps and communication (map is to deliver a simple and clear message).

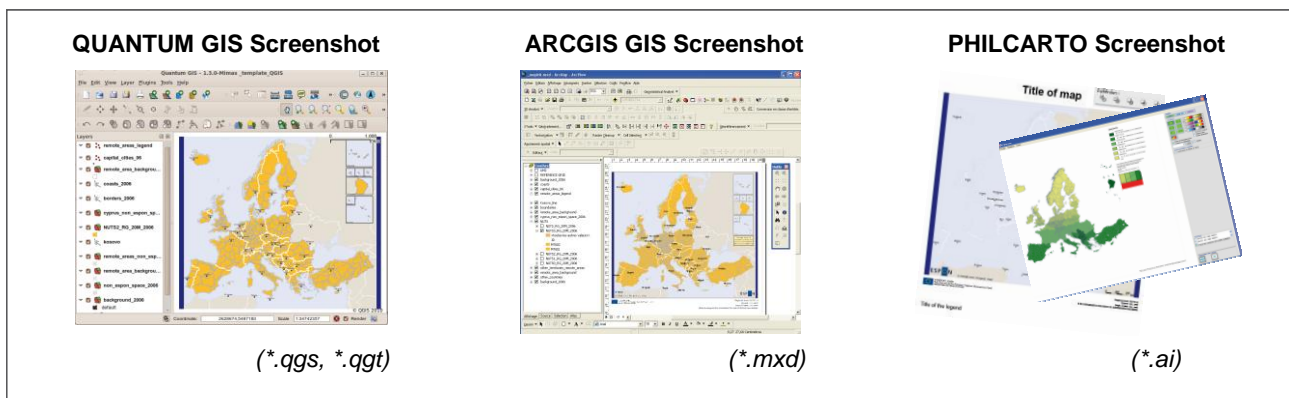
---

<sup>1</sup> Béguin M., Pumain D., 2003, *La représentation des données géographiques – statistique et cartographie*, Armand Colin, 192p.

# 1. Maps and ESPON 2013 (Map-Kits)

The ESPON Map-Kit Tool is composed by a **set of 6 specific mapkits**, adapted to different geographical levels, from local to global. In any case, each mapkit contains geometries (e.g. Eurogeographics, GISCO, ...) and cartographical templates.

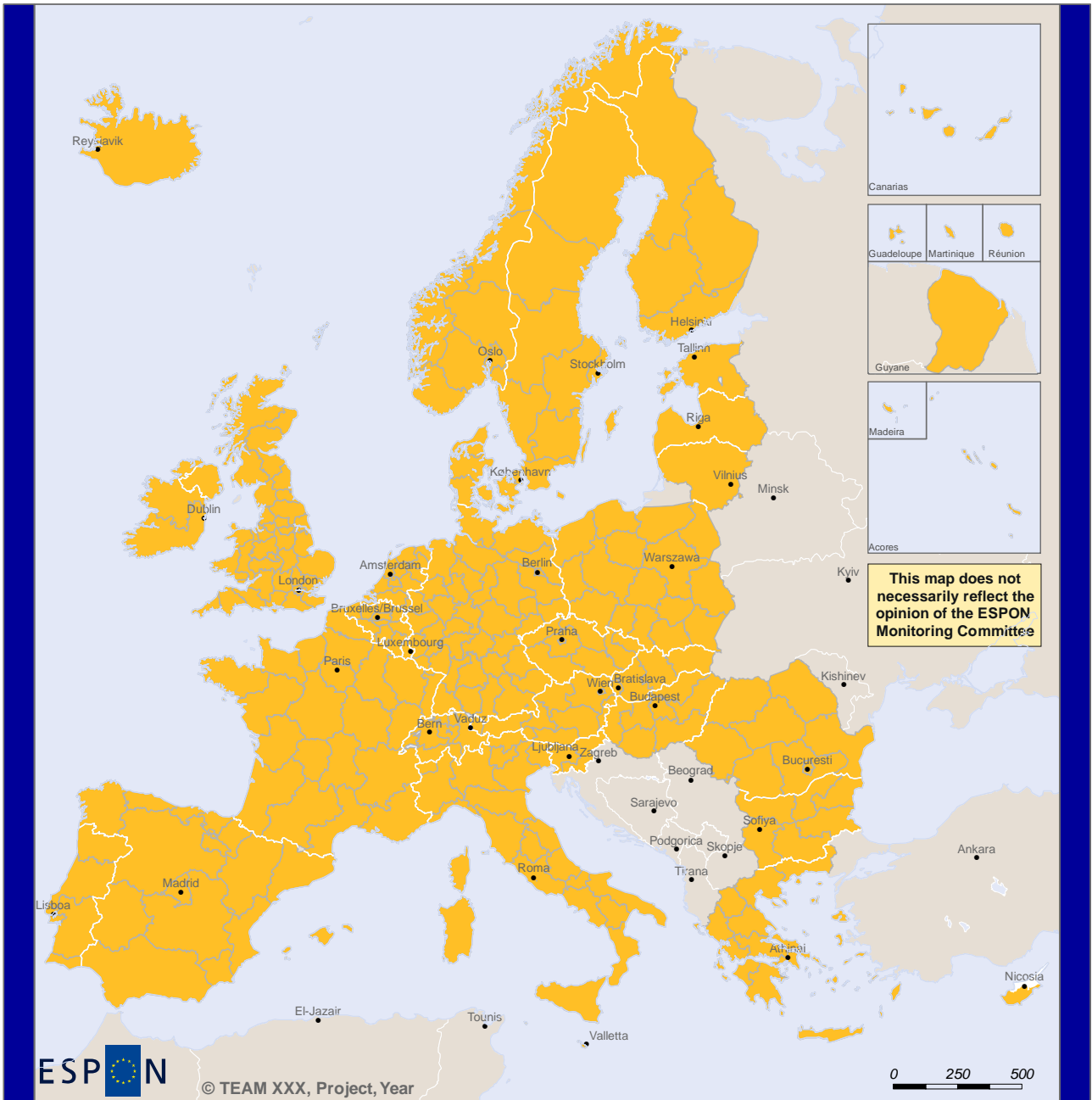
Compatible with the ESPON 2013 database, all these elements are included in an **ARCGIS mxd** document, which is an easy way to make harmonized maps. But, the map kit is also available in an open source format. Quantum GIS (**QGIS**) is a user friendly Open Source Geographic Information System (GIS) licensed under the GNU General Public License. QGIS is an official project of the Open Source Geospatial Foundation (OSGeo). It runs on Linux, Unix, Mac OSX, and Windows and supports numerous vector, raster, and database formats and functionalities. It is possible to download the application on the following URL: <http://www.qgis.org/en.html>. It is also possible to use a third application to make thematic maps: **Philcarto**. It is not a GIS application, is it a free tool dedicated to thematic mapping and spatial analysis. The application and the documentation is downloadable on this URL: <http://philcarto.free.fr/Inscriptions.html>



On each map kit, some elements absolutely have to appear on the map layout

- The EC publicity requirements, with the following sentence: "EUROPEAN UNION; Part-financed by the European Regional Development Fund; INVESTING IN YOUR FUTURE".
- The ESPON logo.
- The MC disclaimer: "This map does not necessarily reflect the opinion of the ESPON Monitoring Committee".
- Team, Project, Date.
- The Regional level and the NUTS version (e.g. NUTS3 2006).
- Data sources (e.g. ESPON 2013 DATABASE).
- Origin of data (e.g., European communities, June 2009) Eurogeographics copyright: © Eurogeographics Association for administrative boundaries..
- The Scale.

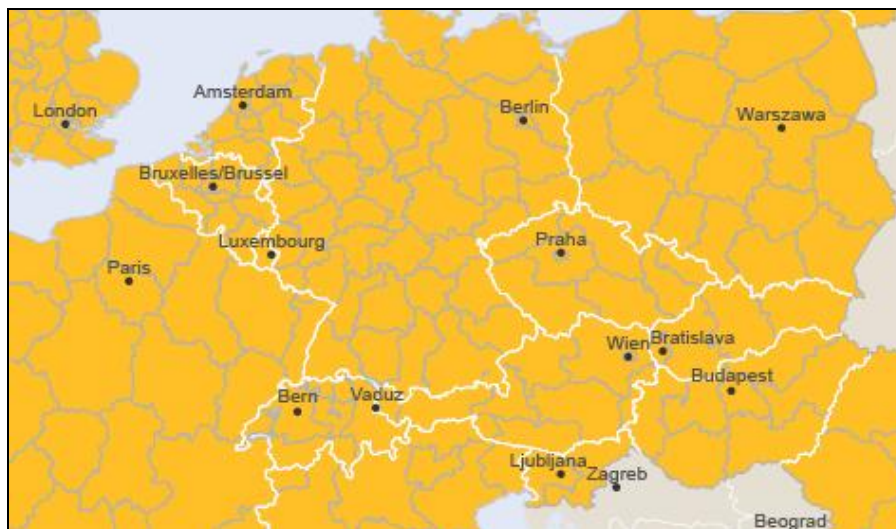
# 1.1 The "ESPON SPACE" Map-Kit




**EUROPEAN UNION**  
 Part financed by the European Regional Development Fund  
**INVESTING IN YOUR FUTURE**

Regional level: NUTS xx  
 Source: xxx, year  
 Origin of data: xxx, year  
 © EuroGeographics Association for administrative boundaries

- The “ESPON SPACE” map kit is used to make maps on the ESPON area defined in the current program, which is composed by all the Member States of European Union (27 countries) plus Switzerland, Norway, Iceland and Liechtenstein.
- The projection is based on the ETRS-LAEA system: ETRS89 Lambert Azimuthal Equal Area Coordinate Reference System. This projection is the standard in Europe for pan-European statistical mapping at all scales. In particular, this projection is used by the European Environment Agency. Parameters: latitude of origin 52° N, longitude of origin 10° E, false northing 3 210 000.0 m, false easting 4 321 000.0 m. (EPSG code: 3035)
- On the map, 43 capital cities have to be written on the map. Vilnius, Minsk, Dublin, Berlin, Amsterdam, Warszawa, London, Bruxelles/Brussel, Kyiv, Praha, Paris, Wien, Budapest, Bern, Beograd, Bucuresti, Sofiya, Tirana, Madrid, Ankara, Helsinki, Zagreb, Nicosia, Luxembourg, Bratislava, Tallinn, Sarajevo, Skopje, Athina, Kishinev, K°benhavn, Lisboa, Oslo, Reykjavik, Riga, Roma, Stockholm, Valletta, Ljubljana, El-Jazair, Tounis, Podgorica and Vaduz.



The localisation of each capital city is shown by a black bullet point. Except for Malta, the name of the city is always above the bullet point. For a better visibility, Valetta is written slightly on the right of the bullet point.

- Remote territories of France (Martinique, Guadeloupe, Guyane française, Réunion), Spain (Canarias) and Portugal (Acores, Madeira) are territories members of the European Union. They have to be represented on maps even when data are not available.

- Cyprus is represented in two different colours. The North area appears in white as “no data”.

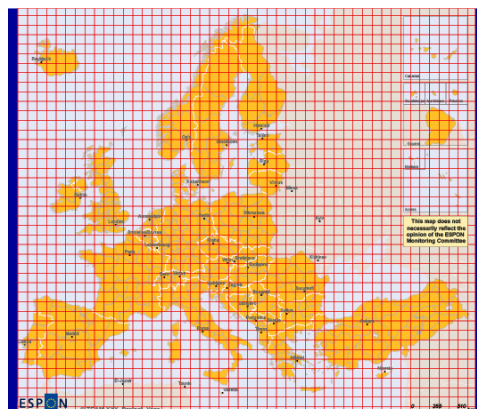


- To ensure the visibility of Malta islands on the maps, we do not use light blue coast-line that could be reduce, on the map, the size of

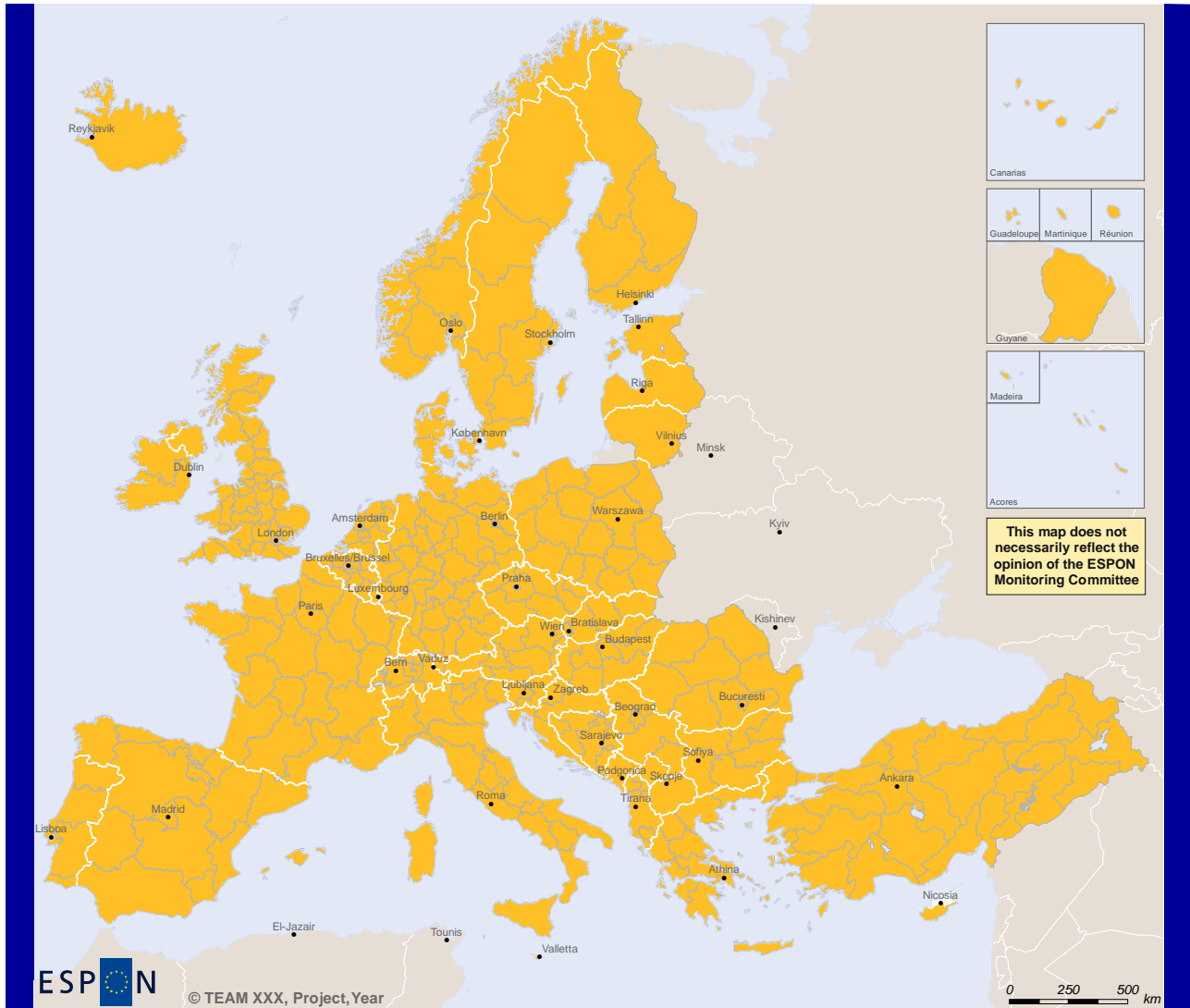


this country. Moreover, the line of the Malta polygon is drawn as thinner as possible.

- The fact that the projection of the MAP KIT is the same as the projection used by EEA (EPSG 3035) ensure the compatibility between EEA reference grids and the ESPON template defined. These grids are included in the Map Kit. As a consequence, it is possible to use them in the same European Map Kit as previously.



## 1.2 The "ESPON SPACE AND CANDIDATES" Map-Kit



This map does not necessarily reflect the opinion of the ESPON Monitoring Committee


 EUROPEAN UNION  
 Part-financed by the European Regional Development Fund  
 INVESTING IN YOUR FUTURE

Regional level: NUTS xx  
 Source: xxx, year  
 Origin of data: xxx, year  
 © EuroGeographics Association for administrative boundaries

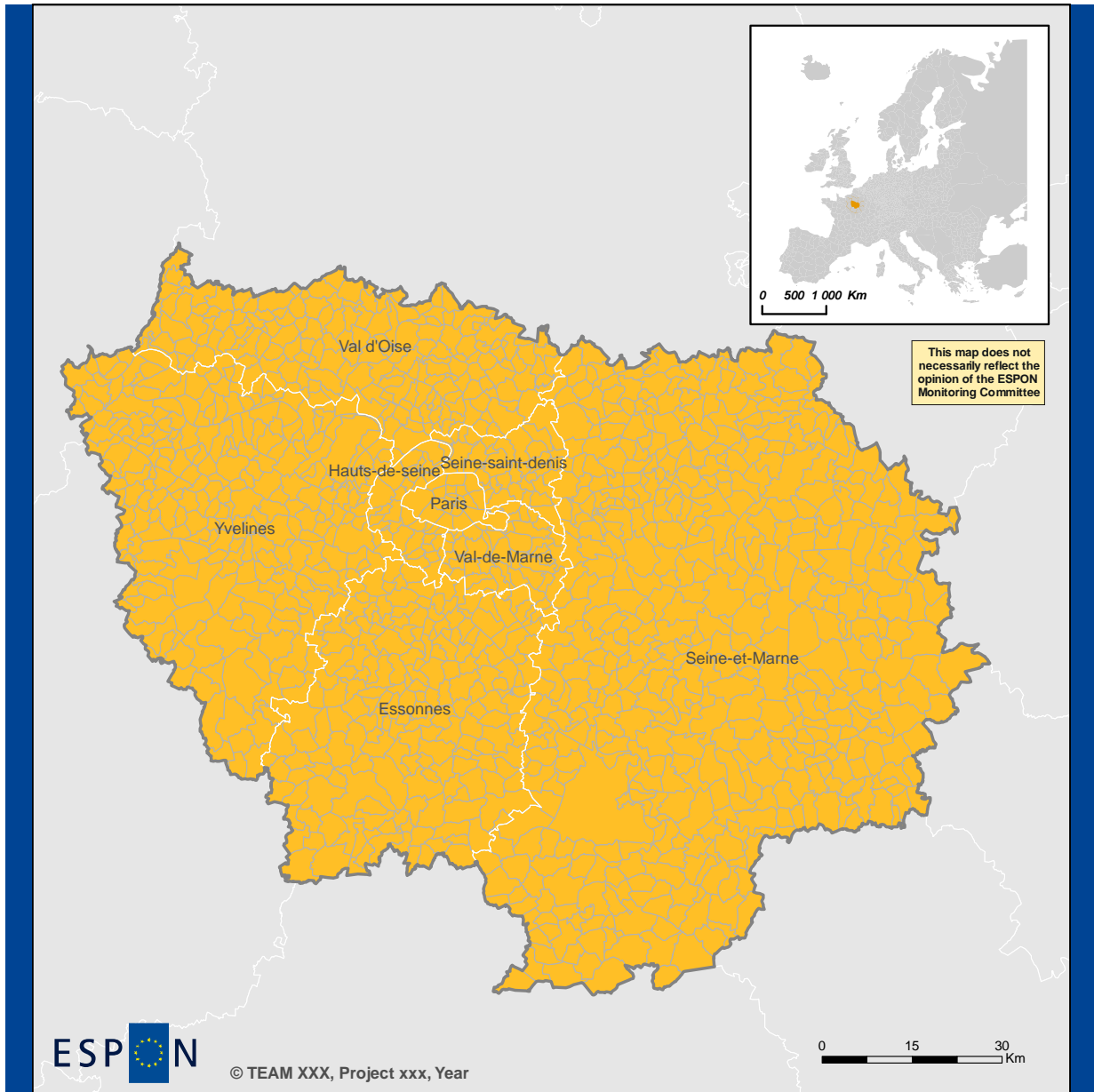
- More than the ESPON area, the map-kit includes the candidate countries and the Western Balkans. For Candidate Countries (Croatia, FYROM, Turkey), the NUTS system already exists. But, concerning the western Balkans, a system named "SIMILAR TO NUTS" has been created (Albania, Kosovo, Montenegro, Bosnia-Herzegovina, Serbia).



- Concerning the rules of cartography, drawing of borders, for some countries, must follow precise rules for political reason. In general, ESPON follows the rules established by European Commission. When these rules do not exist at EU level (for example because of lack of consensus) the rules of UN are used as reference. According to these considerations, we have to use always the reference to the UN resolution when referring to Kosovo, i.e. under UNSCR 1244/99. On the map, the borders of Kosovo are thinner (0.20 pt) than the other boundaries (0.30 pt) and the name of the city of Pristina is not written.
- The projection is based on the ETRS-LAEA system: ETRS89 Lambert Azimuthal Equal Area Coordinate Reference System (EPSG code: 3035).
- Classically, the map has to content the ESPON elements (disclaimer, layout, data source, scales...).



### 1.3 The "ZOOM-IN" Map-Kit



 EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

Local level: LAU2  
Source: xxx, year  
Origin of data: xxx, year  
© EuroGeographics Association for administrative boundaries

- For Priority 2 projects, a “Local” map-kit has been created.
- This template consists in 2 parts: the main map and the localisation map. The main map is an extract of the LAU2 base map. The map of localisation indicates where exactly the study area is located in Europe. This element is positioned on the top-right of the template.
- To be useful, this map have to be represented with the good local or national projection (e.g. for France: LAMBERT93).
- Classically, the map has to content the ESPON elements (disclaimer, layout, data source, scales...).

## 1.4 The "ZOOM-OUT" Map-Kit



 EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

Local level: xxx  
Source: xxx, year  
Origin of data: xxx, year  
© ESRI for administrative boundaries

In order to allow comparisons between Europe and other countries of the World at regional level, a specific map-kit has been created.

- Take China, for example, its template must be divided in two parts:  
The main map shows the case study at regional level and its neighbourhood (Mongolia, Russia, India etc. in grey). The localisation map on the top right of the template indicates the location of the study area in the World. The localisation map is based on orthographic projection.
- For any other part of the World you must adapt the template to your case study (geographical coverage, adapted projection...)
- Classically, the map has to contain the ESPON elements (disclaimer, layout, data source, scales...).

## 1.5 The "GLOBAL" Map-Kit

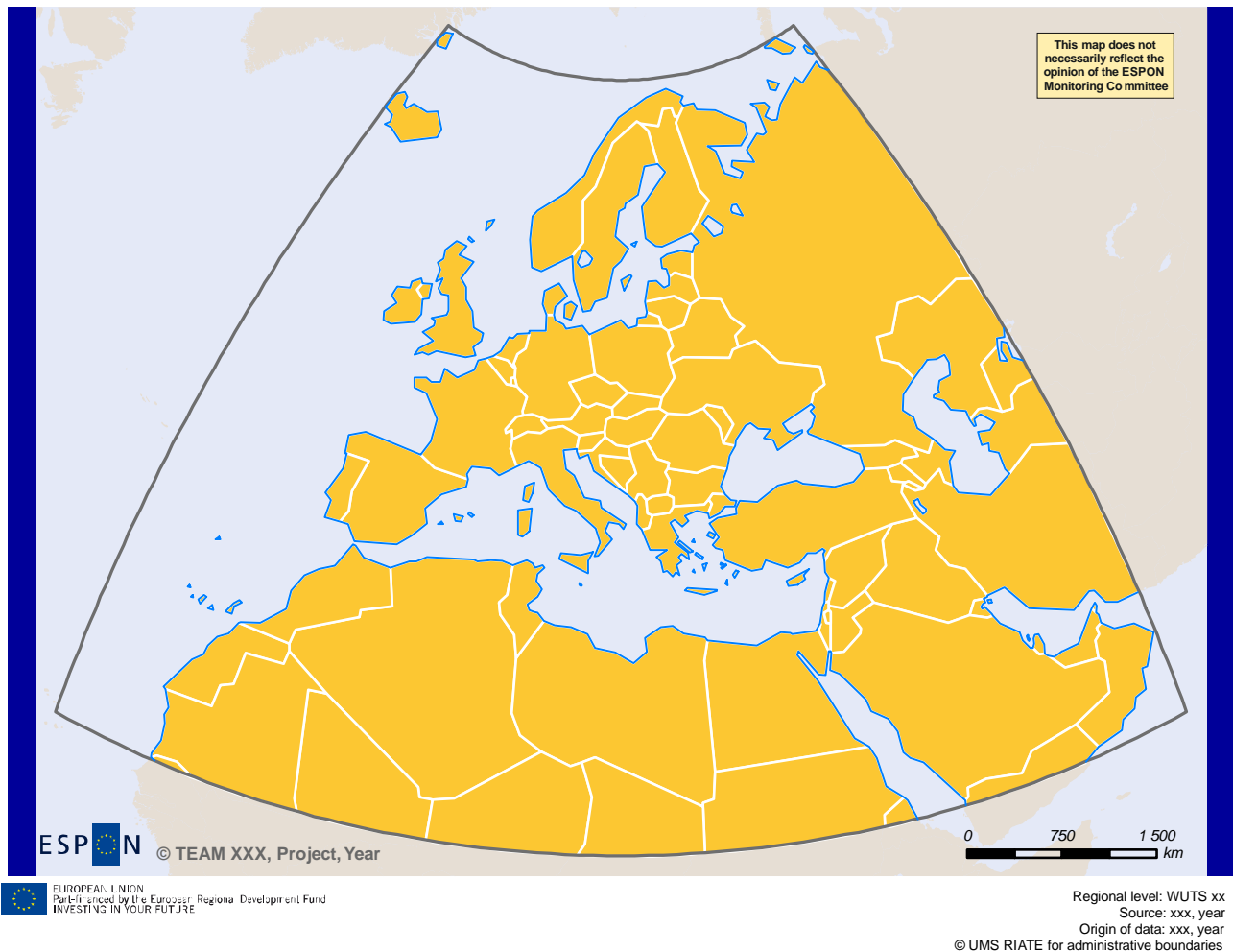


EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

Source: xxx, year  
Origin of data: xxx, year  
© UMS RIATE for administrative boundaries

- Developed during the first period of the program by the project ESPON 3.4.1 "Europe in the World", this specific map-kit is also available. It will probably be improved by teams in charge of the project "Continental territorial structures and flows (Globalisation)".
- The map kit includes different kind of administrative boundaries compatible with different world databases:
  - WUTS: The World Unified Territorial System is a hierarchical system of division of the world. Developed in the project ESPON 3.4.1, this delineation is compatible with the World database of the ESPON 2006 program.
  - CHELEM: The CHELEM is a database constructed by the CEPII. It contains 3 specific databases (The GDP Database, The International Trade Database and the Balance of Payments Database). In the map kit, the CHELEM delineation fit to theses 3 databases.
  - WORLD\_24\_UNIT: This geometry is composed by a partition of the world in 242 territorial units. Consequently, you can easily use this partition with a lot of world databases (e.g UNEP data portal, IMF, UNPP, WDI).
- The projection is based on a north polar projection (lambert azimuthal equal area)
- Classically, the map has to content the ESPON elements (disclaimer, layout, data source, scales...).

## 1.6 The "EUROMED" Map-Kit



- Developed during the first period of the program by the project ESPON 3.4.1 "Europe in the World", this specific map-kit aims to represent information on the pan-European area. This area, named Euromed, is composed by 74 countries, from Niger to Russia.
- Based on the WUTS system, this map kit proposes also a hierarchical system of the world territorial delineations.
- The projection is based on Lambert Azimuthal Equal Area (18°E, 50°N)
- Again, the map has to content the ESPON elements (disclaimer, layout, data source, scales...).

## 2 Enhancing information

### 2.1 Differentiation of data type

Many possibilities exist to show data on map. Choosing relevant representation is not an obvious task and has to be considered seriously. Indeed, choosing the wrong type of map can completely misrepresent the data. It is important to keep in mind that **the choice in cartography is always dependant on the type of data**. It is possible to identify four main types of data:

1. Qualitative data
2. Quantitative data with absolute values
3. Quantitative data with ratios values
2. Ordinal (or ranked) data

For each type of data it is possible to relate it to a **geographical reference: points, lines or areas**.

There are many possibilities to show correctly data on maps. The aim of this paper is not to present all types of correct visualisation, but an extract of the most usual and efficient ones.

#### 2.1.1 Qualitative data

A data is qualitative when its value is a nominal one with qualitative differences: components do not allow establishing range relations between them.

For example, considering the different geographical references:

Points: location universities by type (university, polytechnics...) – **Figure 1**

Lines: communication network without hierarchy (ferry connections, main roads) – **Figure 2**

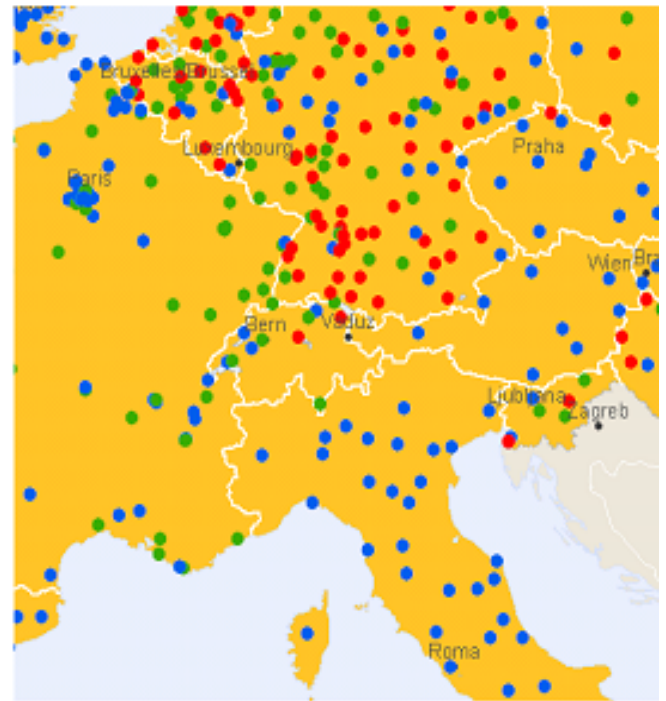
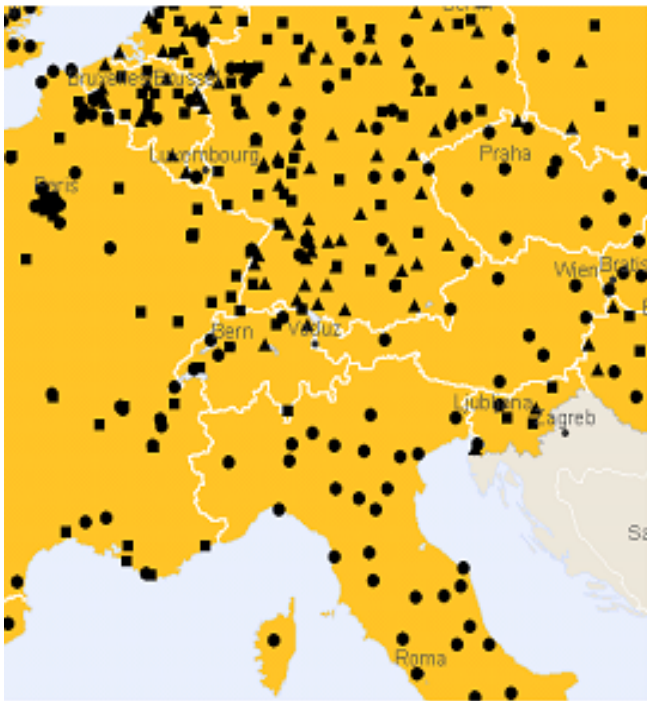
Areas: results from typology (rural area, urban area...) – **Figure 3**

Qualitative data have to be shown such a manner that do not suggest rank either quantity. Two possibilities: use **geometric symbols** or **differential colour** in order to **differentiate** the different elements of the map.

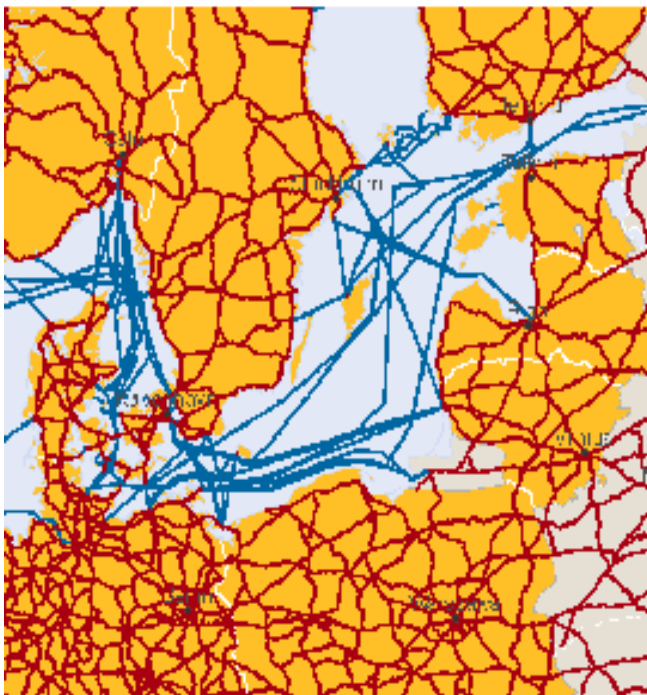
With points (figure 1) the most efficient is to show information by colour or geometric symbols. It is important to use a limited quantity of symbols or colours to make the map understandable.

For lines or areas, differential colours should be used (figure 2 and 3).

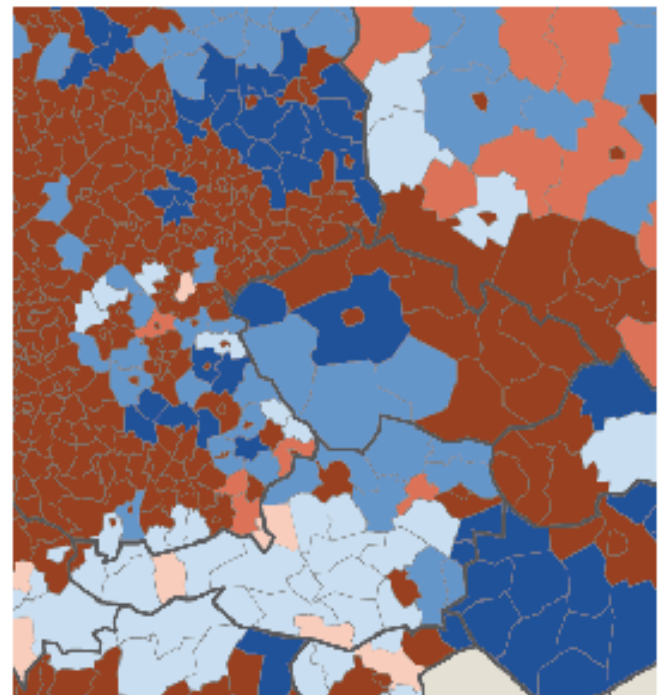




**Figure 1 - Universities by types - Two possibilities**  
**Good map = points + symbols or points + colours**



**Figure 2 - Mains roads and ferry connections**  
**Good map = line + colours**



**Figure 3 - Results from Urban-Rural typology**  
**Good map = areas + colours**

## 2.1.2 Quantitative data with absolute values

Quantitative data with absolute values means concrete **quantity**; the sum of the different values can be calculated and has a real sense. For example, population, GDP, CO2 emissions are absolute quantitative data if we consider the number of inhabitants, number of euros or tons of gas emissions.

For example, considering the different geographical references:

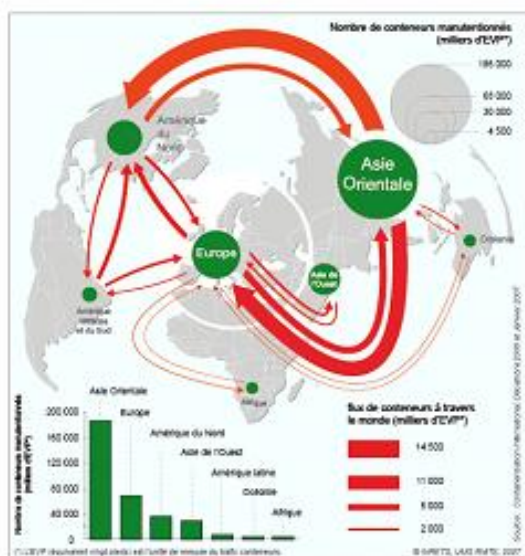
Points: Cities of Europe (number of inhabitants)

Lines: Containers flows across the world (millions tons) – **Figure 4**

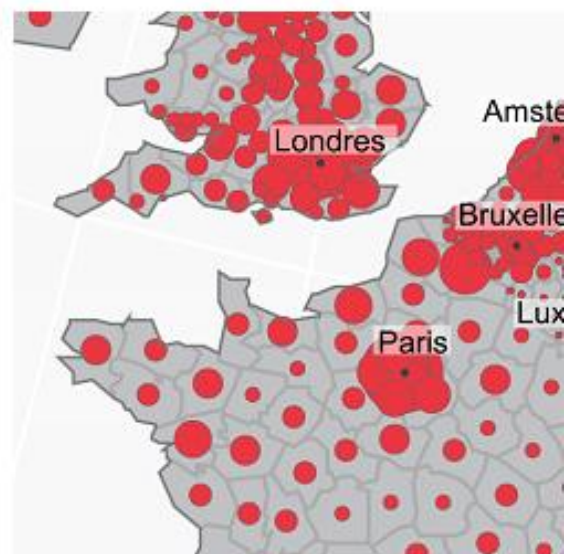
Areas: Population of NUTS 3 – **Figure 5**

Whatever the type of geographical objects (points, line, areas), the cartography of quantitative data with absolute values has to **respect the quantity** and differences of proportionality. For points or areas objects, the most common representation is to use maps with area **proportional circles**. The circled area is proportional to the size of the data value.

The map showing data in line format (**figure 4**) has to use lines of different width. The width of the line is proportional to the data value.



**Figure 4 - Containers flows across the World**  
Good map = line + variation of line size



**Figure 5 - Population of NUTS 3 in Europe**  
Good map = dot + proportional variation of size

### 2.1.3 Quantitative data with interval or ratio values

The ratio values are calculated and expressed a series of ratios or proportional values, such as percentage, per km, per inhabitant. This kind of data is the most common.

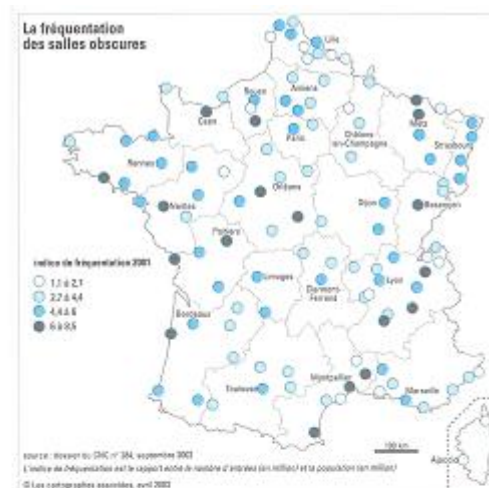
For example, considering the different geographical references:

Points: Cities of France (cinema attendance index) **Figure 6**

Lines: GDP per inhabitants discontinuities (relative difference between two territories) – **Figure 7**

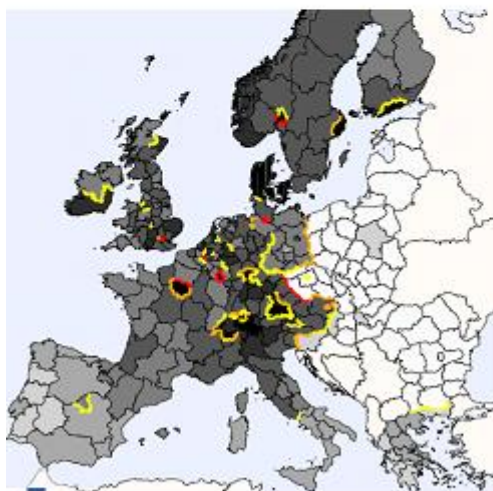
Areas: Abstention, European elections 2009, in Ile-de-France municipalities – **Figure 8**

For ratios values, the most relevant representation is a choropleth map where density is linked to the class of the data value for each area. The efficiency of the map depends on the range between the least dense (lightest) area and the densest (darkest) area. When correctly applied, percentage or densities that are twice as high are represented by a grey value that is twice as dark.



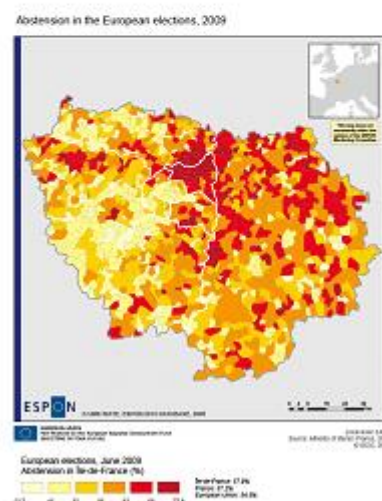
**Figure 6: Cinema attendance Index in French main cities**

*Good map = dots + variation of colours*



**Figure 7: GDP per inhabitants discontinuities**

*Good map = lines + variation of colours*



**Figure 8: Abstention European votes in Île-de-France**

*Good map = area + variation of colours*

## 2.1.4 Ordinal or ranked data

Ordinal data are categorical data where there is a logical ordering to the categories. A good example is the Likert scale that you see on many surveys: 1=strongly disagree; 2=Disagree; 3=Neutral; 4=Agree; 5=strongly agree. Another example could be found with modalities like first, second, third etc., or small, medium and high.

For example, considering the different geographical references:

Points: Typology of Functional Urban Areas – MEGA, national FUA, regional FUE

### **Figure 9**

Lines: Road hierarchy – **Figure 10**

Areas: Degree of policentricity – **Figure 11**

The representation of these data is based on the expression of natural modalities order. Considering the different geographical references (point, line or area) you can only use 2 graphics variables: grey value or the intensity of a colour. They allow denoting differences in intensity of a phenomenon and expressing order between geographical areas, points or lines. Because differences in grey value or in intensity of colour are used, a hierarchy or order between ordinal modalities can be perceived.



**Figure 9:** Typology of Functional Urban Areas  
Good map: points + variation of colour (or size)



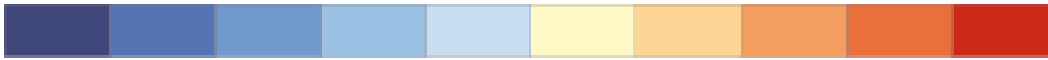
**Figure 10:** Road hierarchy in Europe  
Good map: lines + variation of colour



**Figure 11:** Degree of policentricity in Europe  
Good map: areas + variation of colour

## 2.2 When using two variations of colour?

It is sometimes necessary to show a phenomenon by a variation of two colours fundamentally different:



This kind of representation is very useful since it allows making more differentiation between the classes of the map. However, it is possible to use these oppositions of colours only if the **break has an objective sense** in the dataset, for instance:

- Opposition between negative and positive values (decrease and increase of population between two periods)
- Values above/under the average value or median value of the dataset (level of accessibility above or under the EU27 average)
- Values above/under a value which have a concrete reality (unemployment rate under/above the threshold of 10 %).

Opposition of variation of two colours should be used only for quantitative data with ratio values and ranked data.

To ensure the **harmonisation** of all maps produced by ESPON projects, it is important that also the use of colours is being guided in the case of opposite colours. In general, it is advised not to combine red and green in one map in order to serve the colour-blind people. Other general rules do not exist. The choice of opposite colors is very subjective and cultural. However, it is quite confusing if two different ESPON maps are published where red has a positive meaning in one of the maps and a negative meaning in the other map. Therefore, in the case of ESPON maps with opposite colours, it is decided to have the following principle as guideline:

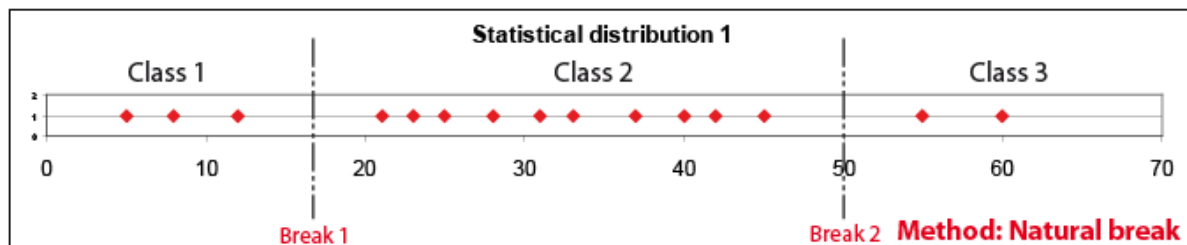
When combining red (warm colours) and blue or green (cold colours), **red is 'not good'/'negative'** and **blue is 'good'/'positive'**

## 2.3 Choice of data ranges

Nevertheless, this kind of representation introduces always a **loss of information** since it transforms a complex statistical distribution into a limited number of classes. Information becomes more generalised and simplified. The accuracy of original values is lost, but **this operation is needed in order to present a synthetic overview of the dataset**. Indeed, a good class division will focus on what is the main content of the dataset, and minimise the loss of accuracy by generalisation. Further below you will find five different classes dividing methods ranging data values. Of course it is also possible to combine different methods, in particular when there are an important number of records. This step before mapping is needed for quantitative values with ratios only.

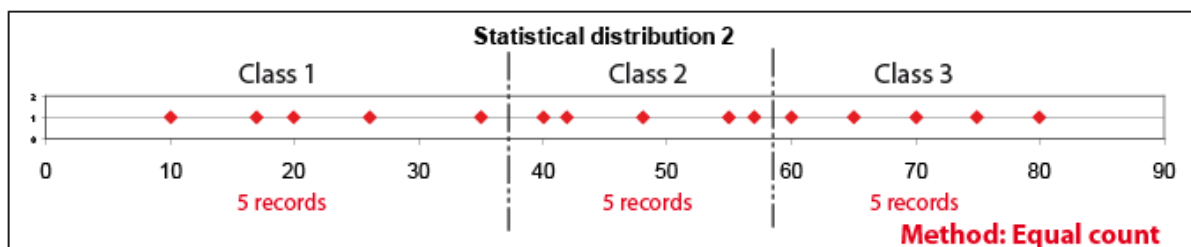
### 2.3.1 Natural Break

This method sets the breakpoint to “natural points” in the dataset. The strength of this method is that it increases the information content. **This method is suited when important breaks** describe the dataset.



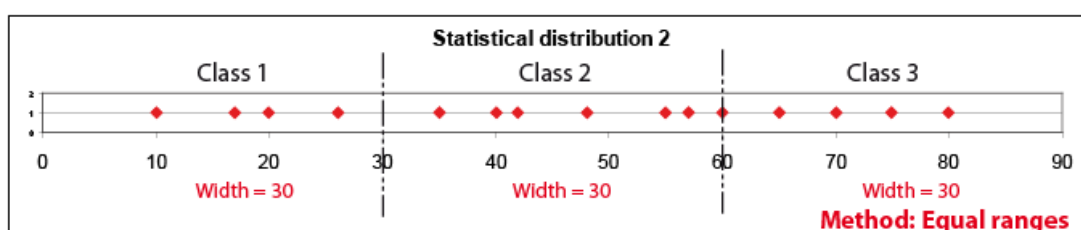
### 2.3.2 Equal Count or quantile

**Equal range contains approximately the same number of records.** With 5 classes, each contains 20 % of the total number of the data values. **This method is suited for comparing one dataset with datasets from other themes.** If the data deviate from a linear distribution, the absolute class width will show large variations. Equal count methodology does not take into account exceptional values in the distribution.



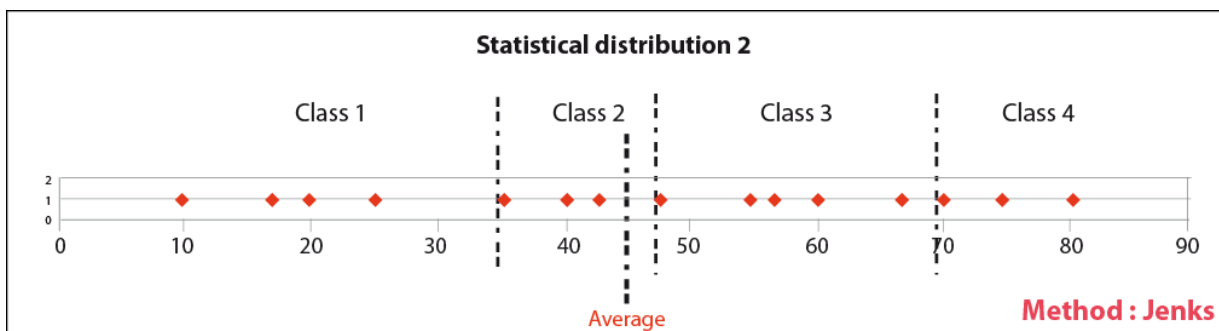
### 2.3.3 Equal Ranges

**The difference between the top and bottom values in each range is the same.** This means that we can use values like 0-20; 20-40 etc. or calculate the width of the dataset, and divide by the number of classes wanted. In this case the lowest class will start with the lowest value; the width between the classes will be the same, and the top of the highest value in the dataset. **This method is suited for datasets with a smooth linear distribution.** If the method is used on dataset that are not linear distributed, you will have some classes with many values and others with few or no values.



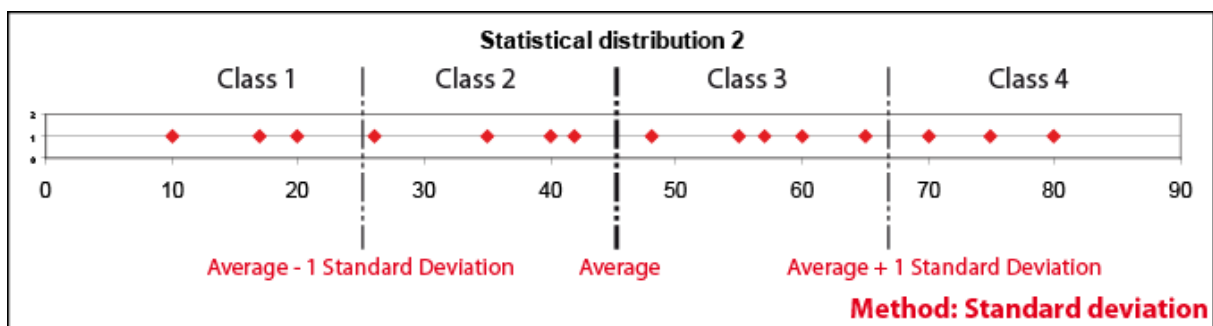
### 2.3.4 Jenks method

The Jenks' optimization method is also known as the goodness of variance fit (GVF). It is a data classification to determine the best arrangement of values into classes. It is used to minimize the squared deviations of the class means while maximizing each class's deviation from the means of other groups. The method first specifies an arbitrary grouping of the numeric data, and then observations are moved from one class to another ensuring maximum homogeneity within groups and maximum heterogeneity between groups.



### 2.3.5 Standard Deviation

**The class borders are calculated from the mean value and the standard deviation.** Standard deviation is a way to describe statistical dispersion. The width of the class is equal to the standard dispersion (or an half depending on the number of classes expected). **This method is suited for normal distributed datasets only.**



### 2.3.6 Geometric progression

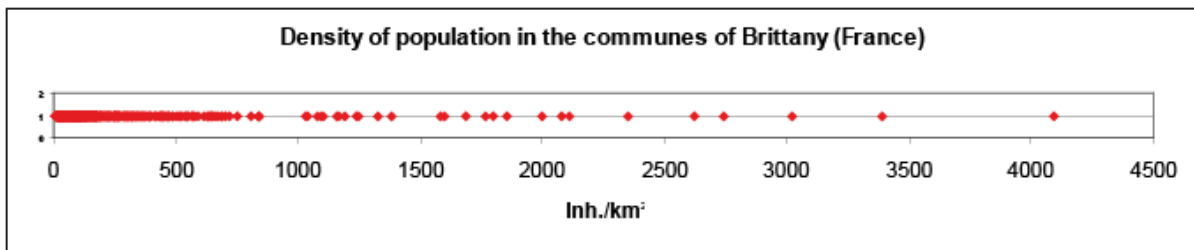
**The widths of the class follow a geometric progression.** To calculate the width of the different class, it is necessary to estimate the geometric ratio, such as:

$$\log R = (\log_{10} \text{Max} - \log_{10} \text{Min}) / \text{number of classes wanted}$$

$$R = 10^{\log r}$$

Width of the Classes = (min, min x R); (min x R; min x R x R) and so on.

**This method is suited for uneven distribution** and particularly distribution described by a lot of low values and few high values, such as density of population distribution.



From the example of Brittany, the data ranges, following the geometric progression, should be in 6 classes:

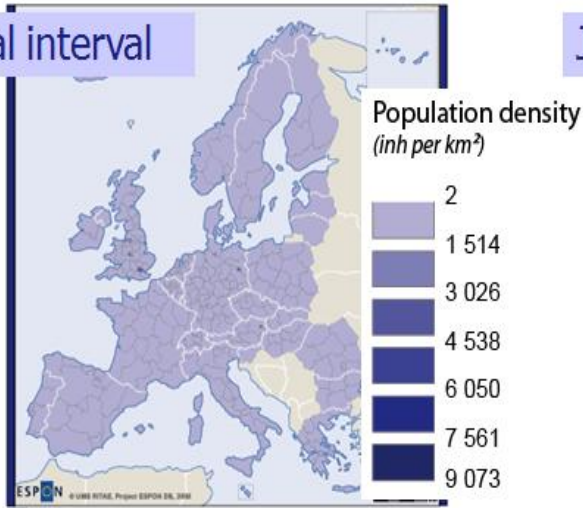
Class	Class boundaries	Number of communes
1	[9;25[	128
2	[25;70[	626
3	[70;190[	343
4	[190;525[	117
5	[525;1470[	39
6	[1470;4100[	15

**Whatever the method chosen for ranging the distribution, it is important to use smooth values for the break, in order to understand and memorize easier the sense of the map, e.g. use 30 instead of 29,77; 1500 instead of 1508 etc.**

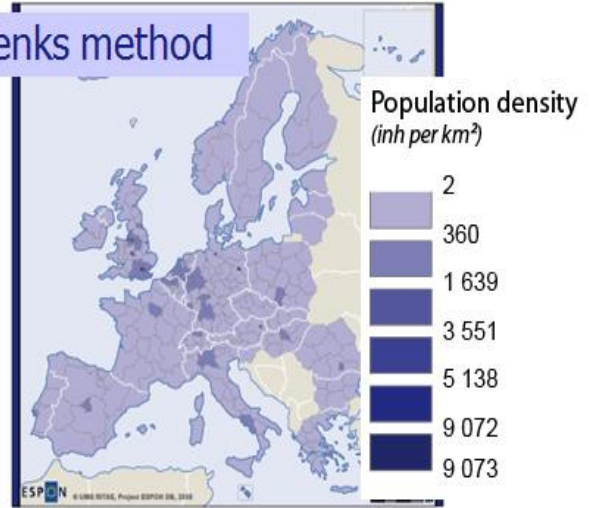
**Figure 12** shows the importance of the choice of data range on the visualisation of phenomena.



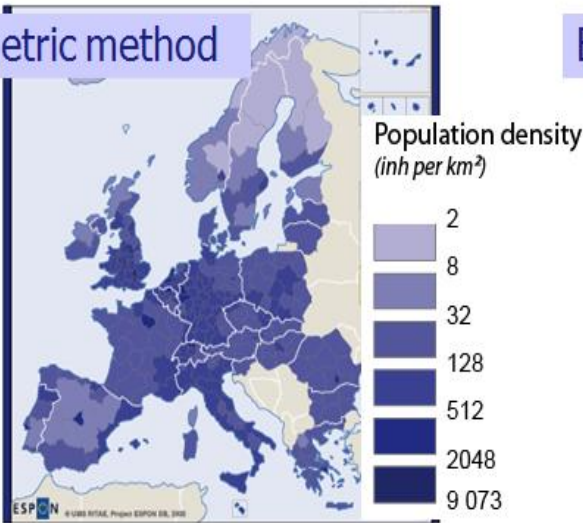
Equal interval



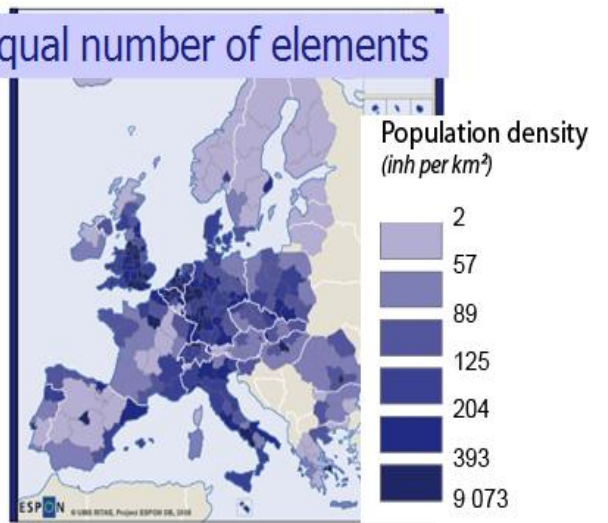
Jenks method



Geometric method



Equal number of elements



**Figure 12:** Result and efficiency are dependent upon the data classification method

### 3 Maps are tool for communication

As we explain in the introduction of this technical report: "Maps are perhaps as fundamental to society as language and the written word. They are the preeminent means of recording and communicating information about the location and spatial characteristics of the natural world and of society and culture<sup>2</sup>".

Maps are produced all over the world and used by people as different as scientists, researchers, scholars, governments or businesses. These maps are most of the time statistical ones connected with the environment, the economy, the politics, the society etc.

**The biggest strength of these maps is to allow an effective and relevant communication of the information.** However, cartography is a special type of visual communication that does require some preliminary learning: a special purpose language for describing spatial relationships. "The analogy with language also helps explain why training in principles of effective cartography is so important--it allows us to communicate more effectively. Without knowledge of some of these basic principles, the beginning cartographer is likely to be misunderstood or cause confusion<sup>2</sup>".

Of course, cartographers must pay special attention to coordinate systems, map projections, and issues of scale and direction but that's not the first issue of map as a tool for communication. Maps are symbolic abstractions and representations. **The first question when mapping is related to know how to simplify, generalize, represent and symbolize the relationships being represented with graphics symbols.** In other words, what is a good map?

If a design is always more effective than a long speech, the measure of a good map is how well it conveys the right information to its readers and how well it communicates with its audience. This raises a series of questions that must be addresses at the start of a map conception: What is the motive, intent, or goal of the map? Who will read the map? Where will the map be used? What data is available for the composition of the map?

Beyond aesthetic characteristics, the communication also passes by a complete and effective layout: some elements must appear within the base map and the thematic representation, a complete legend, explicit title and source, a precise date of data or even a scale.

---

<sup>2</sup> Kenneth E. Foote and Shannon Crum, *The Geographer's Craft Project*, Department of Geography, The University of Colorado at Boulder

From data to map, 7 fundamental goals need to be identified to realize a good map:

1. Identify the goal of the map;
2. Identify the audience of the map and where it will be used;
3. Identify the information to be communicated;
4. Identify the geographical reference (point, line or area?);
3. Choose the base map (map projection and scale);
4. Choose the visual variable (symbolic graphic language);
5. Choose layout and identify all the elements to be added.

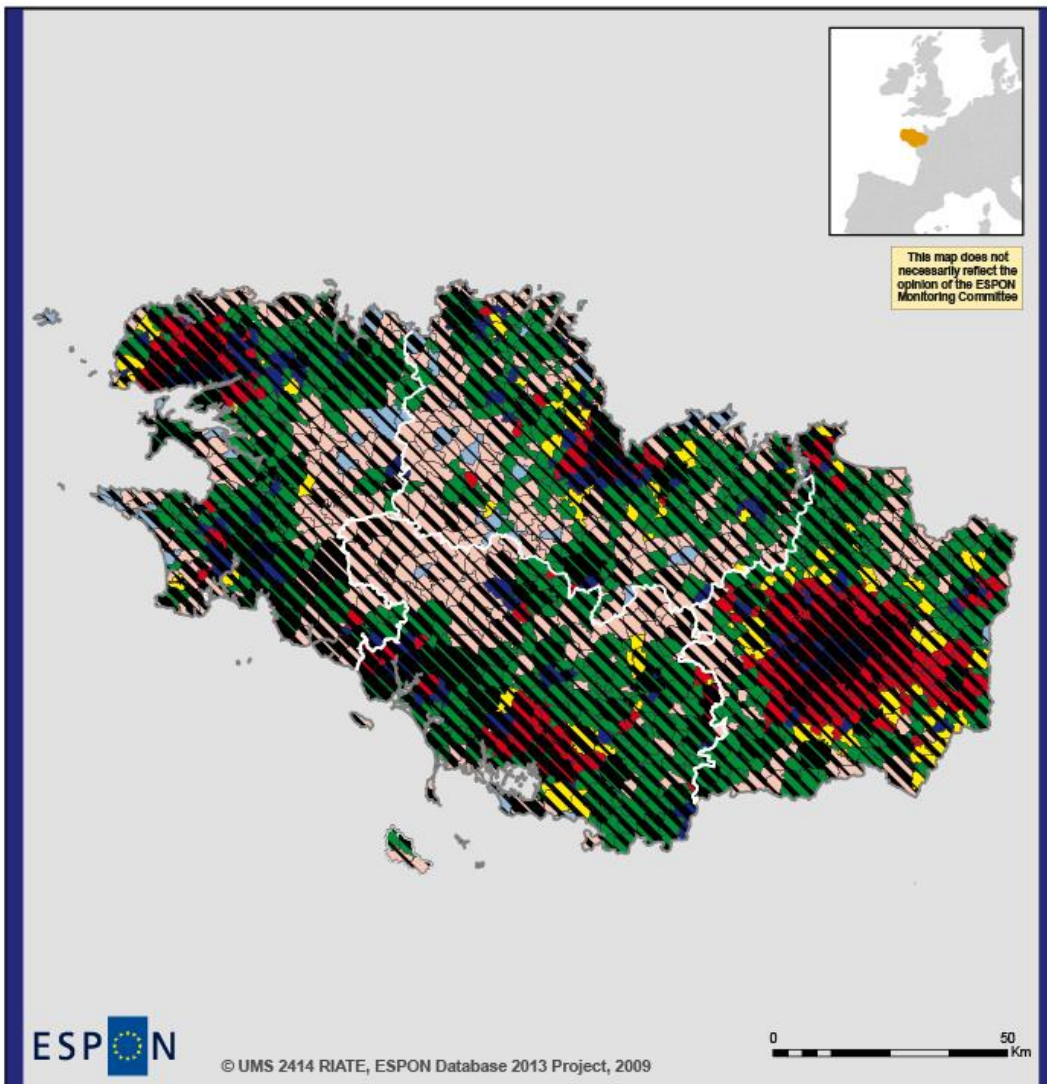
When these different elements are not correctly taking into account, the map will be characterised probably by some mistakes and misunderstandings.

### 3.1 Bad choices in term of representation of the data

Most of the problems of visualisation and map design are generally linked to **bad choices in term of representation of the data** (cf. part 2 of the technical report). When comparing **figures 13 and 14**, which represent the same information, e.g. a typology showing age structure and total population in the municipalities from Brittany (France), it is quite clear that the second map is really clearer than the first one. Two main reasons can explain it (**figure 13**):

- Absolute values (e.g. total population in 2000) don't have to be shown by variation of intensity of black (hachure). This kind of representation does not respect the ratio of proportionality of the indicator, which is fundamental and needed information. Using hachure is also a visual mistake; the map is not readable at all and the representation is not the most efficient. These data have to be shown by proportional symbols, circles for instance.
- This typology, derived from age structure cannot be considered as a qualitative data, since there is an implicit order when considering the progression in term of age. In concrete terms, showing each class by a different colour is not the best solution. To show correctly this data it is important to think about the goal of the map. Here, it is important to represent the municipalities described by high share of young, active and old people. As a consequence, it is important to differentiate these information (3 colours) and also to make possible the analyse of the graduation of the phenomenon (high/medium shares), e.g. using variation of intensity of these 3 colours.

The solution proposed in **figure 14** try to correct these different elements. The most adapted solution for the representation of these data is to combine circles and colours in order to make the map as clear as possible. On top of that, it allows nuancing the interpretation of the map, e.g. Brittany is a region where ageing is important, but it concerns specific small and rural cities.



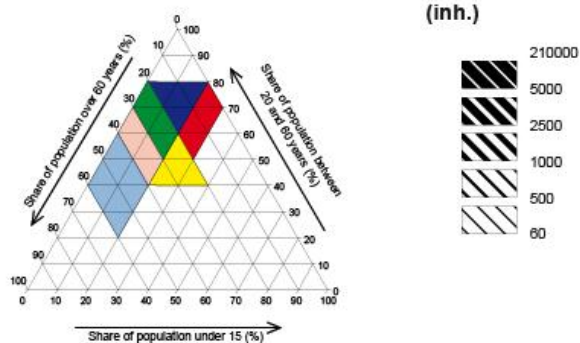
EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

Local level: LAU2  
Source: UMS 2414 RIATE, 2009  
Origin of data: INSEE, 2009  
© EuroGeographics Association for administrative boundaries

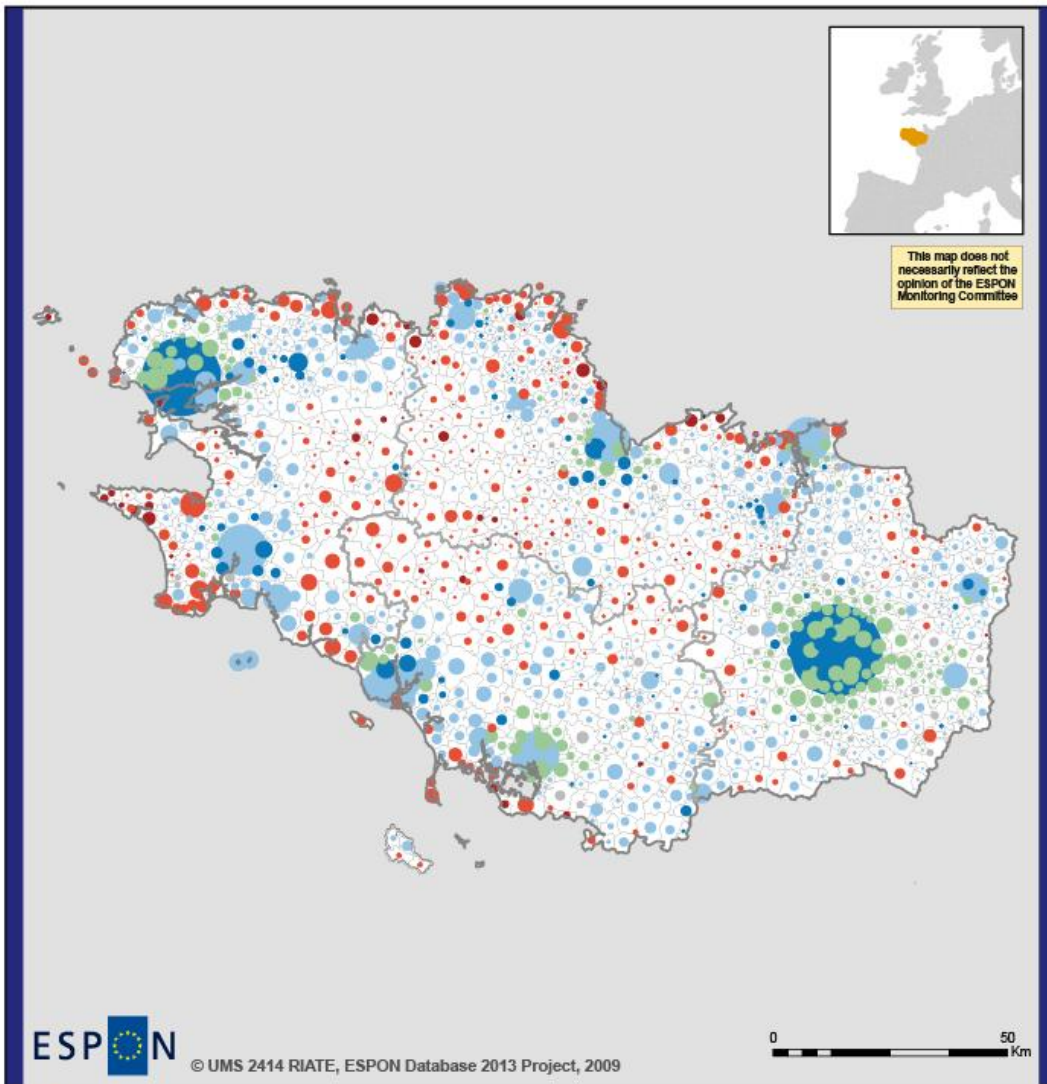
**TYPE OF AGE STRUCTURE IN 2000**

- A) Excedent of young population
  - Type A.1
- B) Excedent of active population
  - Type B.1
  - Type B.2
- C) Excedent of old population
  - Type C.1
  - Type C.2
- D) Medium profile
  - Type D

**TOTAL POPULATION IN 2000  
(inh.)**



**Figure 13: Population and age structure in Brittany (France) – with semiologic problems**

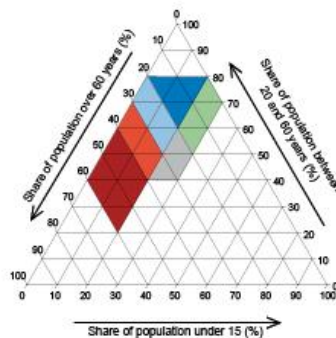


EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

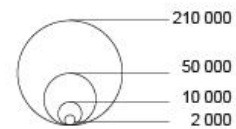
Local level: LAU2  
Source: DG-IPOL, *Shrinking Regions: a paradigm shift in demography and territorial development*, European Parliament, 2008  
Origin of data: INSEE, 2009  
© EuroGeographics Association for administrative boundaries

#### TYPE OF AGE STRUCTURE IN 2000

- A) Excedent of young population
  - Type A.1
- B) Excedent of active population
  - Type B.1
  - Type B.2
- C) Excedent of old population
  - Type C.1
  - Type C.2
- D) Medium profile
  - Type D



#### TOTAL POPULATION IN 2000 (inh.)



**Figure 14:** Population and age structure in Brittany (France) – **without semiologic problems**

## 3.2 Improving the efficiency of the map

Other problem which appears regularly is the degree of complexity of the map. The aim of the maps is to be synthetic. When representing too much information, the eye cannot distinguish the different elements of the map. This kind of figure can be solved by thinking to the design of the map: where is the best location for legend? How using with the most efficiency the place available?

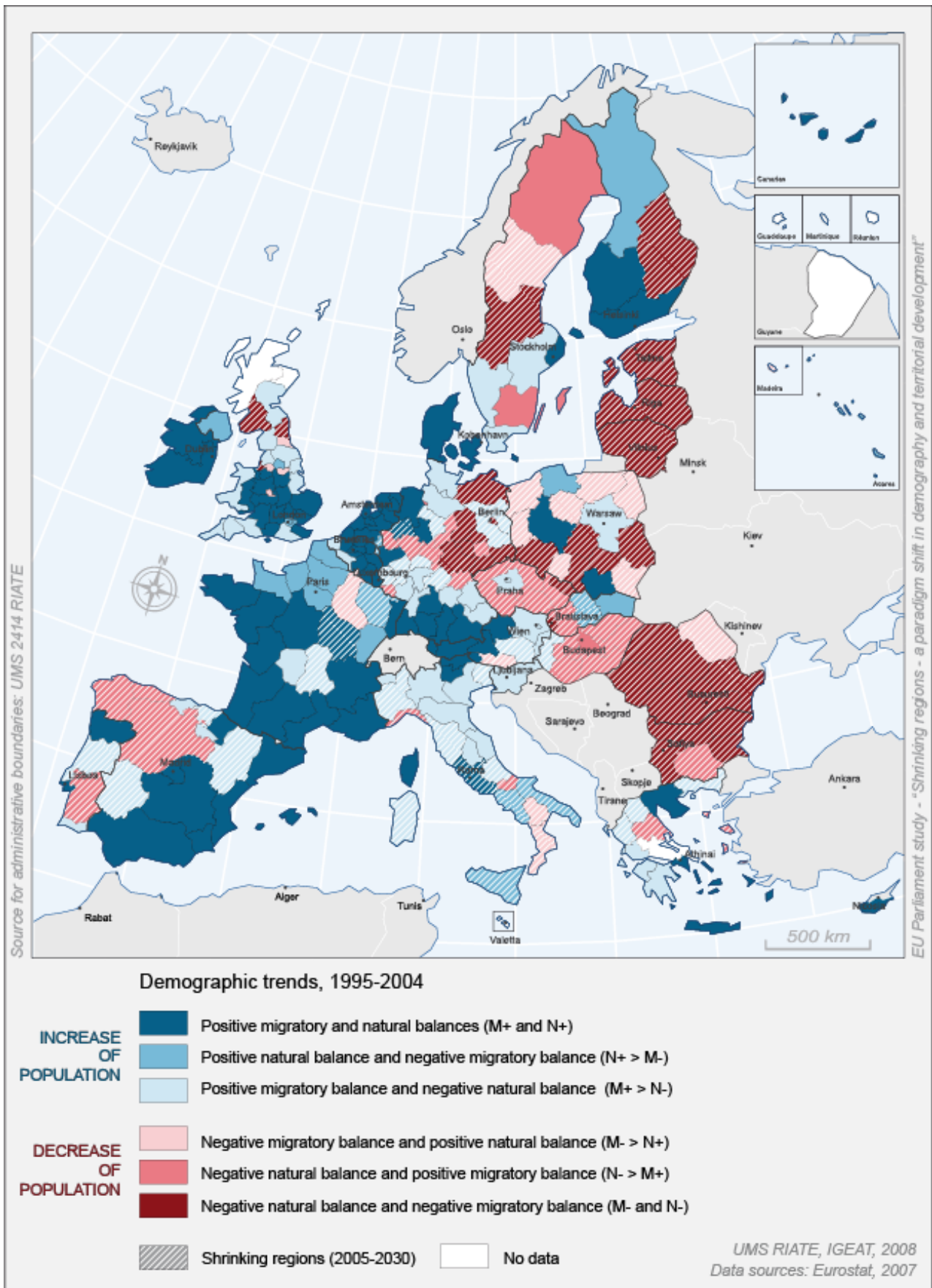
**The figures 15 and 16** show the same information, e.g. a typology of population development by components during the period 1995-2004 in EU27; this data is crossed with expected population evolution in 2030.

**Figure 15** proposes solution which is correct in term of graphic semiology: ordinal data are shown by variation of colour (green/red) and shrinking/non shrinking regions (qualitative data) are represented by the opposition of hachure and no hachure. However, the combination of these two visual variables makes the map hard to interpret and the message become not so clear!

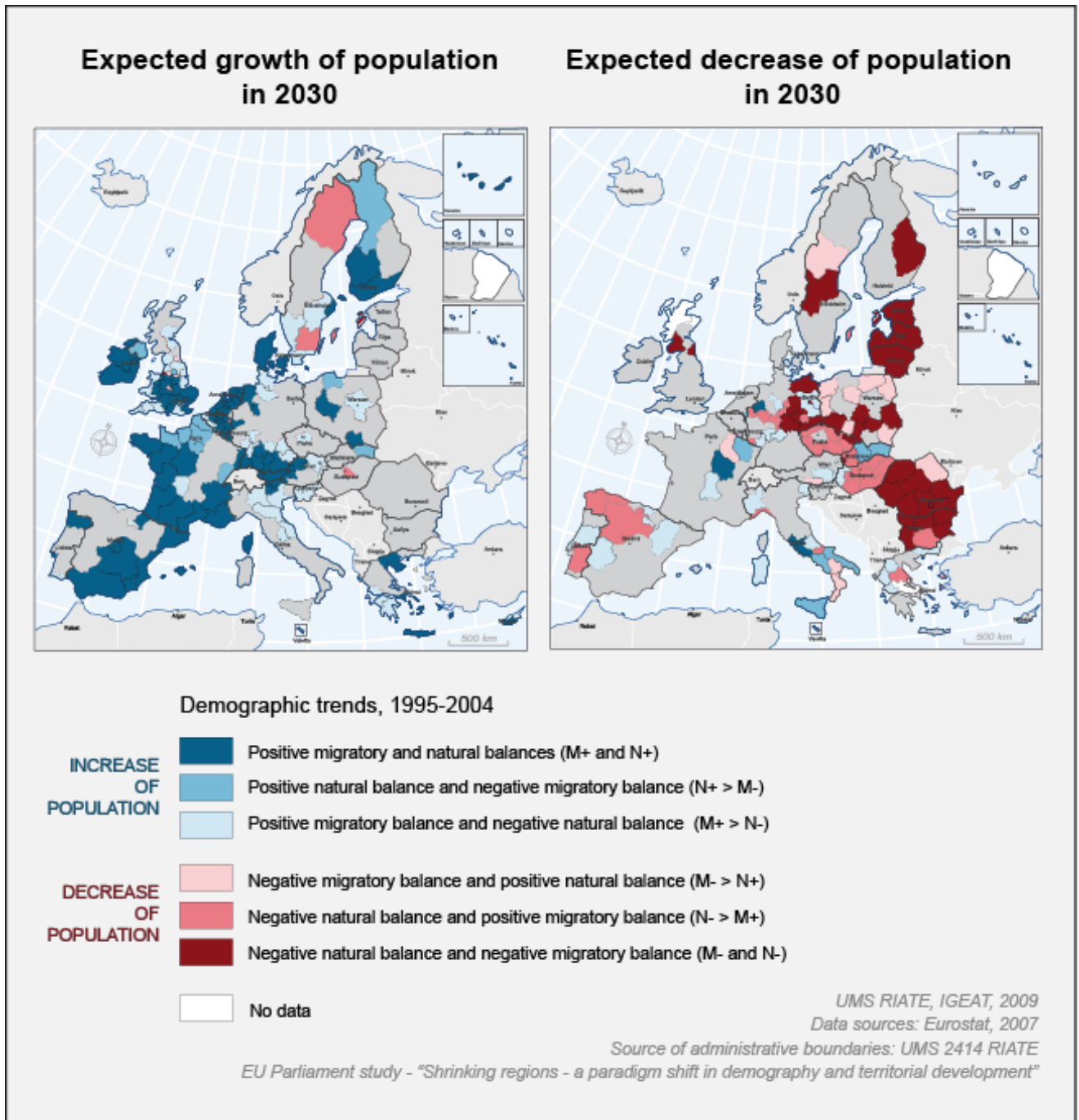
When there is too much information it becomes difficult to be able to synthesise the message of the map. That is why in some cases it is more efficient to split information in two maps instead of concentrating all the elements in a single one. This has been done on figure 16, where the map located on left of the document shows the regions described by an expected growth of population; and the map on the right shows the regions where a demographic decrease is planned. This template allows immediately to observe that during the period 1995-2005 most of the 'shrinking regions' have witnessed a downturn linked to both natural change and a negative migratory balance.

### ***There is never an optimal solution***

Whatever the examples proposed and demonstrated, it is important to keep in mind that there is never a single solution to show information on maps. In fact, each person has his own perception when interpreting graphic documents or pictures. **Map is always a compromise.** But during the creation of the map, is fundamental to try to make the map as understandable as possible. In concrete terms, it is not an obvious task and it is kindly recommended to make different attempts and share the results with other colleagues before saying "OK, my map is ready for the report"!



**Figure 15: Typology of regional growth patterns – Possibility 1**



**Figure 15: Typology of regional growth patterns – Possibility 2**



## ANNEXES

These annexes allow you to choose some efficient graphic variables to communicate differences in size, order or quality.





























































































### ANNEXE 1 - Relation of graphical variables to perceptual characteristics

Graphical variable	Type of data			
	nominal	ordinal	Interval/ratio	quantity
Size		x	x	x
Grey or colour value		x	x	
Grain/texture		x	x	
Colour hue	x			
Orientation	x			
Shape	x			

### ANNEXE 2 - Numbers of categories that can be perceived at a glance

Graphical variable	Point	Line	Area
Size	4	4	5
Grey or colour value	3	4	5
Grain/texture	2	4	5
Colour hue	7	7	8
Orientation	4	2	4
Shape	3	3	3

# ANNEXE 3: Differences in value or lightness

COLOUR INTENSITY			
Blue	Green	Red	Brown
<b>4 classes</b>			
 rgb(0,147,193)	 rgb(31,115,42)	 rgb(235,107,57)	 rgb(126,70,53)
 rgb(118,188,218)	 rgb(100,175,64)	 rgb(246,170,65)	 rgb(195,118,70)
 rgb(208,232,244)	 rgb(191,217,159)	 rgb(255,227,125)	 rgb(229,170,81)
 rgb(235,246,252)	 rgb(230,239,207)	 rgb(255,249,200)	 rgb(255,237,170)
<b>5 classes</b>			
 rgb(0,147,193)	 rgb(18,94,39)	 rgb(229,53,64)	 rgb(126,70,53)
 rgb(118,188,218)	 rgb(60,145,60)	 rgb(235,107,57)	 rgb(195,118,70)
 rgb(167,212,233)	 rgb(129,188,96)	 rgb(246,170,65)	 rgb(229,170,81)
 rgb(208,232,244)	 rgb(191,217,159)	 rgb(255,227,125)	 rgb(255,221,139)
 rgb(235,246,252)	 rgb(230,239,207)	 rgb(255,249,200)	 rgb(255,237,170)
<b>6 classes</b>			
 rgb(0,124,176)	 rgb(18,94,39)	 rgb(229,53,64)	 rgb(97,68,55)
 rgb(0,147,193)	 rgb(60,145,60)	 rgb(235,107,57)	 rgb(126,70,53)
 rgb(118,188,218)	 rgb(107,178,76)	 rgb(246,170,65)	 rgb(195,118,70)
 rgb(167,212,233)	 rgb(151,197,110)	 rgb(255,227,125)	 rgb(229,170,81)
 rgb(208,232,244)	 rgb(200,218,140)	 rgb(255,249,200)	 rgb(255,221,139)
 rgb(235,246,252)	 rgb(239,241,199)	 rgb(255,253,238)	 rgb(255,237,170)
<b>8 classes</b>			
 rgb(0,98,140)	 rgb(11,82,34)	 rgb(173,26,34)	 rgb(97,68,55)
 rgb(0,124,176)	 rgb(31,115,42)	 rgb(207,54,65)	 rgb(126,70,53)
 rgb(0,147,193)	 rgb(62,146,44)	 rgb(229,53,64)	 rgb(165,94,57)
 rgb(68,170,207)	 rgb(100,175,64)	 rgb(235,107,57)	 rgb(195,118,70)
 rgb(118,188,218)	 rgb(145,191,91)	 rgb(246,170,65)	 rgb(219,145,73)
 rgb(167,212,233)	 rgb(180,209,121)	 rgb(255,227,125)	 rgb(229,170,81)
 rgb(208,232,244)	 rgb(200,218,140)	 rgb(255,249,200)	 rgb(255,221,139)
 rgb(235,246,252)	 rgb(239,241,199)	 rgb(255,253,238)	 rgb(255,237,170)

## GREY VALUE

### 4 classes



### 5 classes



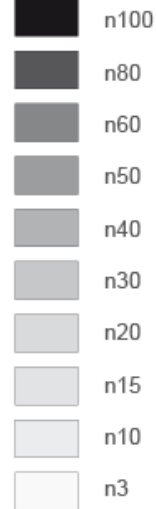
### 6 classes



### 8 classes



### 10 classes



## OPPOSITE COLOURS

 rgb(0,98,140)

 rgb(0,147,193)

 rgb(118,188,218)

 rgb(235,246,252)

 rgb(252,208,211)

 rgb(234,122,133)

 rgb(196,55,79)

 rgb(142,3,17)

 rgb(90,93,122)

 rgb(114,118,159)

 rgb(147,153,199)

 rgb(196,200,226)

 rgb(249,230,239)

 rgb(240,184,210)

 rgb(236,141,181)

 rgb(226,2,128)

 rgb(32,115,43)

 rgb(66,145,44)

 rgb(145,191,92)

 rgb(222,229,157)

 rgb(247,229,196)

 rgb(250,210,147)

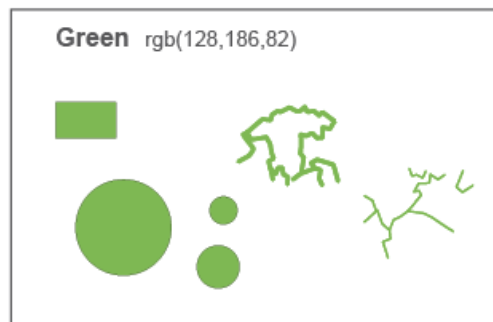
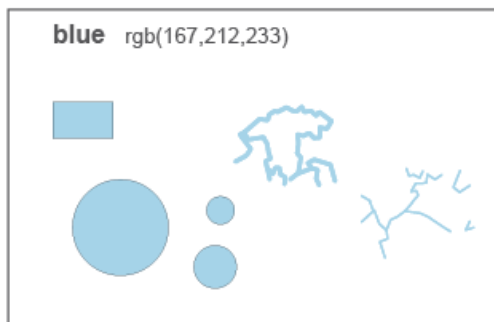
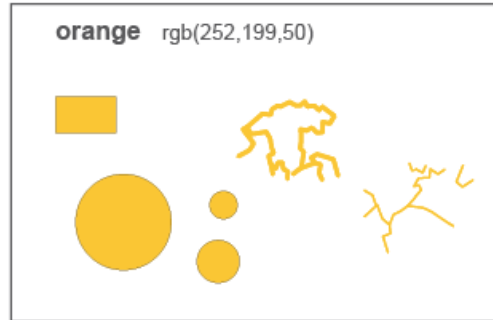
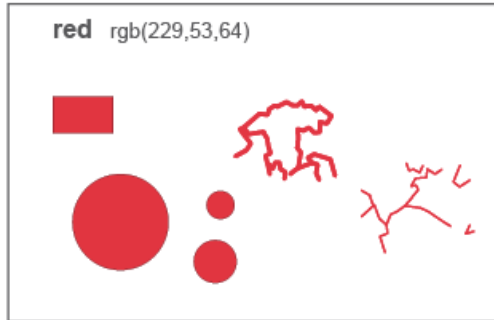
 rgb(244,171,42)

 rgb(175,110,22)

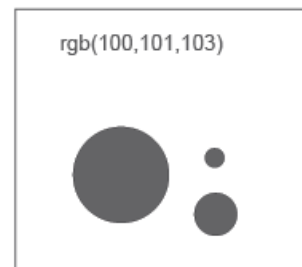
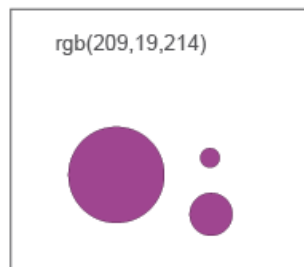
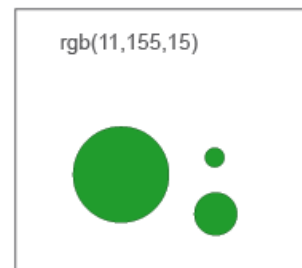
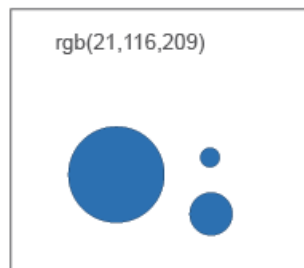
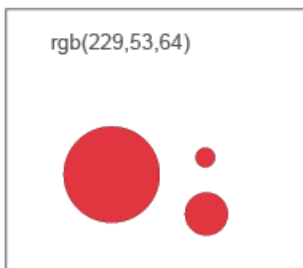
# ANNEXE 4: Colours for differences typology or qualitative value

## QUALITATIVE VALUES

(circles and discontinuities)



(circles)



## References

### • *Litterature*

Béguin M., Pumain D., 2003, *La représentation des données géographiques – statistique et cartographie*, Armand Colin.

Bertin J., 1967, *Sémiologie graphique*, Gauthiers-Villars.

Cambrezy L., de Maximy R. (Ed.), 1995, *La cartographie en débat, représenter ou convaincre*, Editions Kathala et Orstom, Paris

Harris R. L., 1996, *Information graphics, a comprehensive illustrated reference, visual tools for analysing, managing and communicating*, Management Graphics ed., USA

Harley, J. B., 1988, *Maps, knowledge and power*. In COSGROVE, D. (Ed.) *The Iconography of Landscape*. Cambridge, MA, Cambridge University Press.

Kraak M.-J., Ormeling F., 2003, *Cartography, Visualization of Geospatial Data*, 2<sup>nd</sup> edition, Pearson Education, Prentice Hall.

Kraak, M.-J., 1998, *Exploratory cartography, map as tools for discovery*, *ITC Journal* (1), pp.46-54

MacEachren A.M., 1994, *Some truth with maps: a primer on design and symbolization*, Association of American Geographers, Washington DC.

Monmonnier M., 1996, *How to lie with maps*, University of Chicago Press.

Robinson A.H., Morrison J.L., Muehrcke P.C., 1995, *Elements of cartography*, New York, J.Willey & Sons.

Wilkinson L., 1999, *The grammar of graphics*. New York, Springer.

Wood, D., 1992, *The Power of Maps*. New York, The Guildford Press.

Wood C. H., Keller C. P., 1996, *Cartographic design: theoretical and practical perspectives*, Wiley, USA

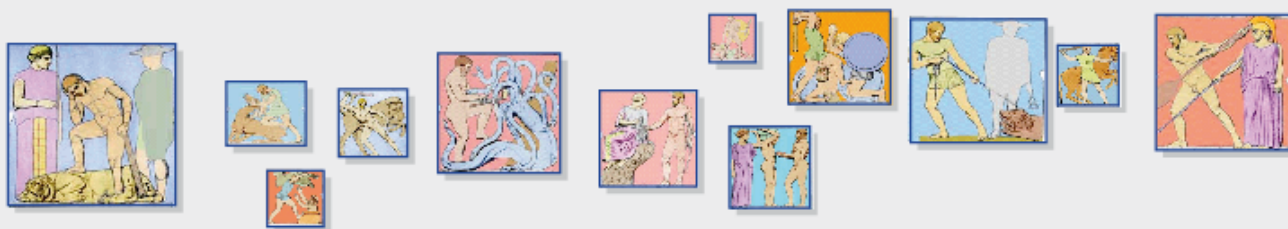
Zanin C., Trémélo M-L, 2003, *Savoir faire une carte: Aide à la conception et à la réalisation d'une carte thématique univariée*, Belin.

### • *Websites*

**Colorbrewer 2.0** is an online tool designed to help people select good color schemes for maps and other graphics: <http://colorbrewer2.org/>

**Philcarto** is a free tool for cartography, available on the net: <http://philcarto.free.fr/>

**Quantum GIS** is an Open Source Geographic Information System. It runs on Linux, Unix, Mac OSX, and Windows and supports numerous vector, raster, and database formats and functionalities: <http://www.qgis.org/>



## MAPPING GUIDE

### *CARTOGRAPHY IN ESPON 2013*

#### CONTENT

- Enhancing information. This part explain how symbolize ESPON 2013 data with the good rules of graphic semiology.
- Maps are tool for communication. This part insists on the fact that a map has necessarily to deliver a clear message.

**ESPON 2013 DATABASE**  
**JANUARY 2011**



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

25 pages

## LIST OF AUTHORS

Christine Zanin, University Paris 7, UMS 2414 RIATE

Nicolas Lambert, UMS 2414 RIATE

Ronan Ysebaert, UMS 2414 RIATE

### **Contact**

[christine.zanin@univ-paris-diderot.fr](mailto:christine.zanin@univ-paris-diderot.fr)

[nicolas.lambert@ums-riate.fr](mailto:nicolas.lambert@ums-riate.fr)

[ronan.ysebaert@ums-riate.fr](mailto:ronan.ysebaert@ums-riate.fr)

tel. + 33 1 57 27 65 32

# TABLE OF CONTENT

<b>Introduction</b> .....	3
<b>1 Enhancing information</b> .....	4
1.1 Differentiation of data type .....	4
1.1.1 Qualitative data .....	4
1.1.2 Quantitative data with absolute values .....	6
1.1.3 Quantitative data with interval or ratio values .....	7
1.1.4 Ordinal or ranked data .....	8
1.2 When using two variations of colour?.....	9
1.3 Choice of data ranges.....	9
1.3.1 Natural Break.....	10
1.3.2 Equal Count or quantile .....	10
1.3.3 Equal Ranges .....	10
1.3.4 Jenks method.....	11
1.3.5 Standard deviation.....	11
1.3.6 Geometric progression.....	12
<b>2 Maps are tool for communication</b> .....	14
2.1 Bad choices in term of representation of the data .....	15
2.2 Improving the efficiency of the map .....	18
<b>ANNEXES</b> .....	21
ANNEXE 1 - Relation of graphical variables to perceptual characteristics .....	21
ANNEXE 2 - Numbers of categories that can be perceived at a glance .....	21
ANNEXE 3: Differences in value or lightness .....	22
References .....	25



# Introduction

Maps are a great way of displaying statistical data. It allows summarizing a complex and important information into clear and compact presentation. They can bring a great help in spotting patterns within data.

Maps are accessible for many reasons. People understand maps (at least, think they do). People like maps because they attract attention and brighten up presentation. Nevertheless, and in a scientific versus, the interest of the representation of geographical information on maps can be summarized in three main points<sup>1</sup>.

**The localisation** is the most elementary subject related to geographic information. It allows answering to question "Where can we find this phenomenon?" The precision of the localisation depends on the quality of this kind of information such as statistical databases, statistical yearbook and so on. Locate a geographical object has generally a sense only if it is possible to compare it to other one "Why this object is located here and not there?". Answers can be read off directly from the map without any other help.

**The comparison:** Geographical objects analysis makes a concrete sense when it is possible to compare them. "What is the situation of this region as compare to the other one?"; "Can we observe geographical pattern, such as discontinuities, concentration?" Maps are useful tools for interpreting and pointing out specific geographical patterns, which are impossible to catch with an only statistical analysis.

**Planning:** Since the relations between European territories are very intensive, territorial planning on a special location must interfere with other territories and have to.

Despite many interests to use maps within ESPON, these kinds of documents have also their limits. Maps always generalise and simplify information. Mapping is more than just rendering; it also getting to know the phenomenon which is to be mapped. That's why mapping is not an easy action. Deliver the right message must remain the first objective of map design and mapping allows you to orchestrate the elements of the map to best convey its message to its audience. Thus, the design of maps is mainly concerned with making choices: the choice of mapping method (proportional symbol or choropleth map, isoline or grid map or even a cartogram), the choice of the aggregation level on which information as to be depicted, the choice on the level of statistic areas and the type of data (absolute or relative representation), the choice of graphic variables (such as differences in size, value, grain, colour, direction and shape) to be used. These choices are fundamental's one, they influence people's conception and visualisation of space.

This technical report is not a formal cartography book but allows everyone to understand easily how to produce an effective and operational map in the ESPON 2013 program. The report is organized in two parts: (i) Enhancing information (mapping methods and graphic semiology); (ii) Maps and communication (map is to deliver a simple and clear message).

---

<sup>1</sup> Béguin M., Pumain D., 2003, *La représentation des données géographiques – statistique et cartographie*, Armand Colin, 192p.

# 1 Enhancing information

## 1.1 Differentiation of data type

Many possibilities exist to show data on map. Choosing relevant representation is not an obvious task and has to be considered seriously. Indeed, choosing the wrong type of map can completely misrepresent the data. It is important to keep in mind that **the choice in cartography is always dependant on the type of data**. It is possible to identify four main types of data:

1. Qualitative data
2. Quantitative data with absolute values
3. Quantitative data with ratios values
2. Ordinal (or ranked) data

For each type of data it is possible to relate it to a **geographical reference: points, lines or areas**.

There are many possibilities to show correctly data on maps. The aim of this paper is not to present all types of correct visualisation, but an extract of the most usual and efficient ones.

### 1.1.1 Qualitative data

A data is qualitative when its value is a nominal one with qualitative differences: components do not allow establishing range relations between them.

For example, considering the different geographical references:

Points: location universities by type (university, polytechnics...) – **Figure 1**

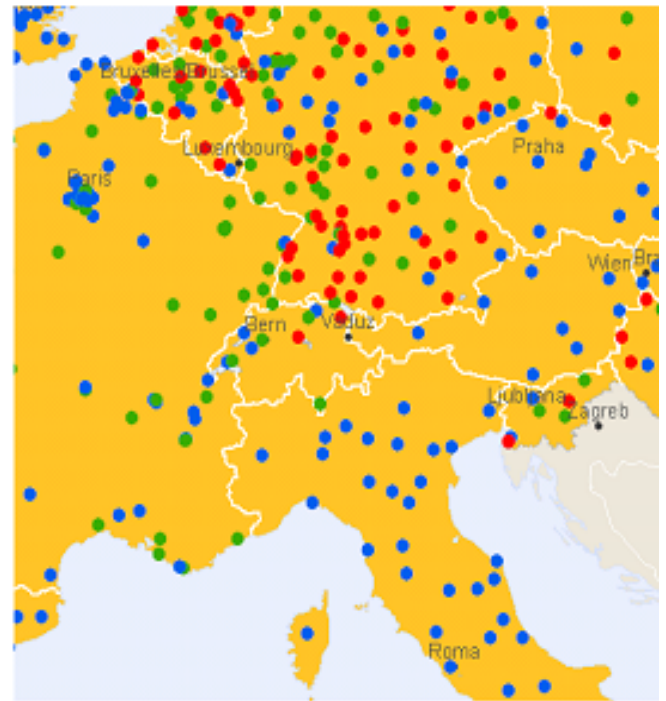
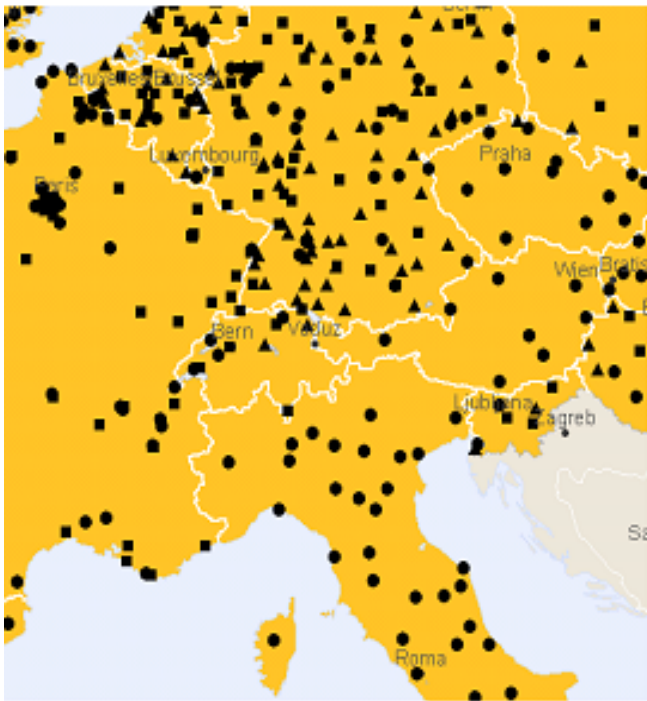
Lines: communication network without hierarchy (ferry connections, main roads) – **Figure 2**

Areas: results from typology (rural area, urban area...) – **Figure 3**

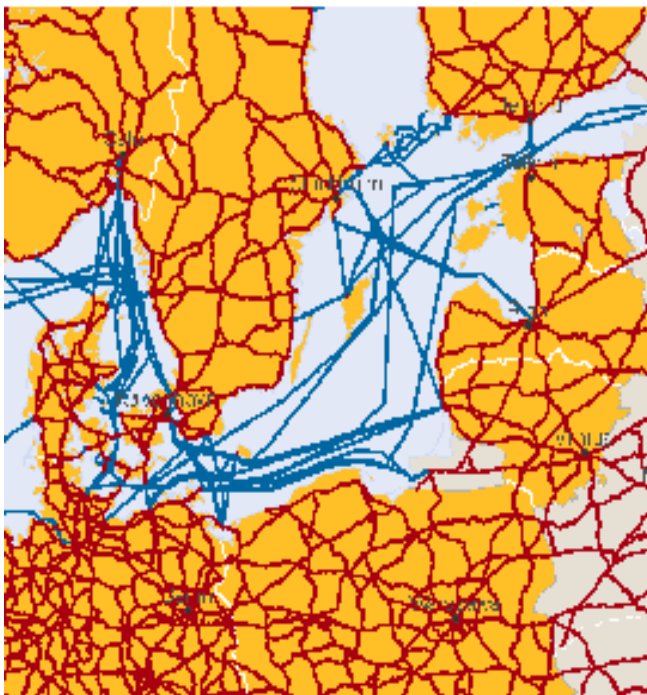
Qualitative data have to be shown such a manner that do not suggest rank either quantity. Two possibilities: use **geometric symbols** or **differential colour** in order to **differentiate** the different elements of the map.

With points (figure 1) the most efficient is to show information by colour or geometric symbols. It is important to use a limited quantity of symbols or colours to make the map understandable.

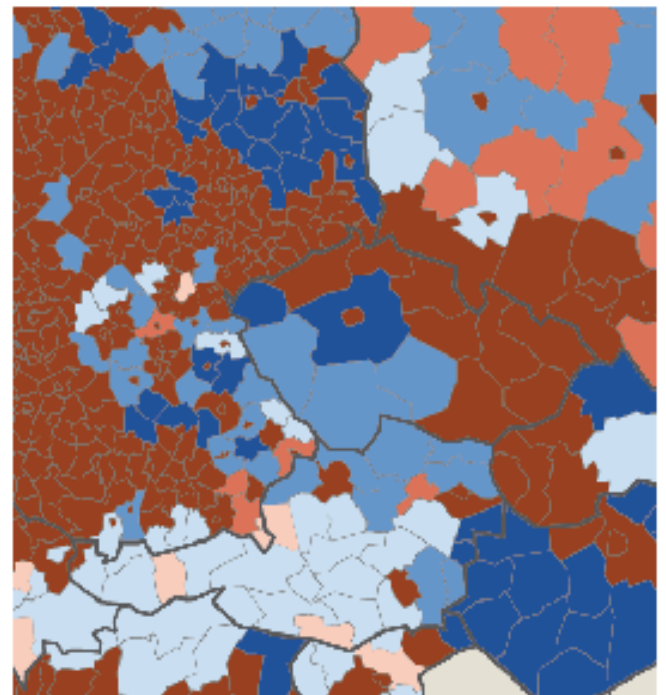
For lines or areas, differential colours should be used (figure 2 and 3).



**Figure 1 - Universities by types - Two possibilities**  
**Good map = points + symbols or points + colours**



**Figure 2 - Mains roads and ferry connections**  
**Good map = line + colours**



**Figure 3 - Results from Urban-Rural typology**  
**Good map = areas + colours**

## 1.1.2 Quantitative data with absolute values

Quantitative data with absolute values means concrete **quantity**; the sum of the different values can be calculated and has a real sense. For example, population, GDP, CO2 emissions are absolute quantitative data if we consider the number of inhabitants, number of euros or tons of gas emissions.

For example, considering the different geographical references:

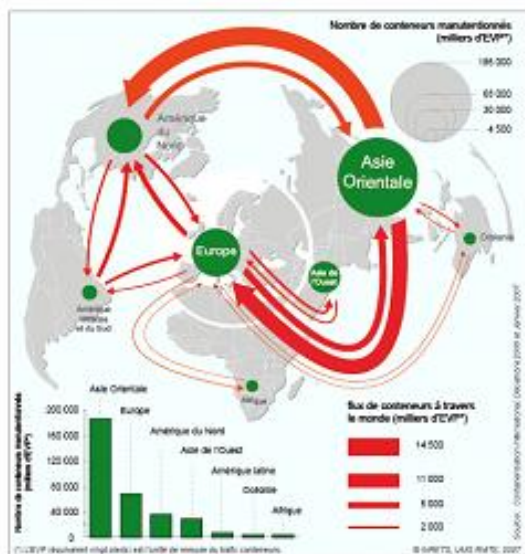
Points: Cities of Europe (number of inhabitants)

Lines: Containers flows across the world (millions tons) – **Figure 4**

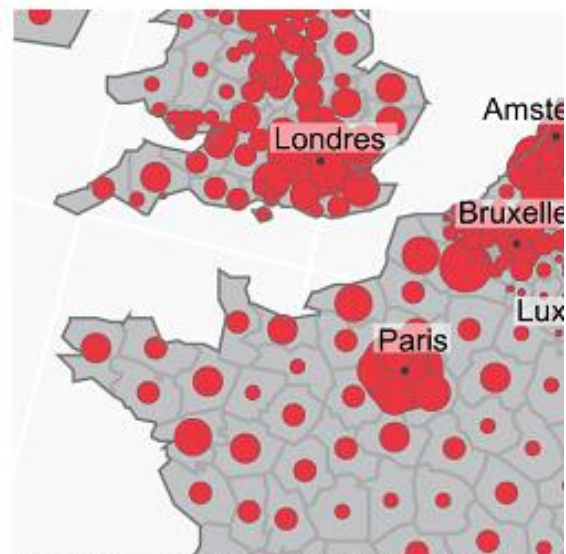
Areas: Population of NUTS 3 – **Figure 5**

Whatever the type of geographical objects (points, line, areas), the cartography of quantitative data with absolute values has to **respect the quantity** and differences of proportionality. For points or areas objects, the most common representation is to use maps with area **proportional circles**. The circled area is proportional to the size of the data value.

The map showing data in line format (**figure 4**) has to use lines of different width. The width of the line is proportional to the data value.



**Figure 4 - Containers flows across the World**  
Good map = line + variation of line size



**Figure 5 - Population of NUTS 3 in Europe**  
Good map = dot + proportional variation of size

### 1.1.3 Quantitative data with interval or ratio values

The ratio values are calculated and expressed a series of ratios or proportional values, such as percentage, per km, per inhabitant. This kind of data is the most common.

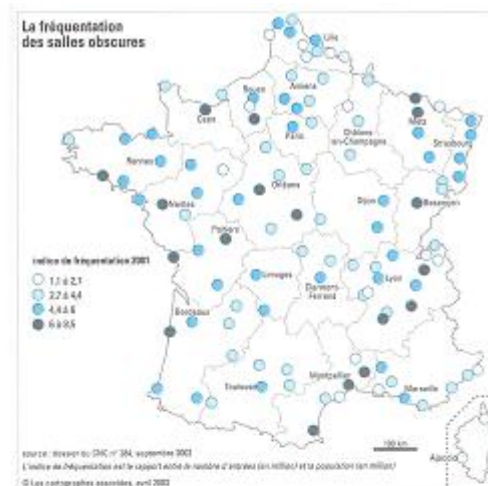
For example, considering the different geographical references:

Points: Cities of France (cinema attendance index) **Figure 6**

Lines: GDP per inhabitants discontinuities (relative difference between two territories) – **Figure 7**

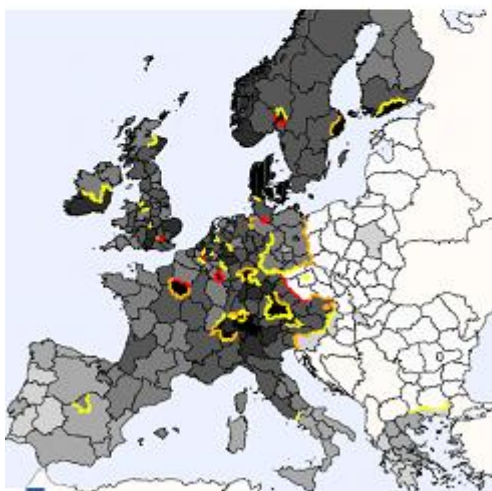
Areas: Abstention, European elections 2009, in Ile-de-France municipalities – **Figure 8**

For ratios values, the most relevant representation is a choropleth map where density is linked to the class of the data value for each area. The efficiency of the map depends on the range between the least dense (lightest) area and the densest (darkest) area. When correctly applied, percentage or densities that are twice as high are represented by a grey value that is twice as dark.



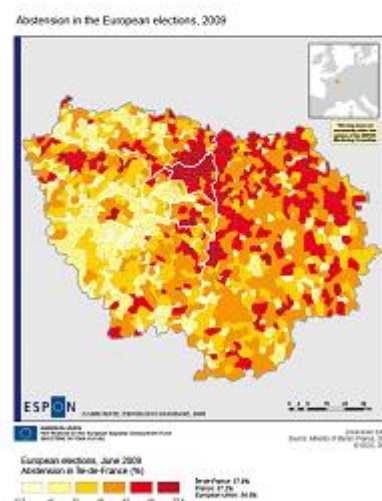
**Figure 6: Cinema attendance Index in French main cities**

*Good map = dots + variation of colours*



**Figure 7: GDP per inhabitants discontinuities**

*Good map = lines + variation of colours*



**Figure 8: Abstention European votes in Île-de-France**

*Good map = area + variation of colours*

### 1.1.4 Ordinal or ranked data

Ordinal data are categorical data where there is a logical ordering to the categories. A good example is the Likert scale that you see on many surveys: 1=strongly disagree; 2=Disagree; 3=Neutral; 4=Agree; 5=strongly agree. Another example could be found with modalities like first, second, third etc., or small, medium and high.

For example, considering the different geographical references:

Points: Typology of Functional Urban Areas – MEGA, national FUA, regional FUE

#### **Figure 9**

Lines: Road hierarchy – **Figure 10**

Areas: Degree of policentricity – **Figure 11**

The representation of these data is based on the expression of natural modalities order. Considering the different geographical references (point, line or area) you can only use 2 graphics variables: grey value or the intensity of a colour. They allow denoting differences in intensity of a phenomenon and expressing order between geographical areas, points or lines. Because differences in grey value or in intensity of colour are used, a hierarchy or order between ordinal modalities can be perceived.



**Figure 9:** Typology of Functional Urban Areas  
Good map: points + variation of colour (or size)



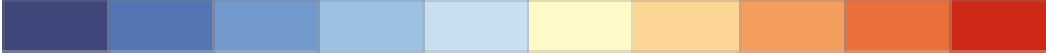
**Figure 10:** Road hierarchy in Europe  
Good map: lines + variation of colour



**Figure 11:** Degree of policentricity in Europe  
Good map: areas + variation of colour

## 1.2 When using two variations of colour?

It is sometimes necessary to show a phenomenon by a variation of two colours fundamentally different:



This kind of representation is very useful since it allows making more differentiation between the classes of the map. However, it is possible to use these oppositions of colours only if the **break has an objective sense** in the dataset, for instance:

- Opposition between negative and positive values (decrease and increase of population between two periods)
- Values above/under the average value or median value of the dataset (level of accessibility above or under the EU27 average)
- Values above/under a value which have a concrete reality (unemployment rate under/above the threshold of 10 %).

Opposition of variation of two colours should be used only for quantitative data with ratio values and ranked data.

To ensure the **harmonisation** of all maps produced by ESPON projects, it is important that also the use of colours is being guided in the case of opposite colours. In general, it is advised not to combine red and green in one map in order to serve the colour-blind people. Other general rules do not exist. The choice of opposite colors is very subjective and cultural. However, it is quite confusing if two different ESPON maps are published where red has a positive meaning in one of the maps and a negative meaning in the other map. Therefore, in the case of ESPON maps with opposite colours, it is decided to have the following principle as guideline:

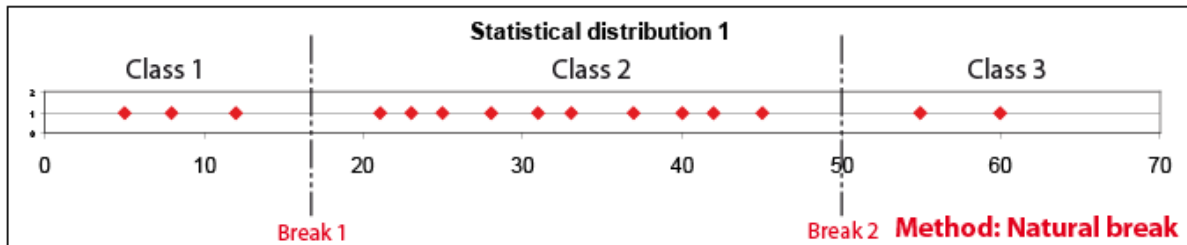
When combining red (warm colours) and blue (cold colours), **red is 'not good'/'negative'** and **blue is 'good'/'positive'**

## 1.3 Choice of data ranges

Nevertheless, this kind of representation introduces always a **loss of information** since it transforms a complex statistical distribution into a limited number of classes. Information becomes more generalised and simplified. The accuracy of original values is lost, but **this operation is needed in order to present a synthetic overview of the dataset**. Indeed, a good class division will focus on what is the main content of the dataset, and minimise the loss of accuracy by generalisation. Further below you will find five different classes dividing methods ranging data values. Of course it is also possible to combine different methods, in particular when there are an important number of records. This step before mapping is needed for quantitative values with ratios only.

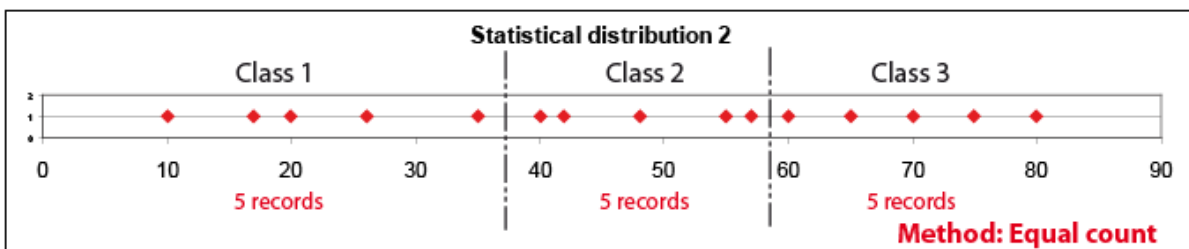
### 1.3.1 Natural Break

This method sets the breakpoint to “natural points” in the dataset. The strength of this method is that it increases the information content. **This method is suited when important breaks** describe the dataset.



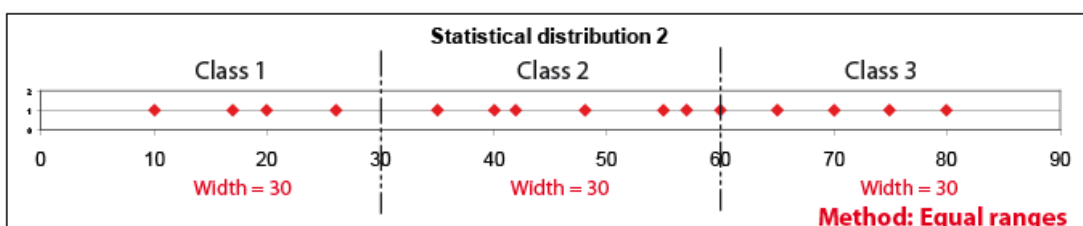
### 1.3.2 Equal Count or quantile

**Equal range contains approximately the same number of records.** With 5 classes, each contains 20 % of the total number of the data values. **This method is suited for comparing one dataset with datasets from other themes.** If the data deviate from a linear distribution, the absolute class width will show large variations. Equal count methodology does not take into account exceptional values in the distribution.



### 1.3.3 Equal Ranges

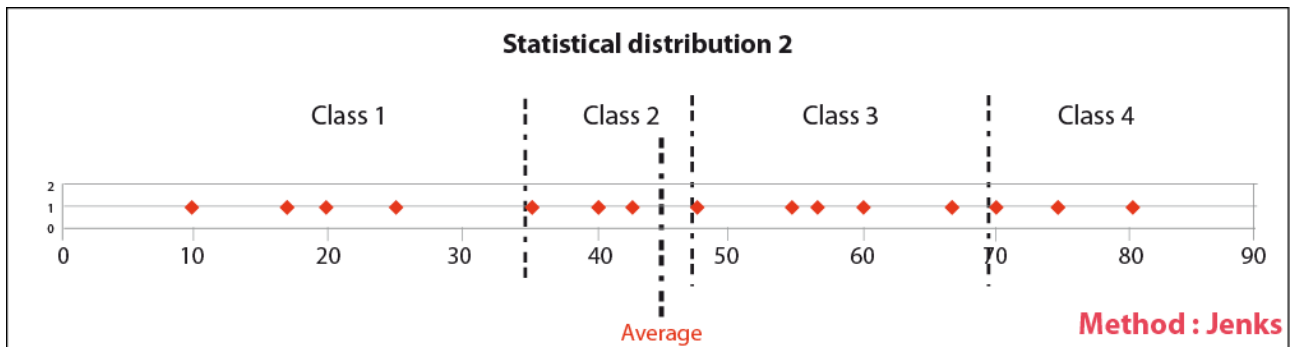
**The difference between the top and bottom values in each range is the same.** This means that we can use values like 0-20; 20-40 etc. or calculate the width of the dataset, and divide by the number of classes wanted. In this case the lowest class will start with the lowest value; the width between the classes will be the same, and the top of the highest value in the dataset. **This method is suited for datasets with a smooth linear distribution.** If the method is used on dataset that are not linear distributed, you will have some classes with many values and others with few or no values.





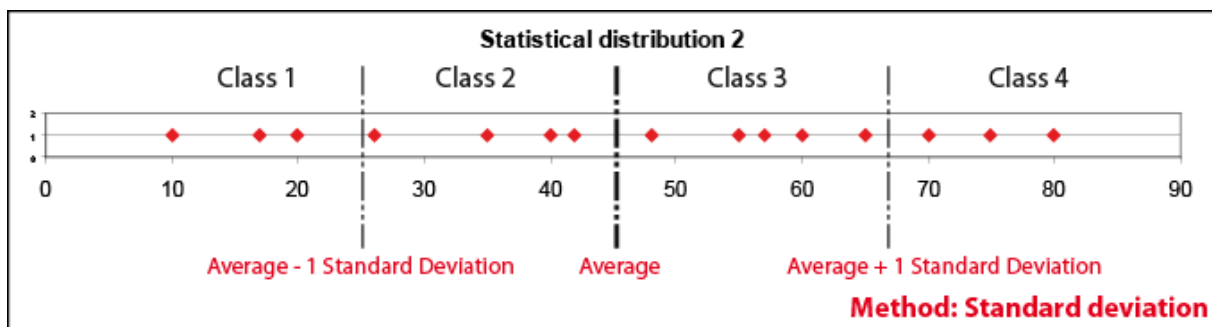
### 1.3.4 Jenks method

The Jenks' optimization method is also known as the **goodness of variance fit (GVF)**. It is a data classification to determine the best arrangement of values into classes. It is used to minimize the squared deviations of the class means while maximizing each class's deviation from the means of other groups. The method first specifies an arbitrary grouping of the numeric data, and then observations are moved from one class to another ensuring maximum homogeneity within groups and maximum heterogeneity between groups.



### 1.3.5 Standard deviation

**The class borders are calculated from the mean value and the standard deviation.** Standard deviation is a way to describe statistical dispersion. The width of the class is equal to the standard dispersion (or an half depending on the number of classes expected). **This method is suited for normal distributed datasets only.**



### 1.3.6 Geometric progression

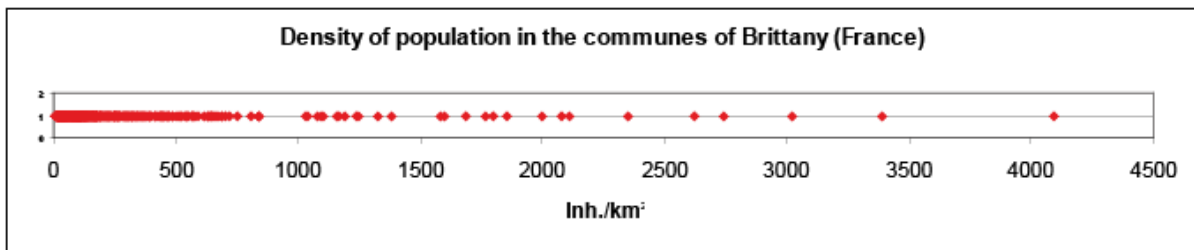
**The widths of the class follow a geometric progression.** To calculate the width of the different class, it is necessary to estimate the geometric ratio, such as:

$$\log R = (\log_{10} \text{Max} - \log_{10} \text{Min}) / \text{number of classes wanted}$$

$$R = 10^{\log r}$$

Width of the Classes = (min, min x R); (min x R; min x R x R) and so on.

**This method is suited for uneven distribution** and particularly distribution described by a lot of low values and few high values, such as density of population distribution.



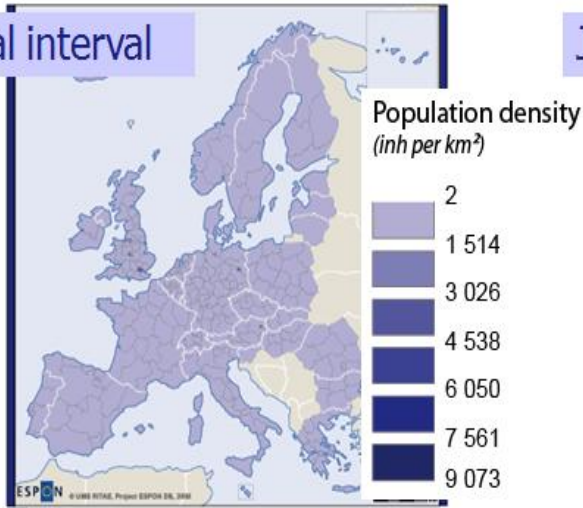
From the example of Brittany, the data ranges, following the geometric progression, should be in 6 classes:

Class	Class boundaries	Number of communes
1	[9;25[	128
2	[25;70[	626
3	[70;190[	343
4	[190;525[	117
5	[525;1470[	39
6	[1470;4100[	15

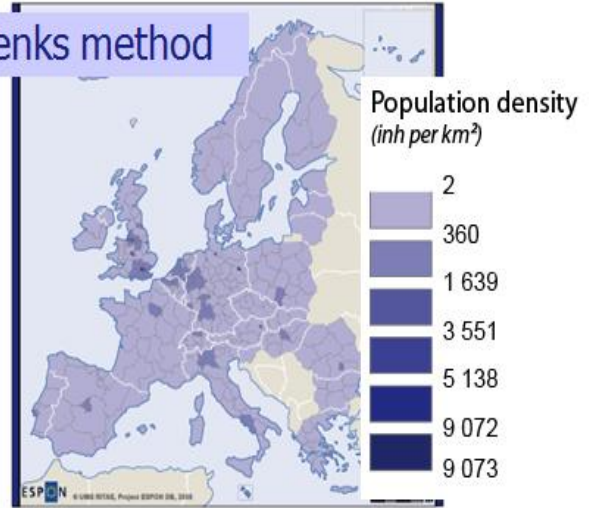
**Whatever the method chosen for ranging the distribution, it is important to use smooth values for the break, in order to understand and memorize easier the sense of the map, e.g. use 30 instead of 29,77; 1500 instead of 1508 etc.**

**Figure 12** shows the importance of the choice of data range on the visualisation of phenomena.

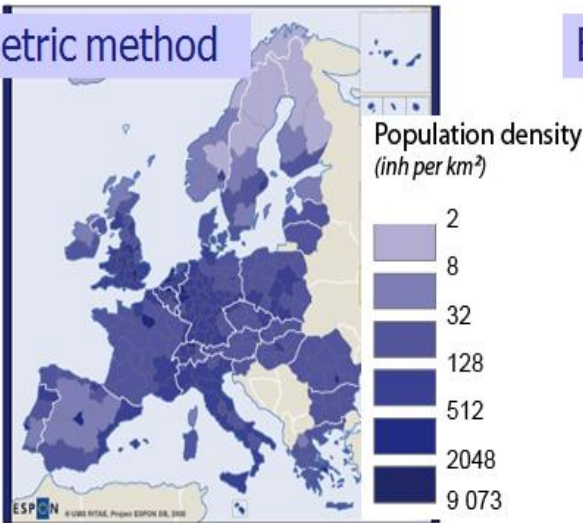
Equal interval



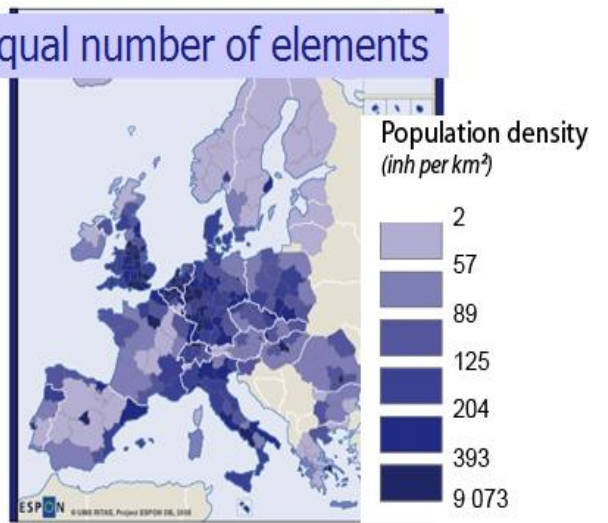
Jenks method



Geometric method



Equal number of elements



**Figure 12:** Result and efficiency are dependent upon the data classification method

## 2 Maps are tool for communication

As we explain in the introduction of this technical report: "Maps are perhaps as fundamental to society as language and the written word. They are the preeminent means of recording and communicating information about the location and spatial characteristics of the natural world and of society and culture<sup>2</sup>".

Maps are produced all over the world and used by people as different as scientists, researchers, scholars, governments or businesses. These maps are most of the time statistical ones connected with the environment, the economy, the politics, the society etc.

**The biggest strength of these maps is to allow an effective and relevant communication of the information.** However, cartography is a special type of visual communication that does require some preliminary learning: a special purpose language for describing spatial relationships. "The analogy with language also helps explain why training in principles of effective cartography is so important--it allows us to communicate more effectively. Without knowledge of some of these basic principles, the beginning cartographer is likely to be misunderstood or cause confusion<sup>2</sup>".

Of course, cartographers must pay special attention to coordinate systems, map projections, and issues of scale and direction but that's not the first issue of map as a tool for communication. Maps are symbolic abstractions and representations. **The first question when mapping is related to know how to simplify, generalize, represent and symbolize the relationships being represented with graphics symbols.** In other words, what is a good map?

If a design is always more effective than a long speech, the measure of a good map is how well it conveys the right information to its readers and how well it communicates with its audience. This raises a series of questions that must be addresses at the start of a map conception: What is the motive, intent, or goal of the map? Who will read the map? Where will the map be used? What data is available for the composition of the map?

Beyond aesthetic characteristics, the communication also passes by a complete and effective layout: some elements must appear within the base map and the thematic representation, a complete legend, explicit title and source, a precise date of data or even a scale.

---

<sup>2</sup> Kenneth E. Foote and Shannon Crum, *The Geographer's Craft Project*, Department of Geography, The University of Colorado at Boulder

From data to map, 7 fundamental goals need to be identified to realize a good map:

1. Identify the goal of the map;
2. Identify the audience of the map and where it will be used;
3. Identify the information to be communicated;
4. Identify the geographical reference (point, line or area?);
3. Choose the base map (map projection and scale);
4. Choose the visual variable (symbolic graphic language);
5. Choose layout and identify all the elements to be added.

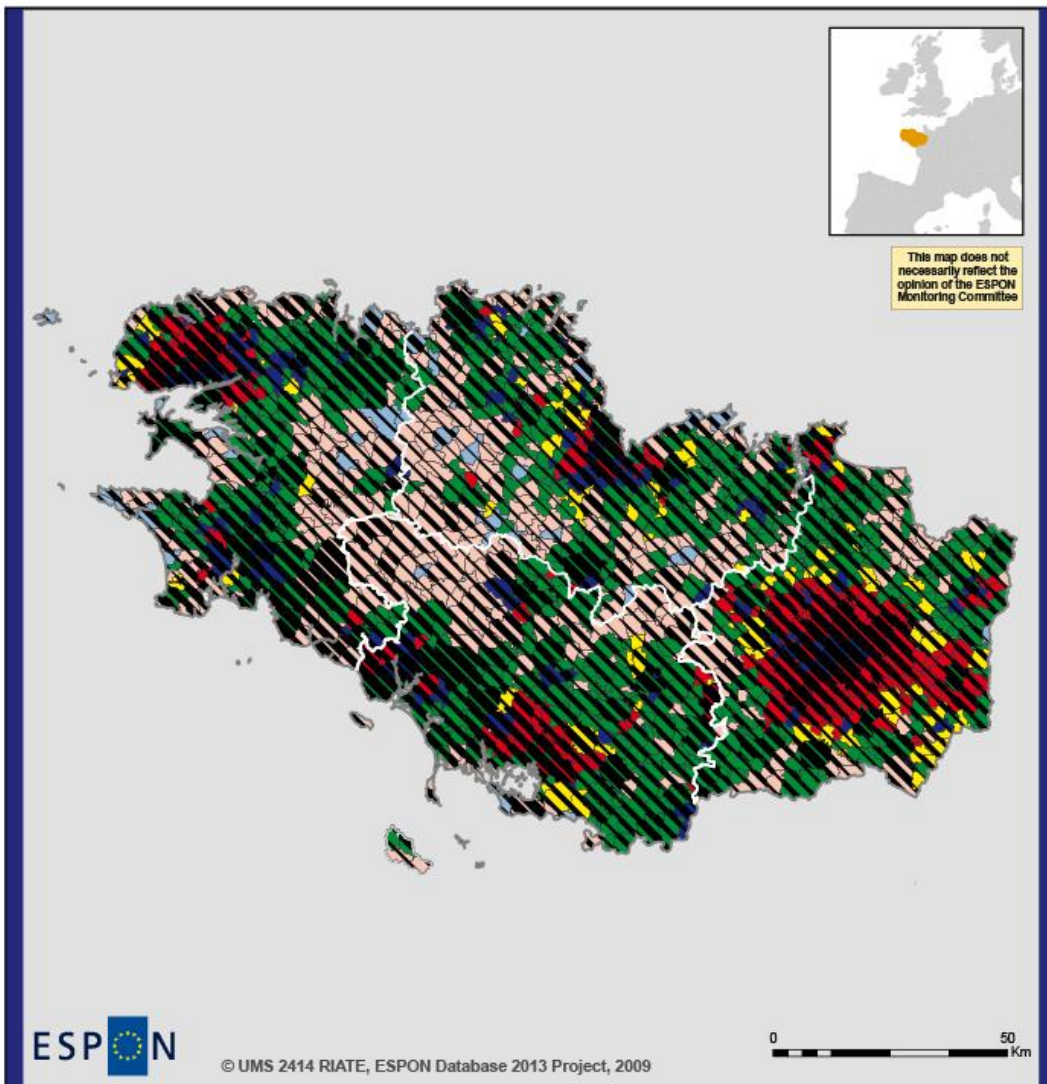
When these different elements are not correctly taking into account, the map will be characterised probably by some mistakes and misunderstandings.

## 2.1 Bad choices in term of representation of the data

Most of the problems of visualisation and map design are generally linked to **bad choices in term of representation of the data** (cf. part 2 of the technical report). When comparing **figures 13 and 14**, which represent the same information, e.g. a typology showing age structure and total population in the municipalities from Brittany (France), it is quite clear that the second map is really clearer than the first one. Two main reasons can explain it (**figure 13**):

- Absolute values (e.g. total population in 2000) don't have to be shown by variation of intensity of black (hachure). This kind of representation does not respect the ratio of proportionality of the indicator, which is fundamental and needed information. Using hachure is also a visual mistake; the map is not readable at all and the representation is not the most efficient. These data have to be shown by proportional symbols, circles for instance.
- This typology, derived from age structure cannot be considered as a qualitative data, since there is an implicit order when considering the progression in term of age. In concrete terms, showing each class by a different colour is not the best solution. To show correctly this data it is important to think about the goal of the map. Here, it is important to represent the municipalities described by high share of young, active and old people. As a consequence, it is important to differentiate these information (3 colours) and also to make possible the analyse of the graduation of the phenomenon (high/medium shares), e.g. using variation of intensity of these 3 colours.

The solution proposed in **figure 14** try to correct these different elements. The most adapted solution for the representation of these data is to combine circles and colours in order to make the map as clear as possible. On top of that, it allows nuancing the interpretation of the map, e.g. Brittany is a region where ageing is important, but it concerns specific small and rural cities.



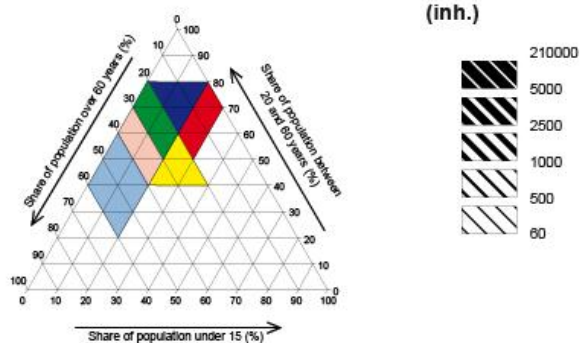
EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

Local level: LAU2  
Source: UMS 2414 RIATE, 2009  
Origin of data: INSEE, 2009  
© EuroGeographics Association for administrative boundaries

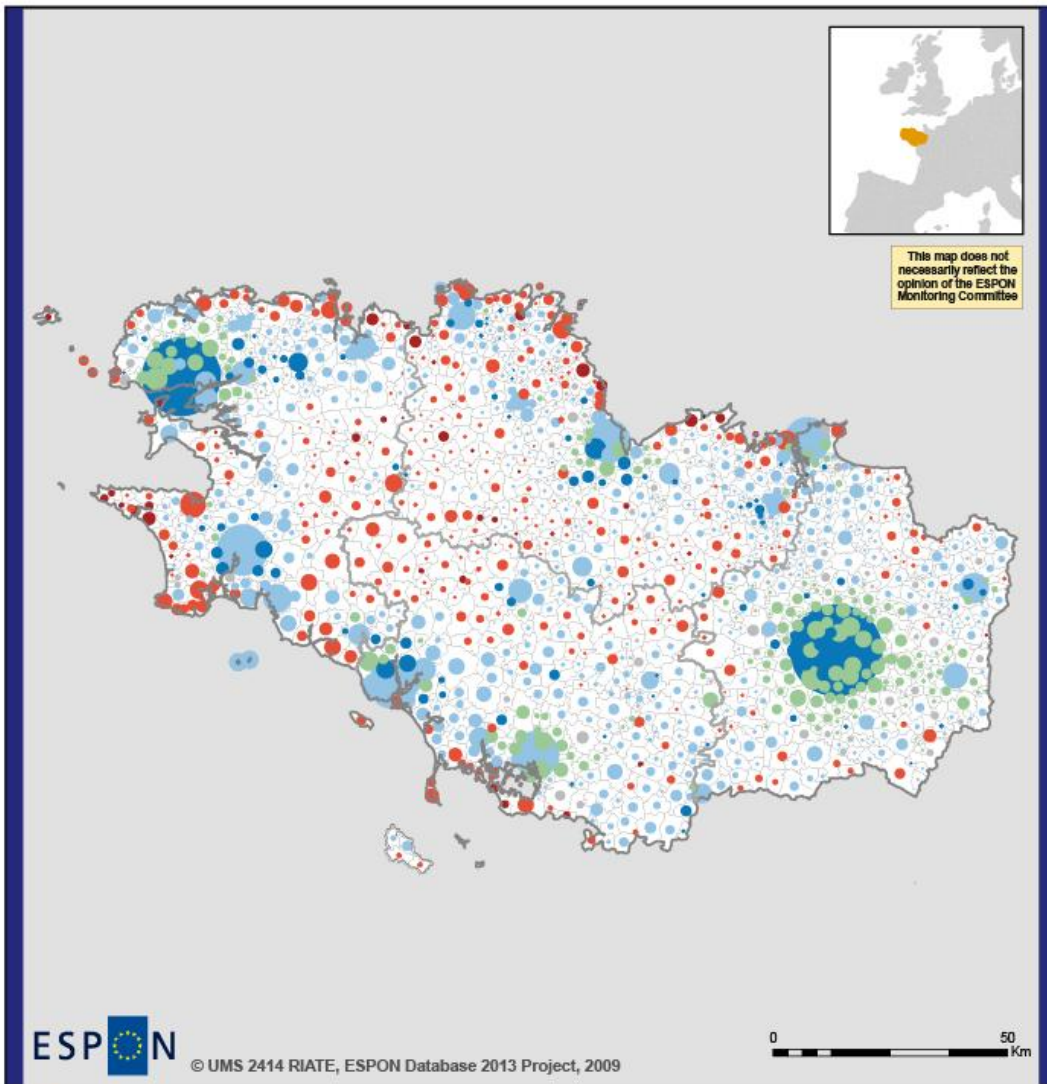
**TYPE OF AGE STRUCTURE IN 2000**

- A) Excedent of young population
  - Type A.1
- B) Excedent of active population
  - Type B.1
  - Type B.2
- C) Excedent of old population
  - Type C.1
  - Type C.2
- D) Medium profile
  - Type D

**TOTAL POPULATION IN 2000  
(inh.)**



**Figure 13: Population and age structure in Brittany (France) – with semiologic problems**



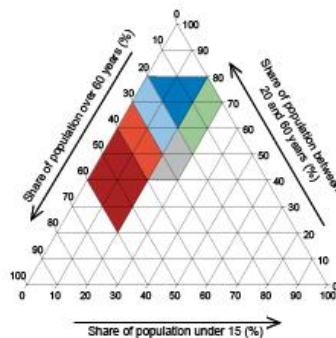
EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

Local level: LAU2  
Source: DG-IPOL, *Shrinking Regions: a paradigm shift in demography and territorial development*, European Parliament, 2008  
Origin of data: INSEE, 2009

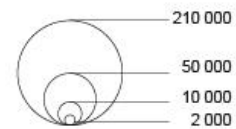
© EuroGeographics Association for administrative boundaries

#### TYPE OF AGE STRUCTURE IN 2000

- A) Excedent of young population
  - Type A.1
- B) Excedent of active population
  - Type B.1
  - Type B.2
- C) Excedent of old population
  - Type C.1
  - Type C.2
- D) Medium profile
  - Type D



#### TOTAL POPULATION IN 2000 (inh.)



**Figure 14: Population and age structure in Brittany (France) – without semiologic problems**

## 2.2 Improving the efficiency of the map

Other problem which appears regularly is the degree of complexity of the map. The aim of the maps is to be synthetic. When representing too much information, the eye cannot distinguish the different elements of the map. This kind of figure can be solved by thinking to the design of the map: where is the best location for legend? How using with the most efficiency the place available?

**The figures 15 and 16** show the same information, e.g. a typology of population development by components during the period 1995-2004 in EU27; this data is crossed with expected population evolution in 2030.

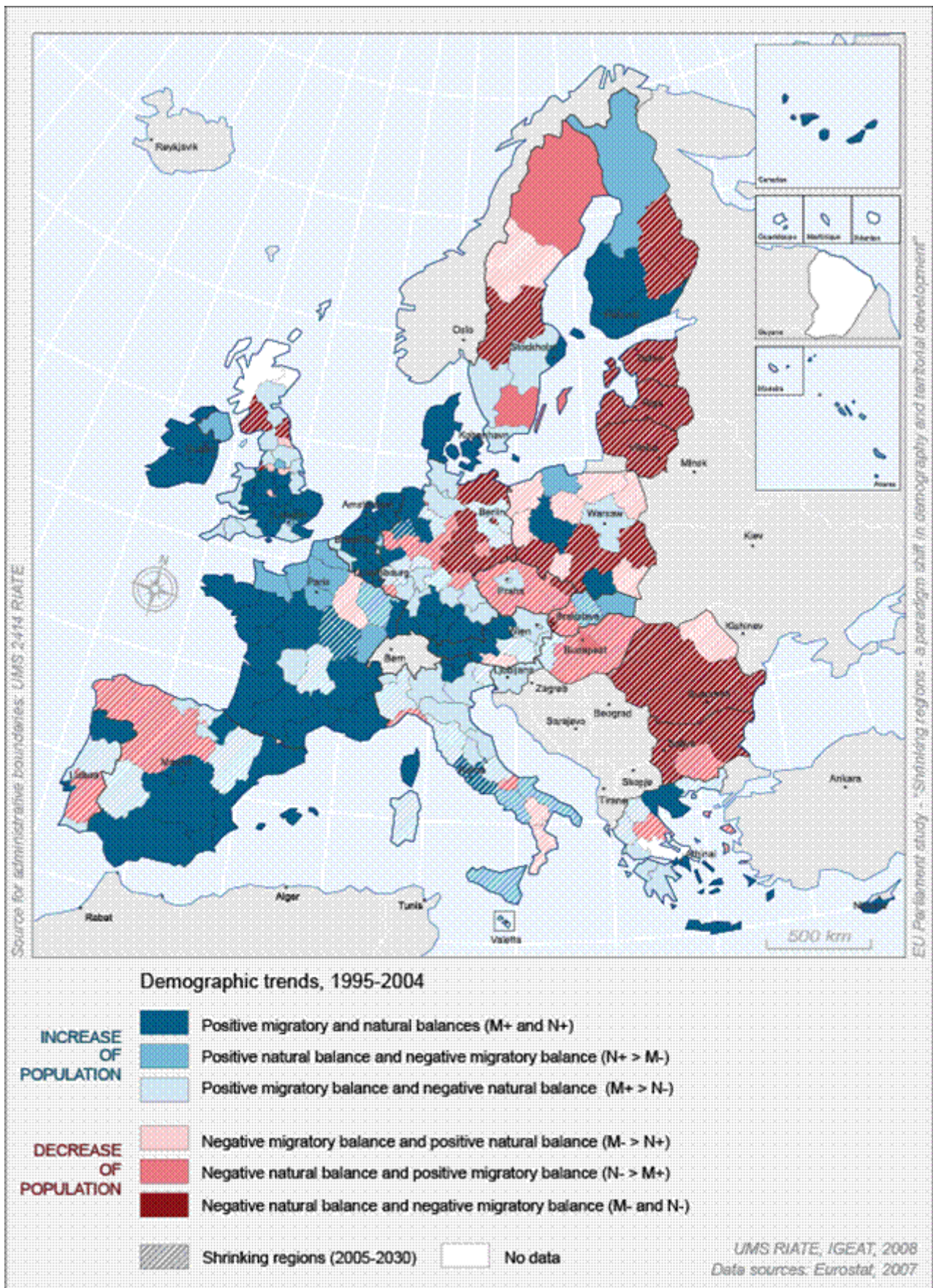
**Figure 15** proposes solution which is correct in term of graphic semiology: ordinal data are shown by variation of colour (green/red) and shrinking/non shrinking regions (qualitative data) are represented by the opposition of hachure and no hachure. However, the combination of these two visual variables makes the map hard to interpret and the message become not so clear!

When there is too much information it becomes difficult to be able to synthesise the message of the map. That is why in some cases it is more efficient to split information in two maps instead of concentrating all the elements in a single one. This has been done on figure 16, where the map located on left of the document shows the regions described by an expected growth of population; and the map on the right shows the regions where a demographic decrease is planned. This template allows immediately to observe that during the period 1995-2005 most of the 'shrinking regions' have witnessed a downturn linked to both natural change and a negative migratory balance.

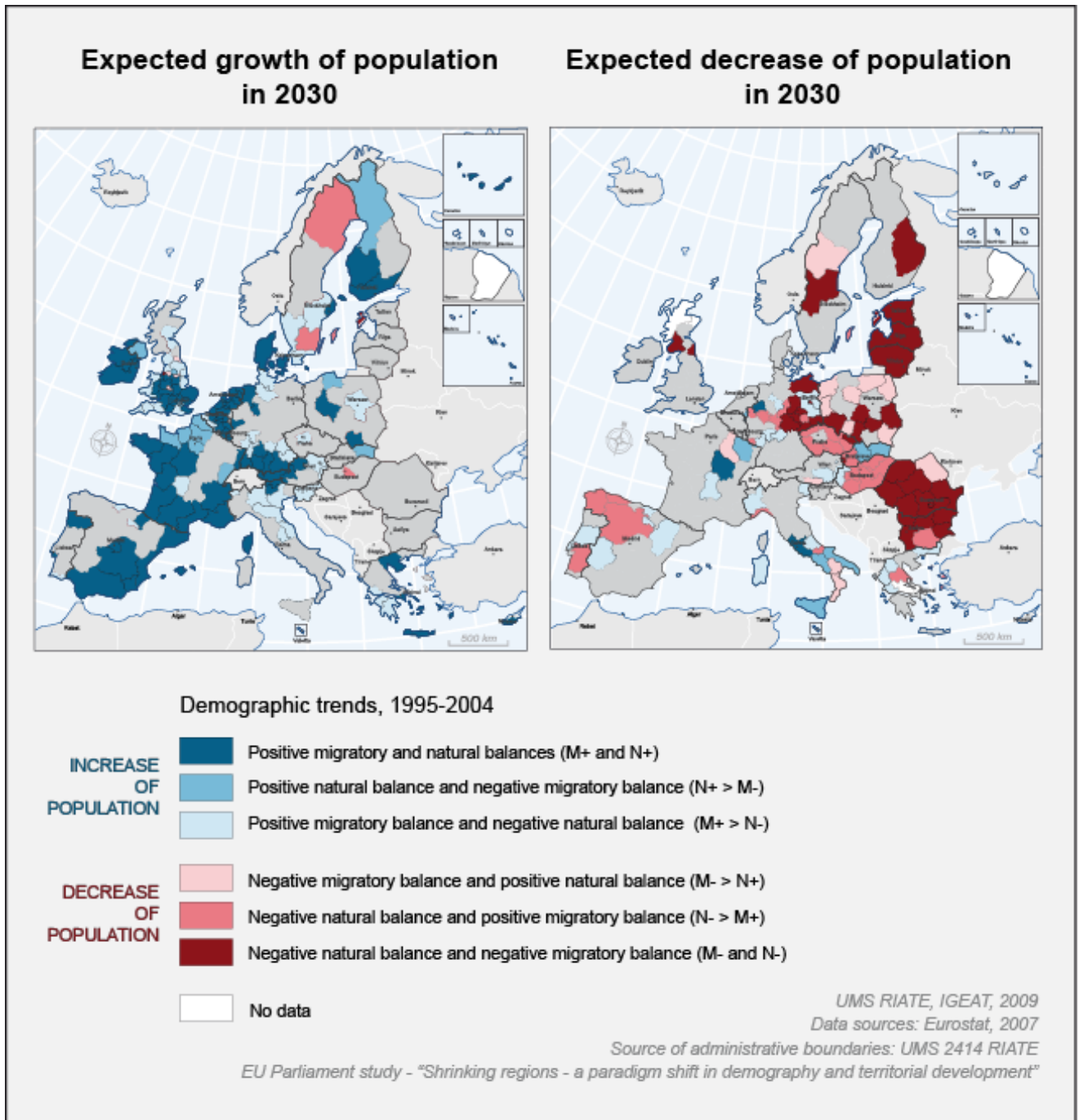
### ***There is never an optimal solution***

Whatever the examples proposed and demonstrated, it is important to keep in mind that there is never a single solution to show information on maps. In fact, each person has his own perception when interpreting graphic documents or pictures. **Map is always a compromise.** But during the creation of the map, is fundamental to try to make the map as understandable as possible. In concrete terms, it is not an obvious task and it is kindly recommended to make different attempts and share the results with other colleagues before saying "OK, my map is ready for the report"!





**Figure 15: Typology of regional growth patterns – Possibility 1**



**Figure 15: Typology of regional growth patterns – Possibility 2**

## ANNEXES

These annexes allow you to choose some efficient graphic variables to communicate differences in size, order or quality.





























































































### ANNEXE 1 - Relation of graphical variables to perceptual characteristics

Graphical variable	Type of data			
	nominal	ordinal	Interval/ratio	quantity
Size		x	x	x
Grey or colour value		x	x	
Grain/texture		x	x	
Colour hue	x			
Orientation	x			
Shape	x			

### ANNEXE 2 - Numbers of categories that can be perceived at a glance

Graphical variable	Point	Line	Area
Size	4	4	5
Grey or colour value	3	4	5
Grain/texture	2	4	5
Colour hue	7	7	8
Orientation	4	2	4
Shape	3	3	3

## ANNEXE 3: Differences in value or lightness

COLOUR INTENSITY			
Blue	Green	Red	Brown
<b>4 classes</b>			
 rgb(0,147,193)	 rgb(31,115,42)	 rgb(235,107,57)	 rgb(126,70,53)
 rgb(118,188,218)	 rgb(100,175,64)	 rgb(246,170,65)	 rgb(195,118,70)
 rgb(208,232,244)	 rgb(191,217,159)	 rgb(255,227,125)	 rgb(229,170,81)
 rgb(235,246,252)	 rgb(230,239,207)	 rgb(255,249,200)	 rgb(255,237,170)
<b>5 classes</b>			
 rgb(0,147,193)	 rgb(18,94,39)	 rgb(229,53,64)	 rgb(126,70,53)
 rgb(118,188,218)	 rgb(60,145,60)	 rgb(235,107,57)	 rgb(195,118,70)
 rgb(167,212,233)	 rgb(129,188,96)	 rgb(246,170,65)	 rgb(229,170,81)
 rgb(208,232,244)	 rgb(191,217,159)	 rgb(255,227,125)	 rgb(255,221,139)
 rgb(235,246,252)	 rgb(230,239,207)	 rgb(255,249,200)	 rgb(255,237,170)
<b>6 classes</b>			
 rgb(0,124,176)	 rgb(18,94,39)	 rgb(229,53,64)	 rgb(97,68,55)
 rgb(0,147,193)	 rgb(60,145,60)	 rgb(235,107,57)	 rgb(126,70,53)
 rgb(118,188,218)	 rgb(107,178,76)	 rgb(246,170,65)	 rgb(195,118,70)
 rgb(167,212,233)	 rgb(151,197,110)	 rgb(255,227,125)	 rgb(229,170,81)
 rgb(208,232,244)	 rgb(200,218,140)	 rgb(255,249,200)	 rgb(255,221,139)
 rgb(235,246,252)	 rgb(239,241,199)	 rgb(255,253,238)	 rgb(255,237,170)
<b>8 classes</b>			
 rgb(0,98,140)	 rgb(11,82,34)	 rgb(173,26,34)	 rgb(97,68,55)
 rgb(0,124,176)	 rgb(31,115,42)	 rgb(207,54,65)	 rgb(126,70,53)
 rgb(0,147,193)	 rgb(62,146,44)	 rgb(229,53,64)	 rgb(165,94,57)
 rgb(68,170,207)	 rgb(100,175,64)	 rgb(235,107,57)	 rgb(195,118,70)
 rgb(118,188,218)	 rgb(145,191,91)	 rgb(246,170,65)	 rgb(219,145,73)
 rgb(167,212,233)	 rgb(180,209,121)	 rgb(255,227,125)	 rgb(229,170,81)
 rgb(208,232,244)	 rgb(200,218,140)	 rgb(255,249,200)	 rgb(255,221,139)
 rgb(235,246,252)	 rgb(239,241,199)	 rgb(255,253,238)	 rgb(255,237,170)

## GREY VALUE

### 4 classes



### 5 classes



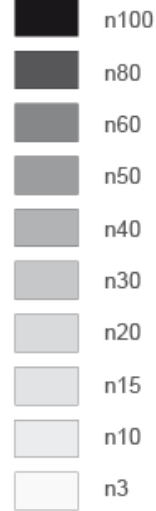
### 6 classes



### 8 classes



### 10 classes



## OPPOSITE COLOURS

 rgb(0,98,140)

 rgb(0,147,193)

 rgb(118,188,218)

 rgb(235,246,252)

 rgb(252,208,211)

 rgb(234,122,133)

 rgb(196,55,79)

 rgb(142,3,17)

 rgb(90,93,122)

 rgb(114,118,159)

 rgb(147,153,199)

 rgb(196,200,226)

 rgb(249,230,239)

 rgb(240,184,210)

 rgb(236,141,181)

 rgb(226,2,128)

 rgb(32,115,43)

 rgb(66,145,44)

 rgb(145,191,92)

 rgb(222,229,157)

 rgb(247,229,196)

 rgb(250,210,147)

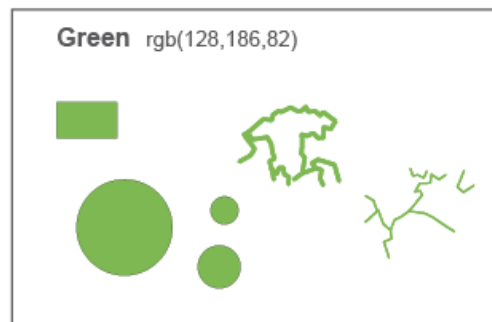
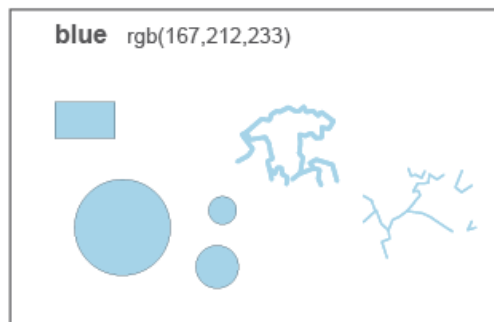
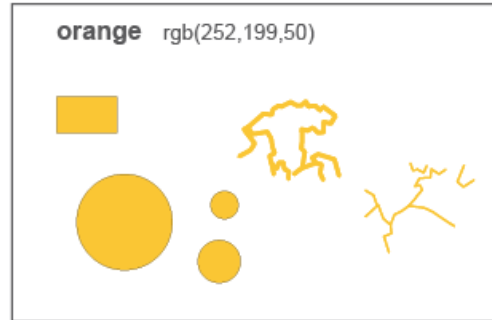
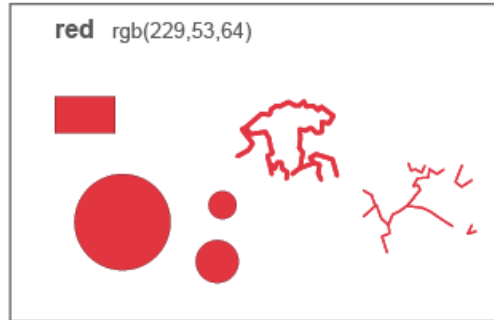
 rgb(244,171,42)

 rgb(175,110,22)

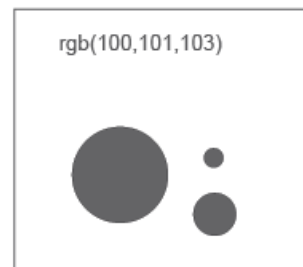
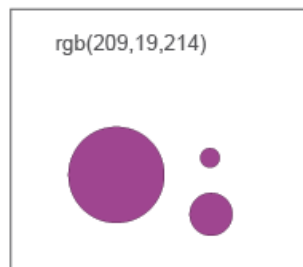
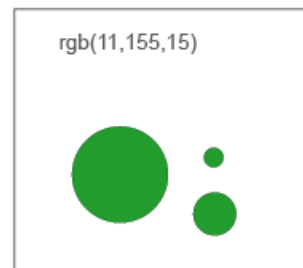
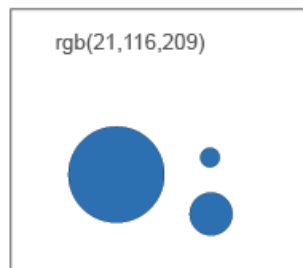
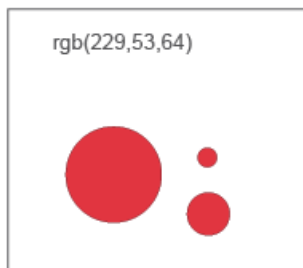
# ANNEXE 4: Colours for differences typology or qualitative value

## QUALITATIVE VALUES

*(circles and discontinuities)*



*(circles)*



## References

### • *Litterature*

Béguin M., Pumain D., 2003, *La représentation des données géographiques – statistique et cartographie*, Armand Colin.

Bertin J., 1967, *Sémiologie graphique*, Gauthiers-Villars.

Cambrezy L., de Maximy R. (Ed.), 1995, *La cartographie en débat, représenter ou convaincre*, Editions Kathala et Orstom, Paris

Harris R. L., 1996, *Information graphics, a comprehensive illustrated reference, visual tools for analysing, managing and communicating*, Management Graphics ed., USA

Harley, J. B., 1988, *Maps, knowledge and power*. In COSGROVE, D. (Ed.) *The Iconography of Landscape*. Cambridge, MA, Cambridge University Press.

Kraak M.-J., Ormeling F., 2003, *Cartography, Visualization of Geospatial Data*, 2<sup>nd</sup> edition, Pearson Education, Prentice Hall.

Kraak, M.-J., 1998, *Exploratory cartography, map as tools for discovery*, *ITC Journal* (1), pp.46-54

MacEachren A.M., 1994, *Some truth with maps: a primer on design and symbolization*, Association of American Geographers, Washington DC.

Monmonnier M., 1996, *How to lie with maps*, University of Chicago Press.

Robinson A.H., Morrison J.L., Muehrcke P.C., 1995, *Elements of cartography*, New York, J.Willey & Sons.

Wilkinson L., 1999, *The grammar of graphics*. New York, Springer.

Wood, D., 1992, *The Power of Maps*. New York, The Guildford Press.

Wood C. H., Keller C. P., 1996, *Cartographic design: theoretical and practical perspectives*, Wiley, USA

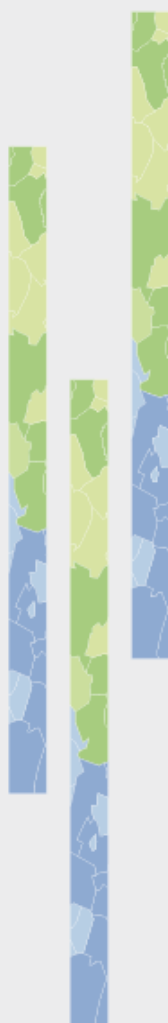
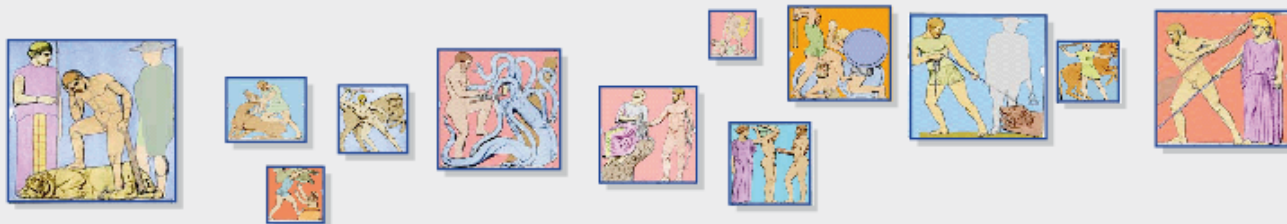
Zanin C., Trémélo M-L, 2003, *Savoir faire une carte: Aide à la conception et à la réalisation d'une carte thématique univariée*, Belin.

### • *Websites*

**Colorbrewer 2.0** is an online tool designed to help people select good color schemes for maps and other graphics: <http://colorbrewer2.org/>

**Philcarto** is a free tool for cartography, available on the net: <http://philcarto.free.fr/>

**Quantum GIS** is an Open Source Geographic Information System. It runs on Linux, Unix, Mac OSX, and Windows and supports numerous vector, raster, and database formats and functionalities: <http://www.qgis.org/>



## WORLD DATABASE

Global data for the ESPON 2013  
Database

### CONTENT

- Description of the ESPON 2013 WORLD DATABASE
- Overview and description of a sample of world databases
- Comparison between world databases and Eurostat databases

**ESPON 2013 DATABASE**  
MARCH 2011



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

50 pages



# LIST OF AUTHORS

Hy Dao, UNEP/GRID-Genève

Andrea de Bono, UNEP/GRID-Genève

Claude Grasland, UMS 2414 RIATE

Nicolas Lambert, UMS 2414 RIATE

Ronan Ysebaert, UMS 2414 RIATE

## Contact

[hy.dao@unepgrid.ch](mailto:hy.dao@unepgrid.ch)

[debono@unepgrid.ch](mailto:debono@unepgrid.ch)

[claudе.grasland@parisgeo.cnrs.fr](mailto:claudе.grasland@parisgeo.cnrs.fr)

[nicolas.lambert@ums-riate.fr](mailto:nicolas.lambert@ums-riate.fr)

[ronan.ysebaert@ums-riate.fr](mailto:ronan.ysebaert@ums-riate.fr)

UNEP/GRID-Genève

Tel. (+ 41) 22 917 82 40

UMS RIATE

Tel. (+ 33) 1 57 27 65 32

## TABLE OF CONTENT

<b>1 Description of the ESPON 2013 WORLD DATABASE .....</b>	<b>4</b>
1.1 Indicators.....	4
<i>Data sources.....</i>	<i>4</i>
<i>Indicators tables .....</i>	<i>4</i>
<i>Fields description.....</i>	<i>4</i>
<i>Metadata.....</i>	<i>4</i>
<i>Notes.....</i>	<i>5</i>
1.2 Country reference tables .....	6
<i>Description of the tables.....</i>	<i>6</i>
<i>Suggested criteria for selecting a representative countries list for the World Database .....</i>	<i>8</i>
1.3 Overview and description of a sample of World databases .....	10
<i>CHELEM Database .....</i>	<i>10</i>
<i>ESPON 2006 EUROPE IN THE WORLD DATABASE: The WUTS System .....</i>	<i>10</i>
<i>UN Standard countries or area and geographical regions .....</i>	<i>11</i>
<i>World Bank: The World Development Indicators (WDI) .....</i>	<i>12</i>
<i>The Global Environment Outlook (GEO) Data Portal .....</i>	<i>13</i>
<b>2 Comparison between World databases and Eurostat databases.....</b>	<b>15</b>
2.1 Methodology.....	15
<i>Verification Phase.....</i>	<i>15</i>
<i>Calculate the differences EIE-EIW.....</i>	<i>16</i>
<i>Thresholds.....</i>	<i>16</i>
<i>Typologies of differences .....</i>	<i>16</i>
2.2 Results .....	18
<i>Total population data .....</i>	<i>18</i>
<i>Population by sex and age groups .....</i>	<i>19</i>
<b>3 Conclusion.....</b>	<b>22</b>
<b>Annex 1. List of EIW indicators.....</b>	<b>23</b>
<b>Annex 2.1. LIST OF GEOGRAPHICAL UNITS FROM CHELEM (PARTITION IN 96 ENTITIES).....</b>	<b>26</b>
<b>Annex 2.2. LIST OF GEOGRAPHICAL UNITS (168 UNITS) FROM ESPON 2006 PROGRAM (EUROPE IN THE WORLD) .....</b>	<b>29</b>
<b>Annex 2.3. LIST OF GEOGRAPHICAL UNITS FROM GEO .....</b>	<b>33</b>
<b>Annex 2.4. LIST OF GEOGRAPHICAL UNITS FROM WDI.....</b>	<b>39</b>
<b>Annex 2.5. LIST OF GEOGRAPHICAL UNITS FROM UN (WPP08).....</b>	<b>44</b>
<b>Databases references .....</b>	<b>50</b>

# Introduction

The first obvious aim of this challenge is to provide data for ESPON projects working at global scale, such as the new projects on "Globalisation" launched in February 2010.

Another important objective is to fill the gaps in some discontinuous time series at NUTS2 or NUTS3 levels by means of disaggregation of time series available at national levels.

The work done by UMS RIATE and expert team UNEP on this challenge is summarised in this technical report entitled "World database".

The first section describes the ESPON 2013 World Database by defining the base indicators considered and by introducing the notion of "units of reference" compiled in the "country reference tables".

Secondly, we have considered the lists of countries from main international "thematic" providers. In fact, the definitions of "what is a country" for each provider do not correspond in several cases. The second section shows concretely this fact.

The section 3 focuses on the linking of World data with Eurostat regional data. Our goal has been to design a methodological tool (named "Gap Tracker") for explaining the differences between global databases and Eurostat data.

This Technical Report shows results attained during the course of the ESPON 2013 Database Project.

# 1 Description of the ESPON 2013 WORLD DATABASE

The version 2.0 of the ESPON 2013 World Database can be subdivided into two main components (described in detail in the next sections):

- *indicators*, i.e. the data values, with a global coverage, mainly from international organizations as well as data provided by Eurostat, which cover the European region;
- *country reference tables*, i.e. the country units and the regional/thematic aggregations used by global providers and Eurostat.

## 1.1 Indicators

We focused our data collection and testing methodologies on three groups of variables: population, economical indicators and carbon dioxide emissions. A complete list of indicators (March 2011) included in the database can be found in the annexes.

### Data sources

- **Population:** United Nations/Population Division with the World Population Prospects (WPP2008)  
<http://www.un.org/esa/population/unpop.htm>
- **CO2 emissions:** UNFCCC includes official data reported by countries (Annex I parties), and [http://unfccc.int/ghg\\_data/items/3800.php](http://unfccc.int/ghg_data/items/3800.php); CDIAC where data are calculated from the energy statistics of the UN yearbook <http://cdiac.ornl.gov/trends/emis/overview.html>
- **Economical data:** World Bank World Development Indicators 2010 edition (WDI 2010) <http://data.worldbank.org/data-catalog/world-development-indicators/wdi-2010>

### Indicators tables

The indicators tables (see Figure 1) include:

- **wpp2008\_data** contains 157 variables extracted from the stocks indicators of the UN WPP2008;
- **co2\_data** contains 8 variables derived from both CDIAC and UNFCCC
- **wdi\_data** contains a set of 6 economical indicators from the WDI
- **eu2009\_pop\_stocks** (Europe, not on Figure 1) contains 42 population stocks indicators extracted from the Eurostat databases. These indicators were used in the comparison phase (see in next chapter)

### Fields description

- "category" (text): indicator full name ex. population sex ratio
- "iso3\_code" (text): original country code
- "id\_metadata" (integer): data source code; linked with table metadata
- "1950", ..., "2050" (double): value for each year

### Metadata

All the indicators are with their metadata entered following the ISO 19115:2003/19139 metadata standard.

Very basic metadata fields (such as “abstract”, “temporal extent”) are included in the metadata table. The complete information (ISO19115) is included in separate xml files, annexed to the database.

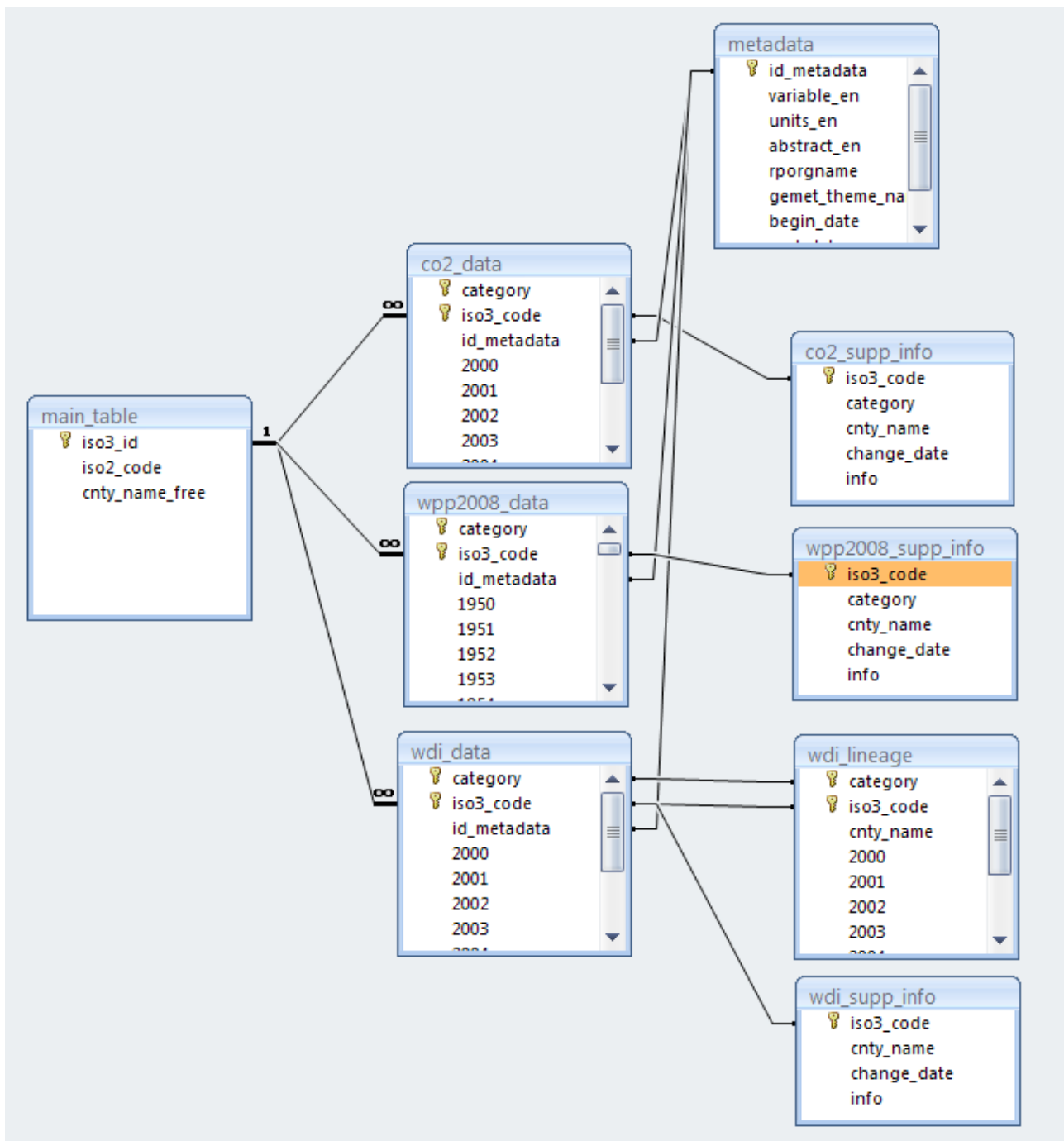
Tables **wdi\_supp\_info**, **co2\_supp\_info**, and **wpp2008\_supp\_info** hold information concerning the country status (e.g. change of names over the time) and its relationships to other territories, if applicable (e.g. United Kingdom includes data for Channel Islands).

Population and CO2 emissions indicators do not display significant data gaps in their time-series and it was not necessary to fill missing values.

In the case of the economical data, several estimates were done by the WDI staff. The table **wdi\_lineage** lists the calculated values.

### **Notes**

Population data from WPP08 from 2009 to 2050 are projections calculated by the Population Division staff on the basis of the “Medium Fertility Variant”.



**Figure 1.** Indicators and metadata relationships

## 1.2 Country reference tables

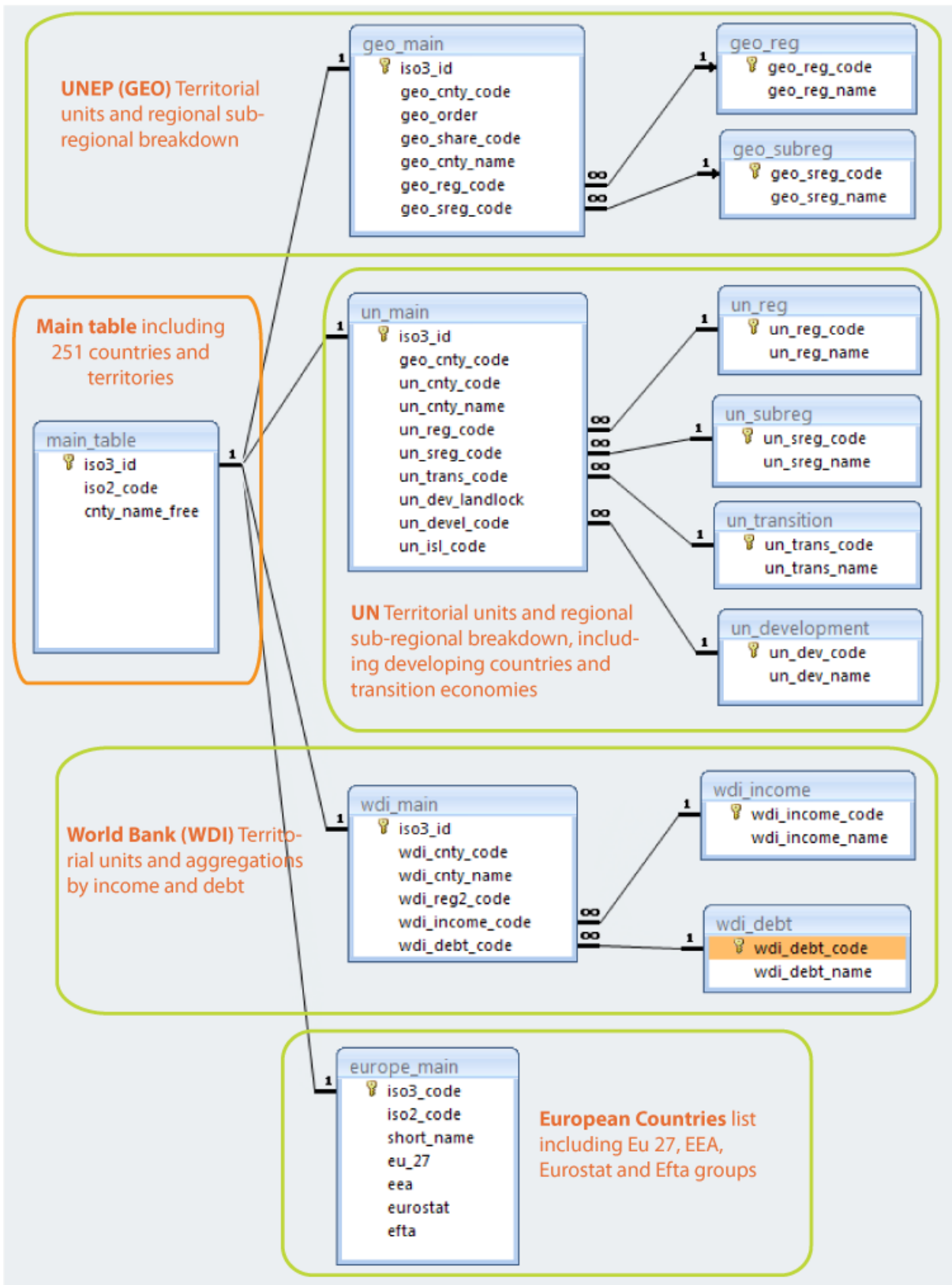
### Description of the tables

This chapter deals with the countries/territories and regional/thematic aggregations as supplied by our main data providers. The data sources are described in detail in the next chapter: "Overview and description of a sample of world databases".

The database includes a table named **main\_table** comprising all the basic territorial units and their codes furnished by our providers. It includes 251 countries/territories and their ISO alpha 3 and 2 codes. Tables **un\_main**, **wdi\_main**, **geo\_main** include countries/territories with their regional and or thematic aggregations respectively from

UN, World Development Indicators and GEO, with their original country codes and the ISO3 code field as primary key.

Table **europa\_main** follows the same structure.



**Figure 2.** reference countries tables and regional aggregations relationships

## **Suggested criteria for selecting a representative countries list for the World Database**

Our aim consists to suggest a list of selected countries/territories to be included in the World Database as reference countries.

These territorial units must be significant in terms of population, area, political status and, of course, within a consistent amount of statistical data over time.

Several approaches can be developed to select them, such as thresholds based on area and/or population size, economy. Every approach has positive and negative aspects.

We consider the official list of countries from four main international thematically focused providers:

- UN (population)
- FAO (land)
- World Bank - World Development Indicators WDI (economics)
- UNEP-GEO (environment)

The number of countries and the definition of "what is a country" for each provider do not correspond in several cases: for example Gibraltar is considered as a separate entity for all sources (GIB), except World Bank that includes them into United Kingdom.

**The first criterion of selection:** a territorial unit must be a UN Member. This is the case for 192 countries, with a correspondence between the UN, FAO, WDI and GEO codes

**The second criterion** is based on the maximum agreement between providers to consider the geographic entity in the same way (58 territories to analyze):

- if all 4 providers agree that the territory is a "separate entity" (individual code), it is selected (18);
- if all 4 providers consider the territory as integrated to a parent entity (no individual code), it is not selected (7).

**The third criterion** is applied for those 59 Territories that escaped from the above selection. In this case, the territorial entity must be recognised by at least 3 providers including UN to enter the list.

Hong Kong and Macau are only listed in the UN database, but for economical and demographic reasons they will be added to the list. Antarctica will be integrated too

In conclusion the suggested list of countries of reference for the World database includes:

- 192 UN member states (with perfect correspondence between UN, FAO, WDI and GEO codes);
- 18 territories complying with the second criterion;
- 18 territories complying with the third criterion;
- territories added as exceptions.

The list of countries is included in the table "**world\_database\_cnty\_list**" in the World Database.



iso3_id	cnty name	Status	un_iso3	geo_iso3	fao_iso3	wdi_iso3	population	area	un	geo	fao	wdi	selected
PCN	PITCAIRN	Non-Self-Governing Territory of the UK	PCN	PCN	PCN	GBR	50	47					
VAT	HOLY SEE (VATICAN CITY STATE)	The City of Vatican	VAT	VAT	VAT	ITA	783	1					
TKL	TOKELAU	Non-Self-Governing Territory of New Zealand	TKL	TKL	TKL	NZL	1'181	10					
NIU	NIUE	Self-Governing Territory in free association with New Zealand	NIU	NIU	NIU	NZL	1'557	260					
FLK	FALKLAND ISLANDS (MALVINAS)	Non-Self-Governing Territory of the UK	FLK	FLK	FLK	GBR	2'998	12173					
SHN	SAINT HELENA	Non-Self-Governing Territory of the UK	SHN	SHN	SHN	GBR	4'578	413					
MSR	MONTERRAT	Non-Self-Governing Territory of the UK	MSR	MSR	MSR	GBR	5'875	102					
SPM	SAINT PIERRE AND MIQUELON	Territorial Collectivity of France	SPM	SPM	SPM	FRA	6'096	242					
AIA	ANGUILLA	Non-Self-Governing Territory of the UK	AIA	AIA	AIA	GBR	14'486	102					
WLF	WALLIS AND FUTUNA	Oversea Territory of France	WLF	WLF	WLF	FRA	15'100	274					
COK	COOK ISLANDS	Self-Governing in free association with New Zealand	COK	COK	COK	NZL	19'531	237					
VGB	VIRGIN ISLANDS, BRITISH	Non-Self-Governing Territory of the UK	VGB	VGB	VGB	GBR	22'539	153					
GIB	GIBRALTAR	Non-Self-Governing Territory of the UK	GIB	GIB	GIB	GBR	31'002	7					
TCA	TURKS AND CAICOS ISLANDS	Non-Self-Governing Territory of the UK	TCA	TCA	TCA	GBR	32'413	430					
CYM	CAYMAN ISLANDS	Non-Self-Governing Territory of the UK	CYM	CYM	CYM	GBR	55'022	262					
ESH	WESTERN SAHARA	Non-Self-Governing Territory	ESH	ESH	ESH		480'000	266000					
NFK	NORFOLK ISLAND	Australian Territory	NFK	NFK	NFK	AUS	-	35					
XAA	CHANNEL ISLANDS	Self-Governing Territory of the UK	XAA	GBR	XAA	XAA	149'288						
MAC	MACAO	Province of China	MAC	CHN	CHN	MAC	513'076						
HKG	HONG KONG	Province of China	HKG	CHN	CHN	HKG	6'947'843						
ATA	ANTARCTICA	The Antarctic Treaty System		ATA			-	1.4E+07					
GGY	GUERNSEY	British Crown Dependency	GGY	GGY	GBR	GBR	65'573	78					
JEY	JERSEY	British Crown Dependency	JEY	JEY	GBR	GBR	91'321	116					
ALA	ÅLAND ISLANDS	Autonomous province of Finland	ALA	FIN	FIN	FIN	-						
SJM	SVALBARD AND JAN MAYEN	Norwegian Territory	SJM	NOR	NOR	NOR	-	61397					
BLM	SAINT BARTHÉLEMY	Oversea Department of France	BLM	FRA	FRA	FRA	-						
WAK	WAKE ISLAND	US Territory		WAK			-	7					
JTN	JOHNSON ATOLL	US Territory	USA	JTN	USA	USA	-	3					
MID	MIDWAY ISLANDS	US Territory	USA	MID	USA	USA	-	6					
CCK	COCOS (KEELING) ISLANDS	Australian Territory	AUS	CCK	AUS	AUS	-	14					
CXR	CHRISTMAS ISLAND	Australian Territory	AUS	CXR	AUS	AUS	-	135					

provider code corresponds to iso3\_id     
 Territories included in the list  
 provider code not corresponds to iso3\_id     
 Territories not included in the list

**Table1.** The third criterion of selection for non UN Members territories

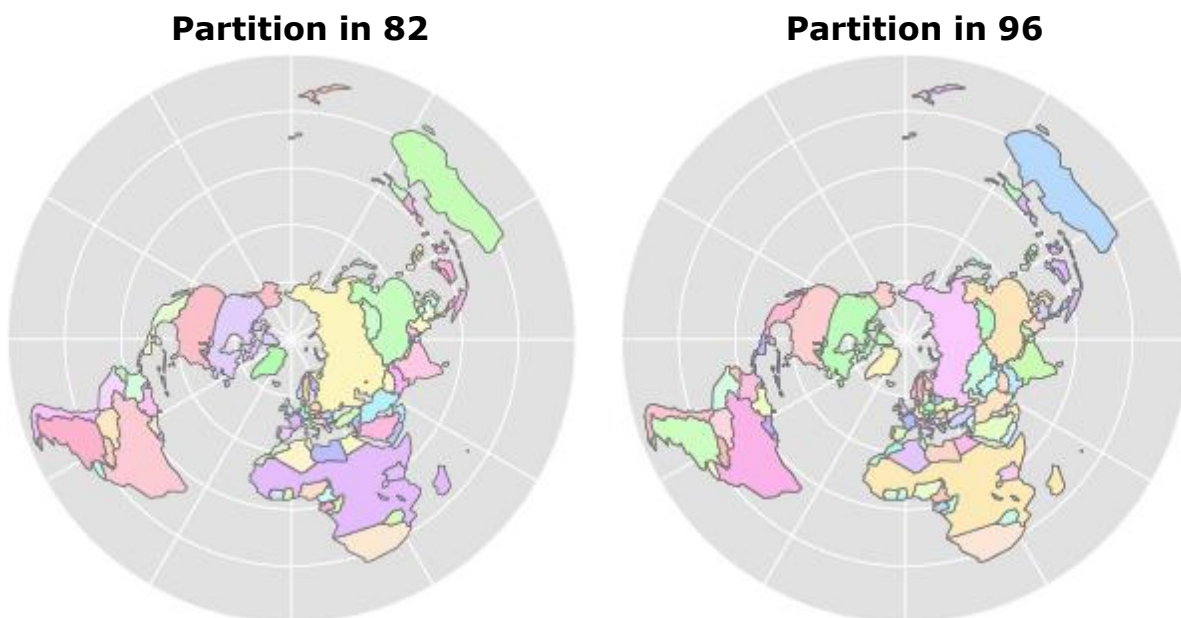
## 1.3 Overview and description of a sample of World databases

Several World databases exist. However, they do not describe their spatial units in a same way. A first work consists to identify the structure of reference of each of databases. The complete description of geographical units is presented in annexe 2.

### CHELEM Database

CHELEM is an economic long term database constructed by the Centre d'Etudes Prospectives et d'Informations Internationales (CEPII). The database covers the period from 1960 to present. The aim of this database is to provide a coherent view of the world economy. This database is composed of 3 sub-databases: international trade (1) GDP (2) and balance of payments (3).

CHELEM is based on a specific geographic classification with two kinds of partition. The partition in 96 zones (available from 1993 onwards) gives the maximal detail for trade. The partition in 82 zones does not detail the countries resulting from former Yugoslavia, USSR and Czechoslovakia.

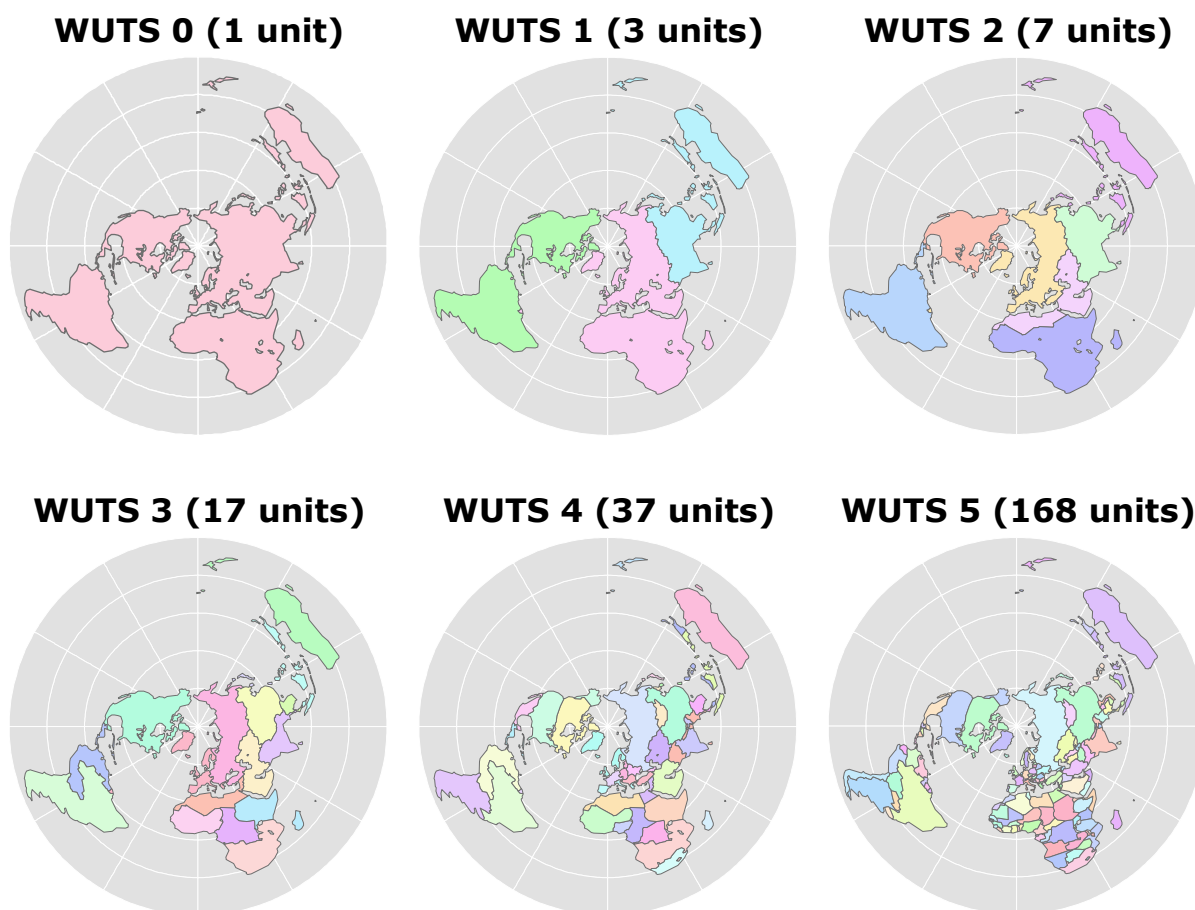


**Figure 3.** CHELEM Database subdivisions

### ESPON 2006 EUROPE IN THE WORLD DATABASE: The WUTS System

Realised in the first ESPON programme, this world database is based on a precise list of 168 states that represent a minimum of 1/10'000<sup>th</sup> of the population, GDP or area of the World. This list of 168 states provides a clear basis for data collection in an harmonised way, all states being identified by a specific code (WUTS CODE).

The WUTS (World Unified Territorial System) is a harmonised hierarchical system of World divisions, which is directly inspired from the NUTS (Nomenclature of Territorial Units for Statistics) created by Eurostat more than 25 years ago in order to provide a single uniform breakdown of territorial units for the production of regional statistics for the European Union. The WUTS is composed of 5 hierarchical levels, from the level of states (WUTS5) to the level of the World (WUTS0).



**Figure 4** : Europe in the World subdivisions

### **UN Standard countries or area and geographical regions**

The list of countries or areas includes those countries or areas for which the Statistics Division of the United Nations Secretariat compiles statistical data. The names of countries or areas refer to their short form used in day-to-day operations of the United Nations and not necessarily to their official name as used in formal documents.

The geographical regions and groupings of countries and areas are not comprehensive but only a selection, which are or may be used in the compilation of statistics. In order to ensure consistency in statistics and for convenience, each country or area is shown in one region only. The macro geographical regions are arranged to the extent possible according to continents. Within these groupings more detailed component of geographical regions are shown.

The group of least developed countries (LDCs), as defined by the United Nations, comprises 49 countries, out of which 33 are in Africa, 10 in Asia, 1 in Latin America and the Caribbean, and 5 in Oceania.

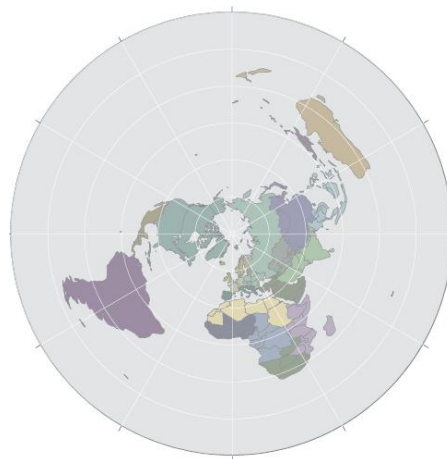
Note that there is no established convention for the designation of "developed" and "developing" countries or areas in the United Nations system: the designations "developed" and "developing" are intended for statistical convenience and do not necessarily express a judgement about the stage reached by a particular country or area in the development process.

Criteria for identification of LDCs and Landlocked developing countries can be found at <http://www.unohrlls.org/en/ldc/related/59/>.

### UN regional Breakdown



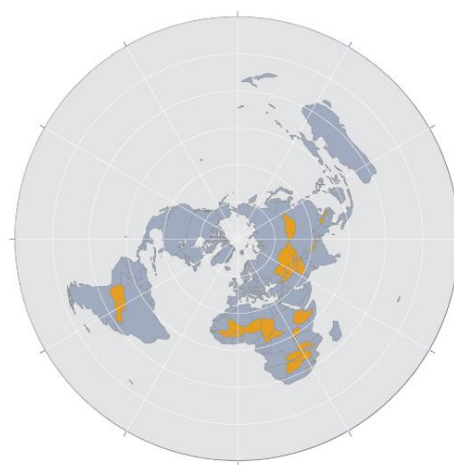
### UN sub-regional Breakdown



### UN Developing countries



### UN Landlocked developing countries



**Figure 5:** United Nations aggregations

### **World Bank: The World Development Indicators (WDI)**

The World Development Indicators (WDI) 2009 is the statistical benchmark that helps measure the progress of development.

The 2009 WDI includes more than 800 indicators organised in 6 sections: World View, People, Environment, Economy, States and Markets, and Global Links.

Data are shown for all World Bank member countries (185), and all other economies with populations of more than 30,000 (209 total)

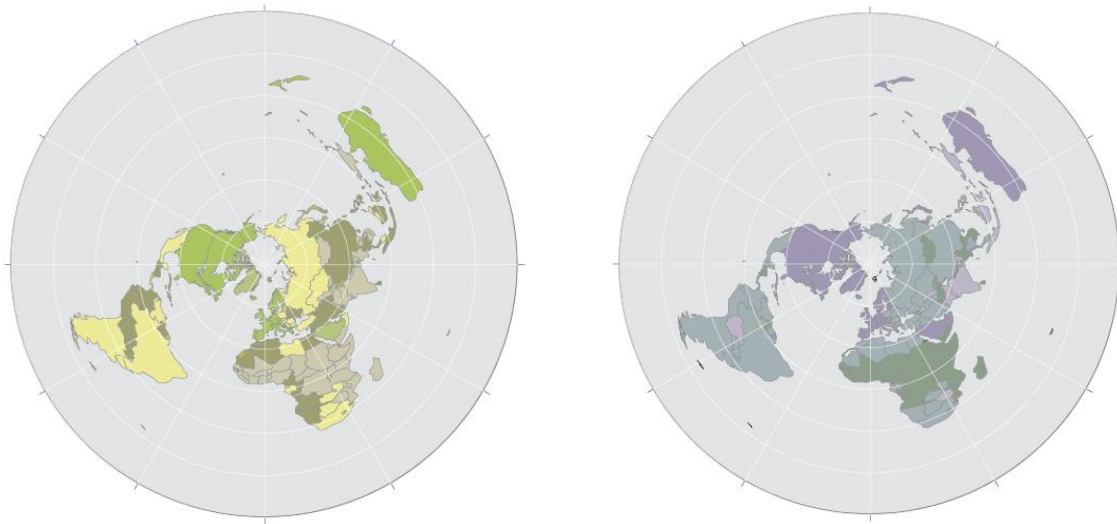
For operational and analytical purposes, the World Bank's main criterion for classifying economies is gross national income (GNI) per capita. Based on its GNI per capita, every economy is classified as low income, middle income (subdivided into lower middle and upper middle), or high income. Other analytical groups based on geographic regions are also used.

Geographic region: classifications and data reported for geographic regions are for low-income and middle-income economies only. Low-income and middle-income economies are sometimes referred to as developing economies. The use of the term is convenient; it is not intended to imply that all economies in the group

are experiencing similar development or that other economies have reached a preferred or final stage of development. Classification by income does not necessarily reflect development status.

WDI aggregations by Income

WDI aggregations by Debt



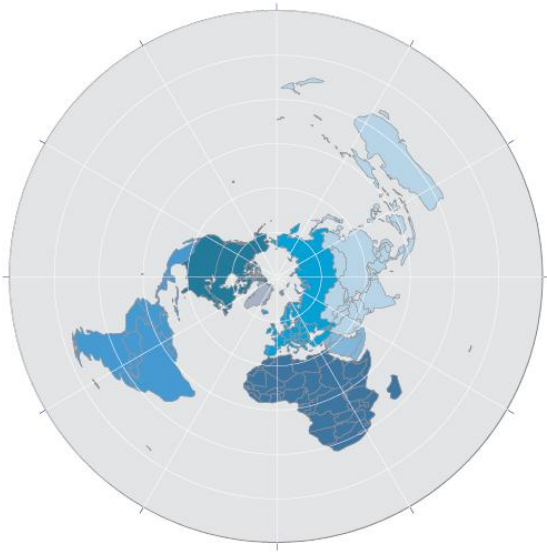
**Figure 6** : WDI aggregations

### **The Global Environment Outlook (GEO) Data Portal**

The GEO Data Portal gives access to a broad collection of harmonised environmental and socio-economic datasets from authoritative sources at global, regional (7), sub-regional (23) and national (237) levels. There is no established convention for the designation of regional and sub-regional groups. Geographical aggregations are arranged to the extent possible according to continents. Some inconsistencies exist: for example French Guyana is incorporated in the South America in regional aggregations, but in a political point of view belongs to Europe.

In the other way Israel could be included by its geographic position in West Asia but it is comprised de facto to the Western Europe group.

GEO regional Breakdown (7)



GEO sub-regional Breakdown (23)



**Figure 7** : GEO aggregations

## 2 Comparison between World databases and Eurostat databases

### 2.1 Methodology

Based on the results of ESPON 2006 Programme, we examined in a systematic way how to combine datasets at World/neighbourhood levels (where basic territorial units are the states) and datasets at European/regional levels (where basic territorial units are NUTS2 or NUTS3 units).

Our focus in this chapter is to explain the observed differences in the indicators values, for the same geographical unit, between the global and European databases. The compared datasets are:

- Europe in the ESPON database (EIE)
  - o Provider: mainly Eurostat
  - o Coverage: Eurostat countries
- Europe in the World database (EIW)
  - o Provider: International Organizations (eg UN, FAO)
  - o Coverage: global but the check/comparison is done only between countries matching with Eurostat coverage

In order to increase compatibility between EIE and EIW datasets, we set up a process of systematic analysis of the differences between the two datasets. This process is called the "Gap Tracker" method. The steps are described below.

#### Verification Phase

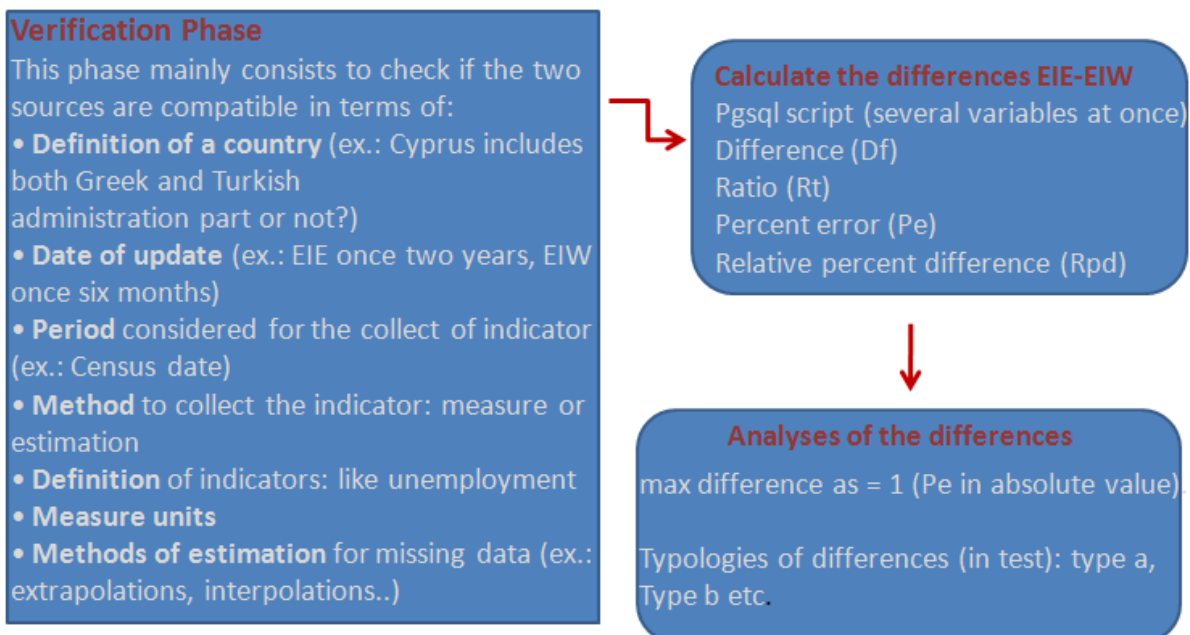
This phase mainly consists to check whether the two sources are compatible in terms of:

- definition of a country (e.g. does the term "Cyprus" include both Greek and Turkish administrated parts or not?);
- date of update (e.g. EIE once every two years, EIW once every six months) ;
- period considered for the data acquisition (e.g. census date) ;
- method to collect the data: measure or estimation ;
- definition of indicators (e.g. criteria for defining unemployment) ;
- measurement units ;
- methods of estimation for missing data (e.g. extrapolations, interpolations..).

# WORLD DICTIONARY OF UNITS

Comparison between world databases and Eurostat databases: preliminary results.

Europe in the ESPON database (EIE) → Europe in the World database (EIW)



**Figure 8** : Summary of the methodology used in the testing phase

## Calculate the differences EIE-EIW

Four simple formulas are used for calculating the differences between EIE and EIW data :

- **Difference (Df)** =  $EIW - EIE$
- **Ratio (Rt)** =  $EIW / EIE$
- **Percent error (Pe)** =  $((EIW - EIE) / EIE) * 100$
- **Relative percent difference (Rpd)** =  $((EIW - EIE) / ((EIW + EIE) / 2)) * 100$

A Postgresql Pgsq script helped to compute these differences simultaneously on several datasets. Other formulas could be easily implemented, if needed.

## Thresholds

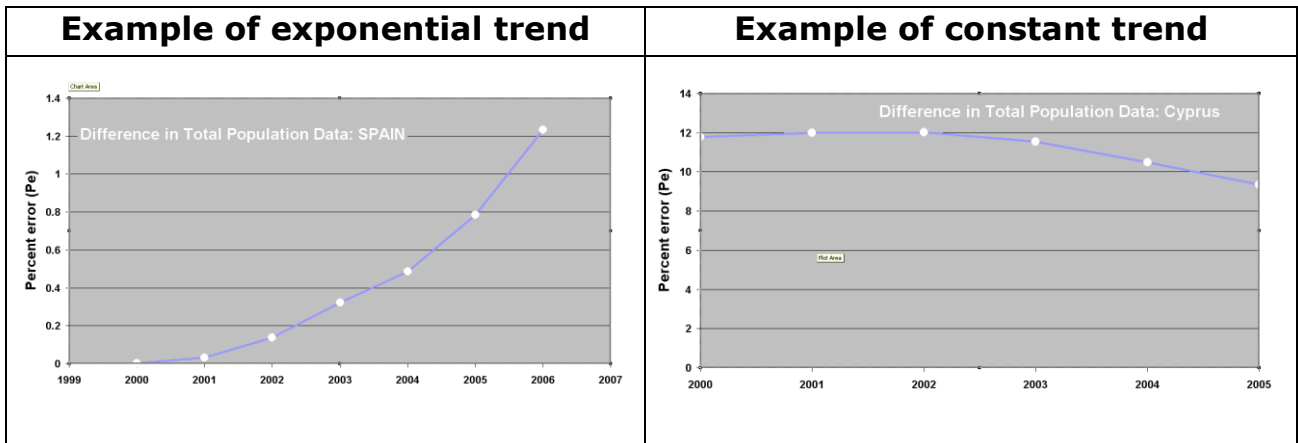
Once the acceptable maximum difference is set at Pe (in absolute value) = 1 (this threshold at 1 can be debated), we can approach the problem under different perspectives: by country, indicator, group of indicators, year, etc.

## Typologies of differences

The idea is to subdivide the Pe in several typologies in order to better characterise the analyses of difference. We introduce three concepts to illustrate the typologies:

- **magnitude** is the numerical value of the difference. It is generally referred to the Percent error (Pe). Translated into a qualitative scale, a Pe from 1% to 3 % is considered as "moderate";
- **range** is the difference between the maximum and minimum magnitude values max in a time series.
- **trend** is the measure of the change of magnitudes over the considered years covered by the indicator. It can be qualified as random, constant, linear, exponential, composite, etc.



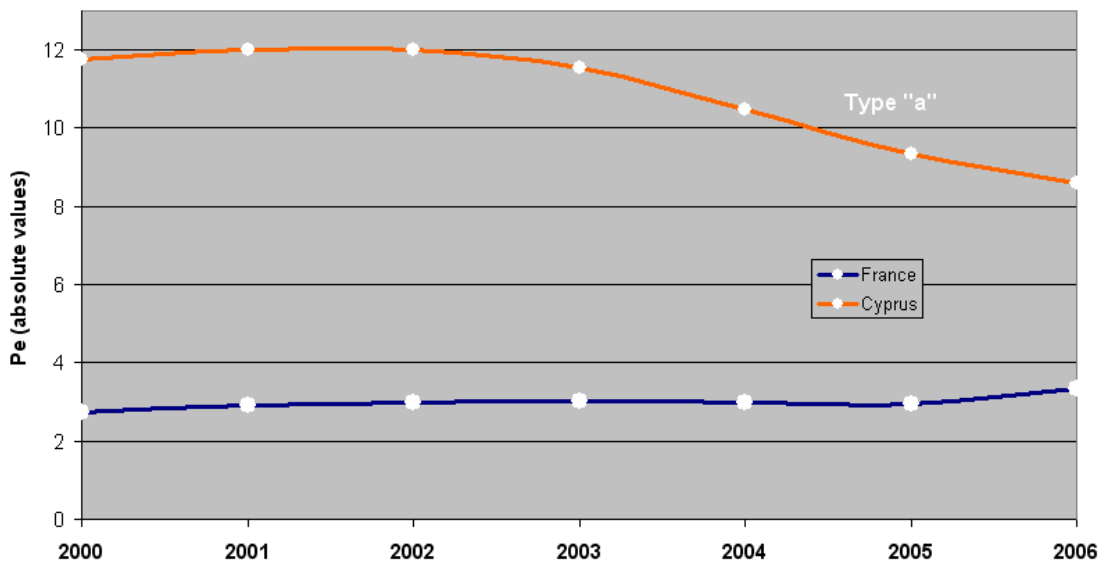


**Figure 9** : Different types of trends

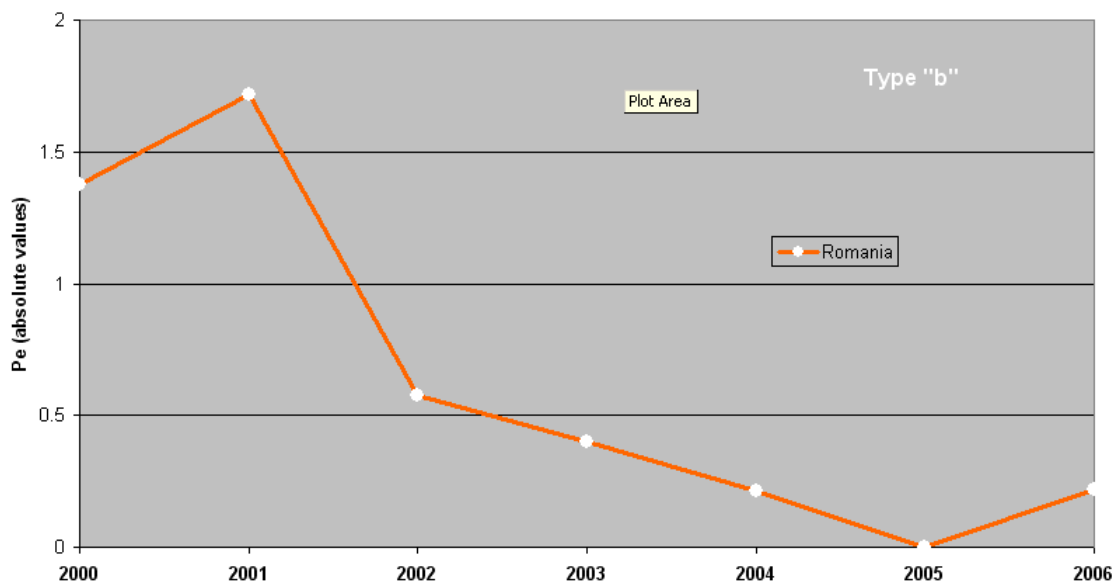
The typologies of differences include:

- type "a" : magnitude moderate to medium, trend constant across all the period of collect, probably due to differences in country definitions (eg. population for Cyprus and France);
- type "b" : magnitude moderate, variable trends, some years without errors. Situation that can have several causes (mixed sources data, interpolations/extrapolations from EIW or/and EIE);
- type "c" : magnitude moderate to high, trend constant to slightly random across all the period of collect due to differences in the indicator definition and to data acquisition methods (measure or calculations).

Example of Type "a": in this case the origin of the discrepancies is the different definition of countries



Example of type "b": something happened in 2000-2001 (census differences? interpolations of values ?)



**Figure 10** : Various types of differences

## 2.2 Results

We analyzed stocks data from the last World Population Prospect (WPP08) including male, female, and both sexes population by age groups.

Although, the majority of Eurostat data on demography is provided by the UN Population Division, some differences between the two sets of data exist, especially for population by sex and age groups.

### Total population data

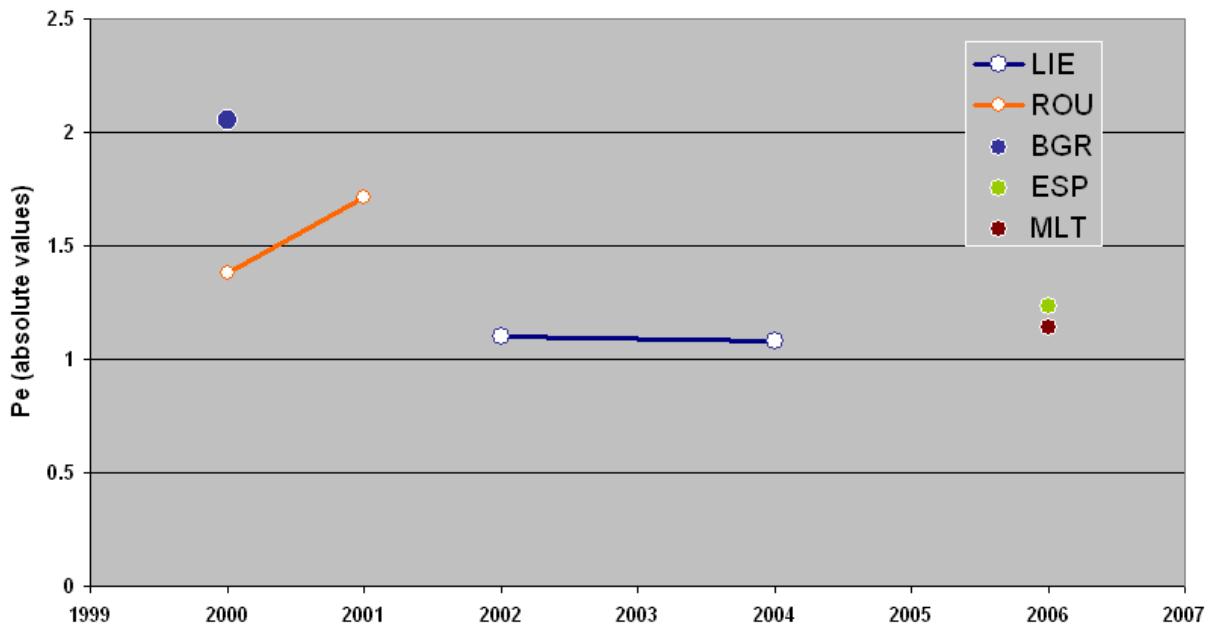
Comparisons from "total population both sexes" indicator give very moderate values of  $P_e$ , i.e. below the threshold of 1 for almost all countries.

Cyprus and France show a distinctive difference of type "a" caused by the different definition of the country: data for Cyprus refer only to the areas of Cyprus controlled by the Government of the Republic of Cyprus in the Eurostat database and France includes the overseas departments (DOM).

Apart from France and Cyprus, only Liechtenstein, Romania, Bulgaria, Spain and Malta display values above 1 (Figure 11).

Data for Spain show an exponentially increasing  $P_e$  through time. This could be caused by secondary readjustment of values.

### Countries with Pe values exceeding 1 (indicator: total population both sexes)

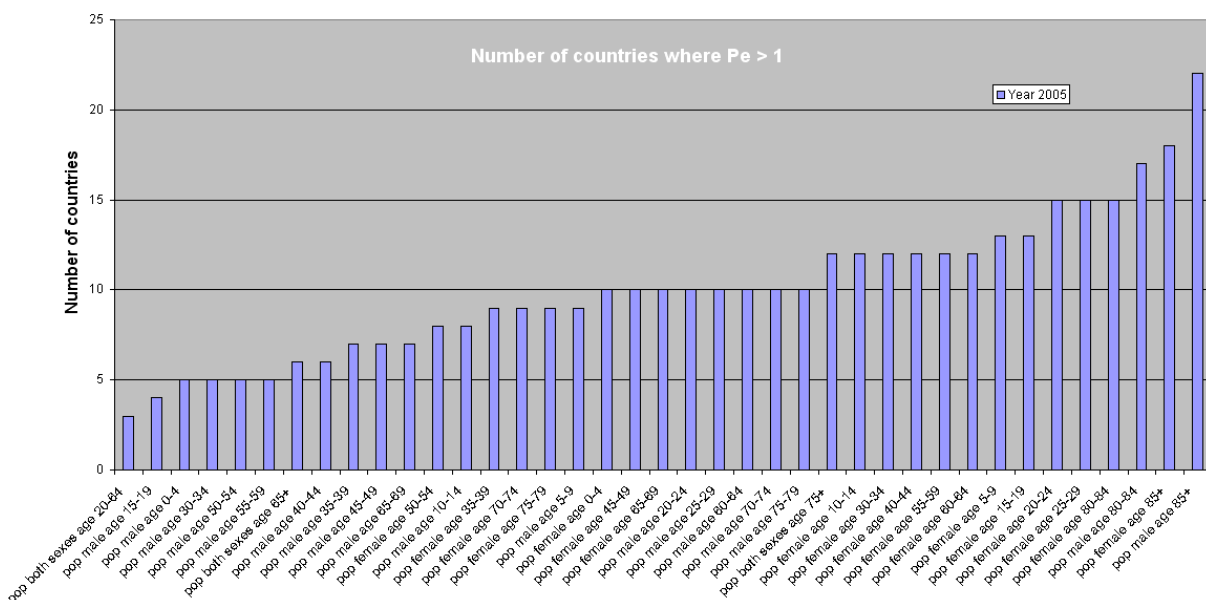


**Figure 11.** Identification of countries where significant differences has been identified

### Population by sex and age groups

This group of indicators shows significant differences between EIW and EIE datasets. Figure 12 shows, per age/sex class, the number of countries where  $Pe > 1$ .

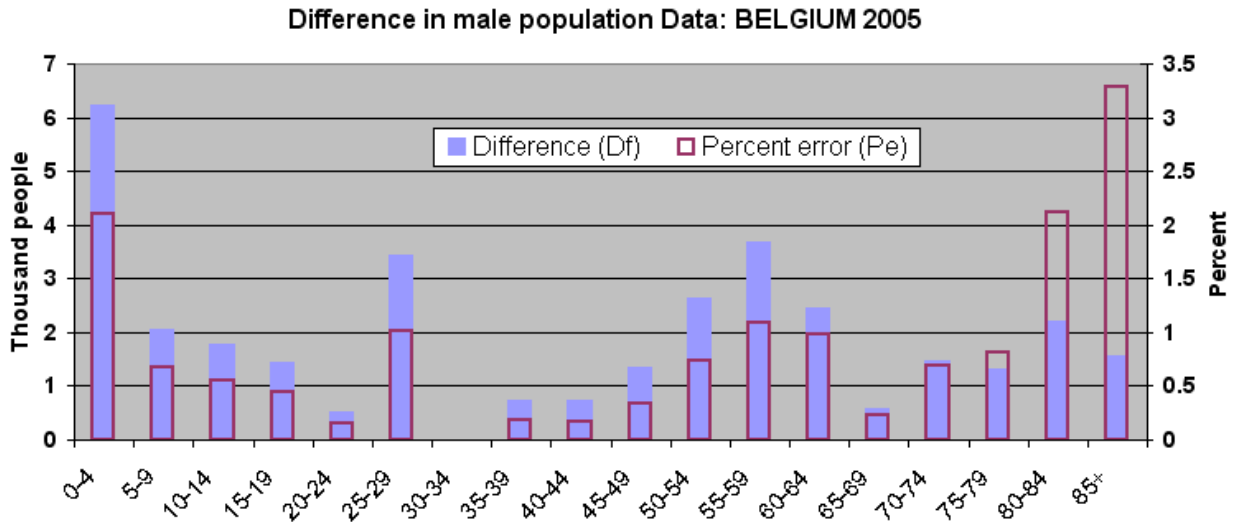
### Number of countries where Pe is greater than 1 (year 2005)



**Figure 12 :** Identification of age-classes where significant differences has been identified

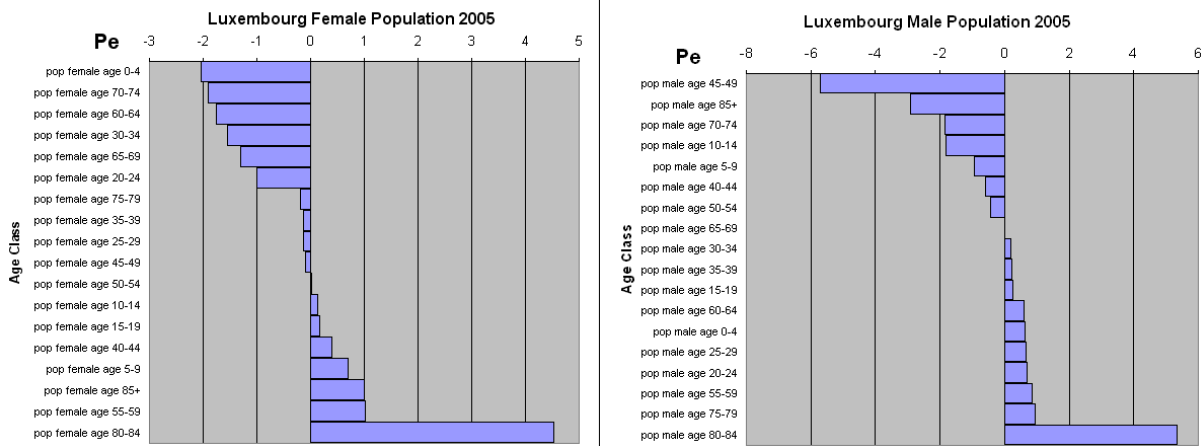
The following tables illustrate some of these inconsistencies per country and per group of variables.

### Absolute and relative differences per age groups fo Belgium (year 2005)



**Figure 13.** Identification of age-classes where significant differences has been identified in Belgium

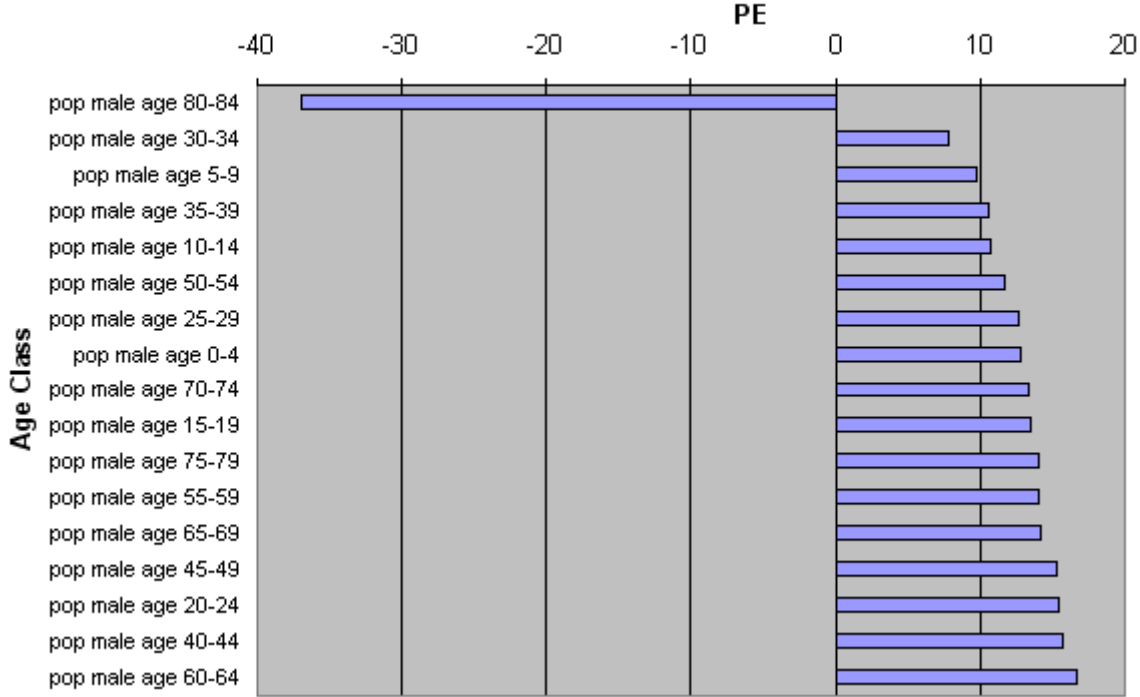
### Differences (Pe) per sex and age groups for Luxembourg (year 2005)



**Figure 14 :** Identification of age-classes where significant differences has been identified in Luxembourg

**Something happened for data on male population over 80 in England...**

### Great Britain Male Population 2005



**Figure 15** : Identification of age-classes where significant differences has been identified in England

Among the 1420 records analysed, all countries show at least one value (but most of time, several values) with a Pe greater than 1. 85 records have a Pe bigger than 10. There are no apparent relationships between errors in different age or sex classes.

### 3 Conclusion

The main achievements of the study are the following:

- a set of criteria for selecting which countries/territories should be included in the "World Dictionary of units";
- a compilation of countries/territories units from four international data sources and a synthetic list of 251 entities based on the selection criteria;
- a methodology for the systematic comparison of values between various data sources, the "Gap tracker" tool;
- the ESPON 2013 World Database version 2.0 (global statistical data + codes of spatial units + links to ESPON DB geometries) itself in MS-Access format, including metadata in ISO 19115:2003/19139.

These data (indicators and list of countries) and methodologies allow for a robust integration of world data into the ESPON 2013 Database.

Further developments should include an approach for taking into account future changes in country definitions, as well as for the inclusion of updated/newer data.

## Annex 1. List of EIW indicators

1 - Data from WPP 2008 population stocks table “wpp\_data”

indicator	temporal extent	indicator	temporal extent
pop female age 80+	1950-2050	pop female age 25+	1950, 1955, ..., 2050
pop male age 80+	1950-2050	pop female age 25-69	1950, 1955, ..., 2050
pop female age 75-79	1950-2050	pop female age 50+	1950, 1955, ..., 2050
pop male age 75-79	1950-2050	pop female age 5-14	1950, 1955, ..., 2050
pop male age 70-74	1950-2050	pop female age 60+	1950, 1955, ..., 2050
pop female age 70-74	1950-2050	pop female age 6-11	1950, 1955, ..., 2050
pop male age 65-69	1950-2050	pop female age 65+	1950, 1955, ..., 2050
pop female age 65-69	1950-2050	pop female age 70+	1950, 1955, ..., 2050
pop female age 60-64	1950-2050	pop female age 75+	1950, 1955, ..., 2050
pop male age 60-64	1950-2050	pop female age 80-84	1950, 1955, ..., 2050
pop female age 55-59	1950-2050	pop female age 85+	1950, 1955, ..., 2050
pop male age 55-59	1950-2050	pop female age 85-89	1950, 1955, ..., 2050
pop female age 50-54	1950-2050	pop female age 90+	1950, 1955, ..., 2050
pop male age 50-54	1950-2050	pop female age 90-94	1950, 1955, ..., 2050
pop female age 45-49	1950-2050	pop female age 95-99	1950, 1955, ..., 2050
pop male age 45-49	1950-2050	pop male age 0-14	1950, 1955, ..., 2050
pop female age 40-44	1950-2050	pop male age 0-17	1950, 1955, ..., 2050
pop male age 40-44	1950-2050	pop male age 0-19	1950, 1955, ..., 2050
pop female age 35-39	1950-2050	pop male age 0-24	1950, 1955, ..., 2050
pop male age 35-39	1950-2050	pop male age 100+	1950, 1955, ..., 2050
pop female age 30-34	1950-2050	pop male age 12-14	1950, 1955, ..., 2050
pop male age 30-34	1950-2050	pop male age 15+	1950, 1955, ..., 2050
pop female age 25-29	1950-2050	pop male age 15-17	1950, 1955, ..., 2050
pop male age 25-29	1950-2050	pop male age 15-24	1950, 1955, ..., 2050
pop female age 20-24	1950-2050	pop male age 15-49	1950, 1955, ..., 2050
pop male age 20-24	1950-2050	pop male age 15-59	1950, 1955, ..., 2050
pop female age 15-19	1950-2050	pop male age 15-64	1950, 1955, ..., 2050
pop male age 15-19	1950-2050	pop male age 18+	1950, 1955, ..., 2050
pop female age 10-14	1950-2050	pop male age 18-23	1950, 1955, ..., 2050
pop male age 10-14	1950-2050	pop male age 20+	1950, 1955, ..., 2050
pop female age 5-9	1950-2050	pop male age 20-64	1950, 1955, ..., 2050
pop male age 5-9	1950-2050	pop male age 20-69	1950, 1955, ..., 2050
pop female age 0-4	1950-2050	pop male age 25+	1950, 1955, ..., 2050
pop male age 0-4	1950-2050	pop male age 25-69	1950, 1955, ..., 2050
pop female age all	1950-2050	pop male age 50+	1950, 1955, ..., 2050
pop male age all	1950-2050	pop male age 5-14	1950, 1955, ..., 2050
pop both sexes age all	1950-2050	pop male age 60+	1950, 1955, ..., 2050
pop both sexes age 0-14	1950, 1955, ..., 2050	pop male age 6-11	1950, 1955, ..., 2050
pop both sexes age 0-17	1950, 1955, ..., 2050	pop male age 65+	1950, 1955, ..., 2050
pop both sexes age 0-19	1950, 1955, ..., 2050	pop male age 70+	1950, 1955, ..., 2050
pop both sexes age 0-24	1950, 1955, ..., 2050	pop male age 75+	1950, 1955, ..., 2050
pop both sexes age 15+	1950, 1955, ..., 2050	pop male age 80-84	1950, 1955, ..., 2050

pop both sexes age 15-17	1950, 1955, ..., 2050
pop both sexes age 15-24	1950, 1955, ..., 2050
pop both sexes age 15-49	1950, 1955, ..., 2050
pop both sexes age 15-59	1950, 1955, ..., 2050
pop both sexes age 15-64	1950, 1955, ..., 2050
pop both sexes age 18+	1950, 1955, ..., 2050
pop both sexes age 18-23	1950, 1955, ..., 2050
pop both sexes age 20+	1950, 1955, ..., 2050
pop both sexes age 20-64	1950, 1955, ..., 2050
pop both sexes age 20-69	1950, 1955, ..., 2050
pop both sexes age 25+	1950, 1955, ..., 2050
pop both sexes age 25-69	1950, 1955, ..., 2050
pop both sexes age 50+	1950, 1955, ..., 2050
pop both sexes age 5-14	1950, 1955, ..., 2050
pop both sexes age 60+	1950, 1955, ..., 2050
pop both sexes age 65+	1950, 1955, ..., 2050
pop both sexes age 70+	1950, 1955, ..., 2050
pop both sexes age 75+	1950, 1955, ..., 2050
pop both sexes age 85+	1950, 1955, ..., 2050
pop both sexes age 90+	1950, 1955, ..., 2050
pop female age 0-14	1950, 1955, ..., 2050
pop female age 0-17	1950, 1955, ..., 2050
pop female age 0-19	1950, 1955, ..., 2050
pop female age 0-24	1950, 1955, ..., 2050
pop female age 100+	1950, 1955, ..., 2050
pop female age 12-14	1950, 1955, ..., 2050
pop female age 15+	1950, 1955, ..., 2050
pop female age 15-17	1950, 1955, ..., 2050
pop female age 15-24	1950, 1955, ..., 2050
pop female age 15-49	1950, 1955, ..., 2050
pop female age 15-59	1950, 1955, ..., 2050
pop female age 15-64	1950, 1955, ..., 2050
pop female age 18+	1950, 1955, ..., 2050
pop female age 18-23	1950, 1955, ..., 2050
pop female age 20+	1950, 1955, ..., 2050
pop female age 20-64	1950, 1955, ..., 2050
pop female age 20-69	1950, 1955, ..., 2050

pop male age 85+	1950, 1955, ..., 2050
pop male age 85-89	1950, 1955, ..., 2050
pop male age 90+	1950, 1955, ..., 2050
pop male age 90-94	1950, 1955, ..., 2050
pop male age 95-99	1950, 1955, ..., 2050
pop median age	1950, 1955, ..., 2050
pop sex ratio age 0-14	1950, 1955, ..., 2050
pop sex ratio age 0-17	1950, 1955, ..., 2050
pop sex ratio age 0-19	1950, 1955, ..., 2050
pop sex ratio age 0-24	1950, 1955, ..., 2050
pop sex ratio age 0-4	1950, 1955, ..., 2050
pop sex ratio age 12-14	1950, 1955, ..., 2050
pop sex ratio age 15+	1950, 1955, ..., 2050
pop sex ratio age 15-17	1950, 1955, ..., 2050
pop sex ratio age 15-24	1950, 1955, ..., 2050
pop sex ratio age 15-49	1950, 1955, ..., 2050
pop sex ratio age 15-59	1950, 1955, ..., 2050
pop sex ratio age 15-64	1950, 1955, ..., 2050
pop sex ratio age 18+	1950, 1955, ..., 2050
pop sex ratio age 18-23	1950, 1955, ..., 2050
pop sex ratio age 20+	1950, 1955, ..., 2050
pop sex ratio age 20-64	1950, 1955, ..., 2050
pop sex ratio age 20-69	1950, 1955, ..., 2050
pop sex ratio age 25+	1950, 1955, ..., 2050
pop sex ratio age 25-69	1950, 1955, ..., 2050
pop sex ratio age 50+	1950, 1955, ..., 2050
pop sex ratio age 5-14	1950, 1955, ..., 2050
pop sex ratio age 60+	1950, 1955, ..., 2050
pop sex ratio age 6-11	1950, 1955, ..., 2050
pop sex ratio age 65+	1950, 1955, ..., 2050
pop sex ratio age 70+	1950, 1955, ..., 2050
pop sex ratio age 75+	1950, 1955, ..., 2050
pop sex ratio age 80+	1950, 1955, ..., 2050
pop sex ratio age 85+	1950, 1955, ..., 2050
pop sex ratio age 90+	1950, 1955, ..., 2050
pop sex ratio age all	1950, 1955, ..., 2050

## 2 - Data from CO2 Emissions (UNFCCC 2009 and CDIAC 2008) table "co2\_data"

indicator	temporal extent
Emissions of CO2 - (National Reports, UNFCCC), Excluding Land Use, Land-Use Change and Forestry	2000-2008
Emissions of CO2 - (National Reports, UNFCCC), Including Land Use, Land-Use Change and Forestry	2000-2008
Emissions of CO2 (CDIAC) from cement production	2000-2008
Emissions of CO2 (CDIAC) from gas flaring	2000-2008



Emissions of CO2 (CDIAC) from gas fuel consumption	2000-2008
Emissions of CO2 (CDIAC) from liquid fuels consumption	2000-2008
Emissions of CO2 (CDIAC) from solid fuels consumption	2000-2008
Emissions of CO2 (CDIAC) from total fuels consumption	2000-2008

### 3 - Data from WDI 2010

indicator	temporal extent
GDP (constant 2000 US\$)	2000-2008
GDP growth (annual %)	2000-2008
GDP, PPP (constant 2005 international \$)	2000-2008
GNI, PPP (current international \$)	2000-2008
Present value of external debt (current US\$)	2000-2008
Unemployment, total (% of total labor force)	2000-2008

## Annex 2.1. LIST OF GEOGRAPHICAL UNITS FROM CHELEM (PARTITION IN 96 ENTITIES)

CHELEM name	Note
United States	United States of America (including Puerto Rico and US Virgin Islands in TRADE, US Samoa, Guam, US Virgin Islands and Puerto Rico in BOP)
Canada	Canada
France	France, Monaco (including French overseas departments in TRADE, and French overseas departments and territories in BOP)
BLEU	Belgium, Luxembourg
Germany	Germany (including East Germany since 1991)
Italy	Italy (including San Marino and the Holy See)
Netherlands	Netherlands
United Kingdom	United Kingdom of Great Britain and Northern Ireland
Ireland	Ireland
Denmark	Denmark
Finland	Finland
Norway	Norway (including Svalbard and Jan Mayen)
Sweden	Sweden
Iceland	Iceland (and Faroe Islands in TRADE)
Austria	Austria
Switzerland	Switzerland (including Liechtenstein in TRADE)
Spain	Spain
Greece	Greece
Portugal	Portugal
Turkey	Turkey
Israel	Israel
Serbia and Montenegro	Federal Republic of Yugoslavia (including Macedonia in TRADE in 1992)
Bosnia and Herzegovina	Bosnia and Herzegovina
Croatia	Croatia
Macedonia, Republic of	Macedonia, Republic of
Slovenia	Slovenia
Others in south Europe	Andorra (in TRADE only), Cyprus, Gibraltar, Malta, West Bank and Gaza (in GDP and BOP only)
Japan	Japan
Australia	Australia
New Zealand	New Zealand
Southafrican Union	Botswana, Lesotho, Namibia, South Africa, Swaziland
Venezuela	Venezuela
Ecuador	Ecuador
Mexico	Mexico
Brazil	Brazil
Argentina	Argentina
Chile	Chile
Colombia	Colombia
Peru	Peru
Bolivia	Bolivia
Paraguay	Paraguay
Uruguay	Uruguay

CHELEM name	Note
Others in America	Anguilla (in BOP and TRADE), Antigua and Barbuda, Aruba, Bahamas, Barbados, Belize, Bermuda, Costa Rica, Cuba, Dominica, Dominican Republic, El Salvador, French Guiana (in GDP only), Grenada, Guadeloupe (in GDP only), Guatemala, Guyana, Haiti, Honduras, Jamaica, Martinique (in GDP only), Montserrat (in BOP and TRADE), Netherland Antilles, Nicaragua, Panama, Puerto Rico (in GDP only), Saint Kitts and Nevis, Saint Lucia, Saint Vincent and the Grenadines, Suriname, Trinidad and Tobago, US Virgin Islands (in GDP only), and all others in America (in TRADE only)
Algeria	Algeria
Morocco	Morocco (including Western Sahara in BOP)
Tunisia	Tunisia
Egypt	Egypt
Libyan Arab Jamahiriya	Libyan Arab Jamahiriya
Saudi Arabia	Saudi Arabia
Gulf nes	Bahrein, Iran, Iraq, Kuwait, Oman, Qatar, United Arab Emirates
Middle East, no OPEC	Jordan, Lebanon, Syria, Yemen
Nigeria	Nigeria
Gabon	Gabon
Cameroon	Cameroon
Cote d'Ivoire	Cote d'Ivoire
Kenya	Kenya
Africa (others)	Congo, Ghana, Mauritius, Reunion (in GDP only), Seychelles, Western Sahara (in GDP and TRADE), Zimbabwe, and all others in Africa (in TRADE only)
African LDCs	Angola, Benin, Burkina Faso, Burundi, Cameroon, Cape Verde, Central African Republic, Chad, Comoros, Cote d'Ivoire, Democratic Republic of Congo (formerly Zaire), Djibouti, Equatorial Guinea, Eritrea, Ethiopia, Gambia, Guinea, Guinea-Bissau, Kenya, Liberia, Madagascar, Malawi, Mali, Mauritania, Mozambique, Niger, Rwanda, Sao Tome and Principe, Senegal, Sierra Leone, Somalia, Sudan, Tanzania, Togo, Uganda, Zambia
Indonesia	Indonesia
India	India
South Korea	Republic of Korea
Hong Kong	Hong Kong Special Administrative Region of China
Singapore	Singapore
Taiwan	Taiwan
Malaysia	Malaysia
Philippines	Philippines
Thailand	Thailand
Pakistan	Pakistan
Brunei Darussalam	Brunei Darussalam
Bangladesh	Bangladesh
Sri Lanka	Sri Lanka
East Asia nes, others	Fiji, French Polynesia (in GDP and TRADE), Guam (in GDP and TRADE), Macao, Mongolia, New Caledonia (in GDP and TRADE), North Korea, Pacific Islands (in GDP and TRADE), Papua New Guinea, Tonga, US Samoa (in GDP and TRADE), Vanuatu, Western Samoa, and all others in Asia and Oceania (in TRADE only)
East Asian LDCs	Afghanistan, Bhutan, Kiribati, Maldives, Myanmar, Nepal, Solomon Islands, Vanuatu, Western Samoa
Russian Federation	Russian Federation
Ukraine	Ukraine
Belarus	Belarus
Kazakhstan	Kazakhstan
Kyrgyzstan	Kyrgyzstan
Caucasus	Armenia, Azerbaijan, Georgia
Other CIS	Moldova, Tajikistan, Turkmenistan, Uzbekistan
Estonia	Estonia

<b>CHELEM name</b>	<b>Note</b>
Latvia	Latvia
Lithuania	Lithuania
Bulgaria	Bulgaria
Czech Republic	Czech Republic
Slovakia	Slovakia
Hungary	Hungary
Poland	Poland
Romania	Romania
Former German Democratic Rep.	Former German Democratic Republic (up to 1990)
Albania	Albania
China, People's Rep.	The People's Republic of China: Mainland
Viet Nam	Viet Nam
Cambodia, Lao PDR	Cambodia, Lao PDR
Miscellaneous	Not elsewhere specified (international organizations in BOP)
World	Total-of-the-33-Areas

## Annex 2.2. LIST OF GEOGRAPHICAL UNITS (168 UNITS) FROM ESPON 2006 PROGRAM (EUROPE IN THE WORLD)

WUTS5 name	Note
Afghanistan	
Angola	
Albania	
United Arab Emirates	
Argentina	
Armenia	
Australia	
Austria	
Azerbaijan	
Burundi	
Belgium	
Benin	
Burkina Faso	
Bangladesh	
Bulgaria	
Bahrain	
Bahamas	
Bosnia and Herzegovina	
Belarus	
Belize	
Bolivia	
Brazil	
Bhutan	
Botswana	
Central African Republic	
Canada	
Switzerland	
Chile	
China	(China main land + Macao + Hong-Kong)
Côte d'Ivoire	
Cameroon	
Congo, Dem. Rep. of the	
Congo	
Colombia	
Costa Rica	
Cuba	
Cyprus	
Czech Republic	
Germany	
Djibouti	
Denmark	

WUTS5 name	Note
Dominican Republic	
Algeria	
Ecuador	
Egypt	
Eritrea	
West Sahara	
Spain	
Estonia	
Ethiopia	
Finland	
Fiji	
France	France (Mainland) + Guadeloupe + Martinique + Guyane + Réunion
Gabon	
United Kingdom	
Georgia	
Ghana	
Guinea	
Gambia	
Guinea-Bissau	
Equatorial Guinea	
Greece	
Greenland	
Guatemala	
Guyana	
Honduras	
Croatia	
Haiti	
Hungary	
Indonesia	
India	
Ireland	
Iran, Islamic Rep. of	
Iraq	
Iceland	
Israel	Israel (without Occupied Palestinian Territories)
Italy	
Jamaica	
Jordan	
Japan	
Kazakhstan	
Kenya	
Kyrgyzstan	
Cambodia	
Korea, Rep. of	
Kuwait	

WUTS5 name	Note
Lao People's Dem. Rep.	
Lebanon	
Liberia	
Libyan Arab Jamahiriya	
Sri Lanka	
Lesotho	
Lithuania	
Luxembourg	
Latvia	
Morocco	Morocco (without Western Sahara)
Moldova, Rep. of	
Madagascar	
Mexico	
Macedonia, TFYR	
Mali	
Malta	
Myanmar	
Mongolia	
Mozambique	
Mauritania	
Mauritius	
Malawi	
Malaysia	
Namibia	
Niger	
Nigeria	
Nicaragua	
Netherlands	
Norway	
Nepal	
New Zealand	
Oman	
Pakistan	
Panama	
Peru	
Philippines	
Papua New Guinea	
Poland	
Puerto Rico	
North Korea	
Portugal	
Paraguay	
Qatar	
Romania	
Russian Federation	

WUTS5 name	Note
Rwanda	
Saudi Arabia	
Serbia/Montenegro	
Sudan	
Senegal	
Singapore	
Sierra Leone	
El Salvador	
Somalia	
Suriname	
Slovakia	
Slovenia	
Sweden	
Swaziland	
Syrian Arab Republic	
Chad	
Togo	
Thailand	
Tajikistan	
Turkmenistan	
Trinidad and Tobago	
Tunisia	
Turkey	
Taiwan	
Tanzania, U. Rep. of	
Uganda	
Ukraine	
Uruguay	
United States	
Uzbekistan	
Venezuela	
Viet Nam	
Occupied Palestinian Territories	
Yemen	
South Africa	
Zambia	
Zimbabwe	



## Annex 2.3. LIST OF GEOGRAPHICAL UNITS FROM GEO

United Nations country name	Notes
Aruba	
Afghanistan	
Angola	
Anguilla	
Albania	
Andorra	
Netherlands Antilles	
United Arab Emirates	
Argentina	
Armenia	
American Samoa	
Antarctic	
Antigua and Barbuda	
Australia	
Austria	
Azerbaijan	
Burundi	
Belgium	
Benin	
Burkina Faso	
Bangladesh	
Bulgaria	
Bahrain	
Bahamas	
Bosnia and Herzegovina	
Belarus	
Belize	
Bermuda	
Bolivia	
Brazil	
Barbados	
Brunei Darussalam	
Bhutan	
Botswana	
Central African Republic	
Canada	
Cocos (Keeling) Islands	
Switzerland	
Chile	
China	Including Macau, Hong Kong and Taiwan

United Nations country name	Notes
Cote d'Ivoire	
Cameroon	
Democratic Republic of the Congo	
Congo	
Cook Islands	
Colombia	
Comoros	
Cape Verde	
Costa Rica	
Cuba	
Christmas Island	
Cayman Islands	
Cyprus	
Czech Republic	
Germany	
Djibouti	
Dominica	
Denmark	
Dominican Republic	
Algeria	
Ecuador	
Egypt	
Eritrea	
Western Sahara	
Spain	
Estonia	
Ethiopia	
Finland	
Fiji	
Falkland Islands (Malvinas)	
France	
Faroe Islands	
Micronesia (Federated States of)	
Gabon	
United Kingdom of Great Britain and Northern Ireland	
Georgia	
Guernsey	
Ghana	
Gibraltar	
Guinea	
Guadeloupe	
Gambia	
Guinea-Bissau	

United Nations country name	Notes
Equatorial Guinea	
Greece	
Grenada	
Greenland	
Guatemala	
French Guiana	
Guam	
Guyana	
Honduras	
Croatia	
Haiti	
Hungary	
Indonesia	
Isle of Man	
India	
Ireland	
Iran (Islamic Republic of)	
Iraq	
Iceland	
Israel	
Italy	
Jamaica	
Jersey	
Jordan	
Japan	
Johnston Atoll	
Kazakhstan	
Kenya	
Kyrgyzstan	
Cambodia	
Kiribati	
Saint Kitts and Nevis	
Republic of Korea	
Kuwait	
Lao People's Democratic Republic	
Lebanon	
Liberia	
Libyan Arab Jamahiriya	
Saint Lucia	
Liechtenstein	
Sri Lanka	
Lesotho	
Lithuania	

United Nations country name	Notes
Luxembourg	
Latvia	
Morocco	
Monaco	
Moldova, Republic of	
Madagascar	
Maldives	
Mexico	
Marshall Islands	
Midway Islands	
The former Yugoslav Republic of Macedonia	
Mali	
Malta	
Myanmar	
Montenegro	
Mongolia	
Northern Mariana Islands	
Mozambique	
Mauritania	
Montserrat	
Martinique	
Mauritius	
Malawi	
Malaysia	
Mayotte	
Namibia	
New Caledonia	
Niger	
Norfolk Island	
Nigeria	
Nicaragua	
Niue	
Netherlands	
Norway	
Nepal	
Nauru	
New Zealand	
Oman	
Pakistan	
Panama	
Pitcairn Island	
Peru	
Philippines	

United Nations country name	Notes
Palau	
Papua New Guinea	
Poland	
Puerto Rico	
Democratic People's Republic of Korea	
Portugal	
Paraguay	
Occupied Palestinian Territory	Including West Bank and Gaza
French Polynesia	
Qatar	
Reunion	
Romania	
Russian Federation	
Rwanda	
Saudi Arabia	
Sudan	
Senegal	
Singapore	
Saint Helena	
Svalbard and Jan Mayen Islands	
Solomon Islands	
Sierra Leone	
El Salvador	
San Marino	
Somalia	
Saint Pierre and Miquelon	
Serbia	Including Kosovo
Sao Tome and Principe	
Suriname	
Slovakia	
Slovenia	
Sweden	
Swaziland	
Seychelles	
Syrian Arab Republic	
Turks and Caicos Islands	
Chad	
Togo	
Thailand	
Tajikistan	
Tokelau	
Turkmenistan	
Timor-Leste	

United Nations country name	Notes
Tonga	
Trinidad and Tobago	
Tunisia	
Turkey	
Tuvalu	
United Republic of Tanzania	
Uganda	
Ukraine	
Uruguay	
United States of America	
Uzbekistan	
Holy See	
Saint Vincent and the Grenadines	
Venezuela	
British Virgin Islands	
United States Virgin Islands	
Viet Nam	
Vanuatu	
Wake Island	
Wallis and Futuna	
Samoa	
Yemen	
South Africa	
Zambia	
Zimbabwe	

## Annex 2.4. LIST OF GEOGRAPHICAL UNITS FROM WDI

WDI country name	Notes
Afghanistan	
Albania	
Algeria	
American Samoa	
Andorra	
Angola	
Antigua and Barbuda	
Argentina	
Armenia	
Aruba	
Australia	
Austria	
Azerbaijan	
Bahamas, The	
Bahrain	
Bangladesh	
Barbados	
Belarus	
Belgium	
Belize	
Benin	
Bermuda	
Bhutan	
Bolivia	
Bosnia and Herzegovina	
Botswana	
Brazil	
Brunei Darussalam	
Bulgaria	
Burkina Faso	
Burundi	
Cambodia	
Cameroon	
Canada	
Cape Verde	
Cayman Islands	
Central African Republic	
Chad	
Channel Islands	
Chile	
China	Unless otherwise noted, data for China do not include data for Hong Kong, Macau, or Taiwan
Colombia	
Comoros	
Congo, Dem. Rep.	
Congo, Rep.	

WDI country name	Notes
Costa Rica	
Cote d'Ivoire	
Croatia	
Cuba	
Cyprus	Data related to GDP, exclude Turkish-controlled area
Czech Republic	
Denmark	
Djibouti	
Dominica	
Dominican Republic	
Ecuador	
Egypt, Arab Rep.	
El Salvador	
Equatorial Guinea	
Eritrea	
Estonia	
Ethiopia	
Faeroe Islands	
Fiji	
Finland	
France	Data related to GDP, include French Guiana, Guadelupe, Martinique and Réunion
French Guiana	
French Polynesia	
Gabon	
Gambia, The	
Georgia	
Germany	
Ghana	
Greece	
Greenland	
Grenada	
Guadeloupe	
Guam	
Guatemala	
Guinea	
Guinea-Bissau	
Guyana	
Haiti	
Honduras	
Hong Kong, China	
Hungary	
Iceland	
India	
Indonesia	
Iran, Islamic Rep.	
Iraq	
Ireland	
Isle of Man	



WDI country name	Notes
Israel	
Italy	
Jamaica	
Japan	
Jordan	
Kazakhstan	
Kenya	
Kiribati	
Korea, Dem. Rep.	
Korea, Rep.	
Kuwait	
Kyrgyz Republic	
Lao PDR	
Latvia	
Lebanon	
Lesotho	
Liberia	
Libya	
Liechtenstein	
Lithuania	
Luxembourg	
Macao, China	
Macedonia, FYR	
Madagascar	
Malawi	
Malaysia	
Maldives	
Mali	
Malta	
Marshall Islands	
Martinique	
Mauritania	
Mauritius	
Mayotte	
Mexico	
Micronesia, Fed. Sts.	
Moldova	Data related to GDP, exclude Transnistria
Monaco	
Mongolia	
Montenegro	
Morocco	
Mozambique	
Myanmar	
Namibia	
Nauru	
Nepal	
Netherlands	
Netherlands Antilles	

WDI country name	Notes
New Caledonia	
New Zealand	
Nicaragua	
Niger	
Nigeria	
Northern Mariana Islands	
Norway	
Oman	
Pakistan	
Palau	
Panama	
Papua New Guinea	
Paraguay	
Peru	
Philippines	
Poland	
Portugal	
Puerto Rico	
Qatar	
Reunion	
Romania	
Russian Federation	
Rwanda	
Samoa	
San Marino	
Sao Tome and Principe	
Saudi Arabia	
Senegal	
Serbia	Where available, data from Serbia and Montenegro are shown separately. However some indicators for Serbia prior to 2006 include data for Montenegro
Seychelles	
Sierra Leone	
Singapore	
Slovak Republic	
Slovenia	
Solomon Islands	
Somalia	
South Africa	
Spain	
Sri Lanka	
St. Kitts and Nevis	
St. Lucia	
St. Vincent and the Grenadines	
Sudan	
Suriname	
Swaziland	
Sweden	
Switzerland	
Syrian Arab Republic	

WDI country name	Notes
Tajikistan	
Tanzania	Data related to GDP, cover mainland Tanzania only
Thailand	
Timor-Leste	
Togo	
Tonga	
Trinidad and Tobago	
Tunisia	
Turkey	
Turkmenistan	
Tuvalu	
Uganda	
Ukraine	
United Arab Emirates	
United Kingdom	
United States	
Uruguay	
Uzbekistan	
Vanuatu	
Venezuela, RB	
Vietnam	
Virgin Islands (U.S.)	
West Bank and Gaza	
Yemen, Rep.	
Zambia	
Zimbabwe	

## Annex 2.5. LIST OF GEOGRAPHICAL UNITS FROM UN (WPP08)

United Nations country name	Notes
Afghanistan	
Aland Islands	
Albania	
Algeria	
American Samoa	
Andorra	
Angola	
Anguilla	
Antigua and Barbuda	
Argentina	
Armenia	
Aruba	
Australia	Including Christmas Island, Cocos (Keeling) Islands, and Norfolk Island.
Austria	
Azerbaijan	
Bahamas	
Bahrain	
Bangladesh	
Barbados	
Belarus	
Belgium	
Belize	
Benin	
Bermuda	
Bhutan	
Bolivia	
Bosnia and Herzegovina	
Botswana	
Brazil	
British Virgin Islands	
Brunei Darussalam	
Bulgaria	
Burkina Faso	
Burundi	
Cambodia	
Cameroon	
Canada	
Cape Verde	
Cayman Islands	
Central African Republic	
Chad	
Channel Islands	Refers to Guernsey, and Jersey.
Chile	

United Nations country name	Notes
China	For statistical purposes, the data for China do not include Hong Kong and Macao, Special Administrative Regions (SAR) of China.
Colombia	
Comoros	
Congo	
Cook Islands	
Costa Rica	
Cote d'Ivoire	
Croatia	
Cuba	
Cyprus	
Czech Republic	
Democratic People's Republic of Korea	
Democratic Republic of the Congo	
Denmark	
Djibouti	
Dominica	
Dominican Republic	
Ecuador	
Egypt	
El Salvador	
Equatorial Guinea	
Eritrea	
Estonia	
Ethiopia	
Faeroe Islands	
Falkland Islands (Malvinas)	
Fiji	
Finland	Including Åland Islands.
France	
French Guiana	
French Polynesia	
Gabon	
Gambia	
Georgia	
Germany	
Ghana	
Gibraltar	
Greece	
Greenland	
Grenada	
Guadeloupe	
Guam	
Guatemala	
Guernsey	
Guinea	
Guinea-Bissau	
Guyana	

United Nations country name	Notes
Haiti	
Holy See	Refers to the Vatican City State.
Honduras	
Hong Kong Special Administrative Region of China	As of 1 July 1997, Hong Kong became a Special Administrative Region (SAR) of China.
Hungary	
Iceland	
India	
Indonesia	
Iran, Islamic Republic of	
Iraq	
Ireland	
Isle of Man	
Israel	
Italy	
Jamaica	
Japan	
Jersey	
Jordan	
Kazakhstan	
Kenya	
Kiribati	
Kuwait	
Kyrgyzstan	
Lao People's Democratic Republic	
Latvia	
Lebanon	
Lesotho	
Liberia	
Libyan Arab Jamahiriya	
Liechtenstein	
Lithuania	
Luxembourg	
Macao Special Administrative Region of China	As of 20 December 1999, Macao became a Special Administrative Region (SAR) of China.
Madagascar	
Malawi	
Malaysia	
Maldives	
Mali	
Malta	
Marshall Islands	
Martinique	
Mauritania	
Mauritius	Including Agalega, Rodrigues, and Saint Brandon.
Mayotte	
Mexico	
Micronesia, Federated States of	
Monaco	

United Nations country name	Notes
Mongolia	
Montenegro	
Montserrat	
Morocco	
Mozambique	
Myanmar	
Namibia	
Nauru	
Nepal	
Netherlands	
Netherlands Antilles	
New Caledonia	
New Zealand	
Nicaragua	
Niger	
Nigeria	
Niue	
Norfolk Island	
Northern Mariana Islands	
Norway	Including Svalbard and Jan Mayen Islands.
Occupied Palestinian Territory	Including West Bank and Gaza
Oman	
Pakistan	
Palau	
Panama	
Papua New Guinea	
Paraguay	
Peru	
Philippines	
Pitcairn	
Poland	
Portugal	
Puerto Rico	
Qatar	
Republic of Korea	
Republic of Moldova	
Réunion	
Romania	
Russian Federation	
Rwanda	
Saint Helena	Including Ascension, and Tristan da Cunha.
Saint Kitts and Nevis	
Saint Lucia	
Saint Pierre and Miquelon	
Saint Vincent and the Grenadines	
Saint-Barthélemy	
Saint-Martin (French part)	
Samoa	

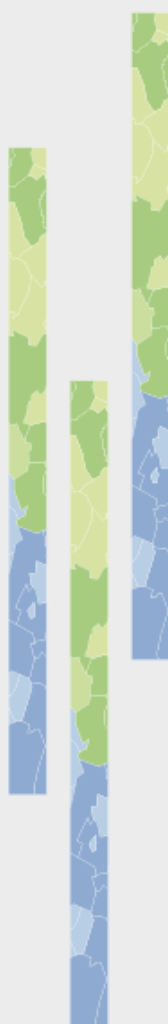
United Nations country name	Notes
San Marino	
Sao Tome and Principe	
Saudi Arabia	
Senegal	
Serbia	
Seychelles	
Sierra Leone	
Singapore	
Slovakia	
Slovenia	
Solomon Islands	
Somalia	
South Africa	
Spain	
Sri Lanka	
Sudan	
Suriname	
Svalbard and Jan Mayen Islands	
Swaziland	
Sweden	
Switzerland	
Syrian Arab Republic	
Tajikistan	
Thailand	
The former Yugoslav Republic of Macedonia	The former Yugoslav Republic of Macedonia.
Timor-Leste	
Togo	
Tokelau	
Tonga	
Trinidad and Tobago	
Tunisia	
Turkey	
Turkmenistan	
Turks and Caicos Islands	
Tuvalu	
Uganda	
Ukraine	
United Arab Emirates	
United Kingdom of Great Britain and Northern Ireland	
United Republic of Tanzania	
United States of America	
United States Virgin Islands	
Uruguay	
Uzbekistan	
Vanuatu	
Venezuela (Bolivarian Republic of)	
Viet Nam	
Wallis and Futuna Islands	



United Nations country name	Notes
Western Sahara	
Yemen	
Zambia	
Zimbabwe	

## Databases references

- **CHELEM Database** (harmonised counts on exchanges and the world economy), built by CEPII (Centre d'Etudes Prospectives et d'Informations Internationales), is known since many years as a precious tool for analysing the global World Economy: <http://www.cepii.fr/francgraph/bdd/chelem.htm>.
- **Geo Data Portal** is the authoritative source for datasets used by UNEP and its partners for the preparation of the Global Environment Outlook (GEO) series of reports and other integrated environmental assessments. Its online database holds more than [500 different variables](#), as national, subregional, regional and global statistics or as geospatial datasets (maps), covering themes like freshwater, population, forests, emissions, climate, disasters, health and GDP. Data can be displayed on-the-fly as [maps](#), [graphs](#), [data tables](#) or downloaded in different formats: <http://geodata.grid.unep.ch/>.
- **United Nations (UN) gateway** to statistical information from databases of the UN and Member States: <http://unstats.un.org/unsd/default.htm>.
- **World Development Indicators Online (WDI)** provides direct access to more than 800 development indicators, with time series for 209 countries and 18 country groups from 1960 to 2008. Data are available at: <http://data.worldbank.org/data-catalog>.



## ANALYSIS OF THE AVAILABILITY AND THE QUALITY OF DATA ON WESTERN BALKANS AND TURKEY

### CONTENT

- **General assessment.** This part discusses the Spatial Administrative Divisions of WB and Turkey and presents the set of data included in the Database as well as the availability and quality of other territorial data for these countries.
- **Assessment per country.** This part presents the assessment of the availability and the quality of data per country of WB and Turkey.
- **Conclusions** on the data availability at NUTS0 to NUTS3 levels and the inclusion of WB and Turkey data in the ESPON Database.

**ESPON 2013 DATABASE**  
**MARCH 2011**



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

50 pages

# LIST OF AUTHORS

Author of the Report and main researcher:

Minas Angelidis, National Technical University of Athens (NTUA)

Contributions:

Gabriella Karka (in parts of the Report), NTUA

Kostas Santimpantakis (in specific parts of data process), NTUA

Epameinondas Tsigkas (in specific parts of data process), NTUA

Vivian Bazoula (in specific parts of data process), NTUA

## Contact

[angelidi@central.ntua.gr](mailto:angelidi@central.ntua.gr)

tel. + 30 210 7721731

## TABLE OF CONTENTS

<b>Introduction and methodological remarks .....</b>	<b>4</b>
<b>1. General assessment .....</b>	<b>7</b>
1.1 The WB and Turkey Spatial Administrative Divisions .....	7
1.2 The data required, the "basic" data included in the ESPON Database and the "additional" data .....	8
<b>2. Assessment per country .....</b>	<b>12</b>
Albania .....	12
Bosnia and Herzegovina.....	15
Croatia .....	19
FYROM .....	22
Serbia .....	25
Montenegro.....	29
Kosovo (Under UN Security Council Resolution 1244) .....	30
Turkey.....	31
<b>3. Conclusions .....</b>	<b>34</b>
Annex-1 Maps .....	36
Annex 2 -Western Balkans and Turkey available territorial data from NSO .....	38
Annex 3 - W. Balkans and Turkey data (2009) from Eurostat / Short presentation	40
Annex 4 – Western Balkans and Turkey data (2009) from Eurostat/ Detailed description .....	44
References - Data sources .....	48

## Abbreviations

CC: Candidate Countries  
FBiH: Federation of Bosnia and Herzegovina  
GDP: Gross Domestic Product  
NUTS: Nomenclature of Territorial Units for Statistics  
PCC: Potential Candidate Countries  
WB: Western Balkans

## List of Tables, Maps and Figures

Table 1.1: NUTS 1, 2, 3 regions in Croatia, FYROM and Turkey .....	7
Table 1.2: "Similar NUTS 1, 2, 3" regions in Albania, Bosnia and Herzegovina, Serbia, Montenegro and Kosovo** .....	7
Table 1.3: Candidate / Potential Candidate countries data included in the ESPON Database per theme, territorial level, time series and source.....	11
Table AL.1: Population per prefecture 2001 .....	12
Map AL.1: Albania similar NUTS2 and 3 units, Population per similar NUTS3 2001 .....	13
Map BH.1: Bosnia and Herzegovina similar NUTS2 and 3 units, Population per similar NUTS3 2001 .....	16
Table BH.1: Official estimate of the population of FBiH cantons ("similar NUTS3") 2007 .....	17
Map CR.1: Croatia NUTS2 and 3 units, Population per NUTS3 2001 .....	20
Table CR.1: Croatian counties (NUTS3) population in 2001 .....	21
Table FY.1 Population 2002 of the FYROM regions / NUTS3 .....	22
Map FY.1: FYROM NUTS2 and 3 units, Population per NUTS3 2001 .....	23
Map SE.1: Serbia similar NUTS2* and NUTS3 units, Population per similar NUTS3 2001 .....	26
Figure SE.2: Serbia "similar NUTS2" according to the 2010 reform .....	27
Map TU.1: Turkey NUTS2 and 3 units, Population per NUTS3 2001 .....	33
Map Annex 1: Population density at NUTS3 level in South-eastern Europe: EU and Candidate / Potential Candidate Countries: Western Balkans and Turkey .....	36
Map Annex 2: EU and Western Balkans and Turkey Population 65 years and over Rate % at NUTS3 or similar NUTS3 level 2008 .....	37
Table Annex 2: Western Balkans and Turkey available (in 2009) territorial data, at NUTS3 or "similar NUTS3" level from National Statistical Organisations (NSO) (censuses, estimations etc) .....	38

## Introduction and methodological remarks

The present Technical Report corresponds to the section 2.8 of the Final Report of the ESPON 2013 Database project: "Enlargement to Neighbourhood" (as well as to the "Challenge 11" of the project).

In this thematic issue we aimed to explore the possibilities to extend the pool of data on the ESPON countries on the **Western Balkans (WB) countries and Turkey** as well as to study how ensure that the relevant data be harmonized with the rest of the ESPON Database.

**Western Balkans countries and Turkey** are **Candidate Countries (CC)** or **Potential Candidate Countries (PCC)**. Specifically:

According to the overall enlargement strategy of the EU document adopted by the Commission on 8.11.2006 ([http://ec.europa.eu/enlargement/countries/index\\_en.htm](http://ec.europa.eu/enlargement/countries/index_en.htm)) **Croatia** and **Turkey** are **Candidate Countries**. In December 2005, the European Council granted the **Former Yugoslav Republic of Macedonia (FYROM)** the status of a **Candidate Country**; The European Council of 16-17.12.2010 agreed to give **Montenegro** the status of **Candidate country**; accession negotiations with these two countries have not started<sup>1</sup>.

**Albania, Bosnia and Herzegovina and Serbia including Kosovo** (Under UN Security Council Resolution 1244) are **Potential Candidate Countries**: See in more detail in the 8.11.06 document and the corresponding following documents.

**Key findings** of our work are:

- In the Database have been integrated datasets with 139 indicators on the CC / PCC for the following "**basic**" themes: GDP, Area, Population, Population density, Age pyramid, Crude Births rate, Crude Deaths rate, Natural growth rate, Active population, Migration and Population sex, for the years 2000 – 2007 and NUTS 0, 1, 2 and 3 or "similar NUTS" (see in next) 1, 2 and 3 levels.

These data are provided from Eurostat and National Statistical Organisations (NSO) and are fully comparable with the data for the ESPON space.

For Croatia, FYROM and Turkey, which have adopted the NUTS classification, available data covers the large majority of themes and years at NUTS3 level, while for Albania, Bosnia and Herzegovina, Montenegro and Serbia including Kosovo there are considerable gaps at NUTS2 and NUTS3 levels.

- We have also examined the availability and quality of data for the CC / PCC provided by **Eurostat, NSO sites and other sources** for a limited number of **additional** themes on demography / society, economy / employment and environment considered of primary importance for territorial analyses at NUTS3 level (for example data from the Labour Force surveys, from Corine Land Cover etc).

From this assessment, we concluded that available data for Croatia, FYROM, Turkey and Serbia covers a large number of themes, years and levels and working on harmonisation is relatively easy, while for the other CC / PCC there are numerous gaps and working on harmonisation of the existing data from NSO is much more difficult.

We should finally mention that Eurostat provides, for these additional themes, data for all CC / PCC at **NUTSO level**. These data are less useful for territorial analyses. They are not

---

<sup>1</sup> The statistics of Montenegro are not actually adapted to the EU statistics rules as for the case of Croatia, FYROM and Turkey. Thus, in the following, we have included Montenegro in the Potential CC.

included in the Database as this last does not contain respective data for the ESPON countries.

## Summary of the methodological issues

In order to ensure a sound comparability of data of the CC and PCC which have not adopted the NUTS classification, we have classified the existing administrative units of these countries at different territorial levels in **“similar NUTS”** territorial units.

We have used for this purpose the **criterion of population potential** of the EU NUTS classification as well as the **overall structure of government** in these countries **with focus on the power of the respective regional and local authorities** and the main features of territorial development in each administrative level per country.

The implementation of this method ensured that the “similar NUTS” divisions correspond almost fully with the respective divisions for the EU countries.

While ESPON 2006 Database as well as several relevant to the CC / PCC studies in the frame of ESPON 2006 focused mainly on NUTS2 data, ESPON 2013 Database project has, in general terms, extended the scope of the Database to NUTS3 level data. In this line, the work on CC / PCC in the project focused on the NUTS3 level.

The method of definition of “similar NUTS” and, more in general, the methodological conclusions of the work on CC / PCC could be used for the future work on the inclusion of data for the Eastern Neighbouring countries (ENC) and the Southern Mediterranean Neighbouring countries (MNC) in the scope of the 2<sup>nd</sup> project of the ESPON database 2013.

## ***NUTS and regional / territorial classification***

We first had to assess ***the conformity of the WB and Turkey spatial administrative divisions to the EU NUTS classification criteria.***

According to Eurostat, **NUTS – Nomenclature of Territorial Units for Statistics, 2010<sup>2</sup>**, the NUTS classification is a hierarchical system for dividing up the economic territory of the EU for the purpose of:

- The collection, development and harmonisation of EU regional statistics:
- Socio-economic analyses of the regions.
  - NUTS 1: major socio-economic regions
  - NUTS 2: basic regions for the application of regional policies
  - NUTS 3: as small regions for specific diagnoses
- Framing of EU regional policies.
  - Regions eligible for aid from the Structural Funds (Objective 1) have been classified at the NUTS 2 level.
  - Areas eligible under the other priority objectives have mainly been classified at the NUTS 3 level.
  - The Cohesion report has so far mainly been prepared at the NUTS 2 level.

---

<sup>2</sup> [http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts\\_nomenclature](http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature) Last update 28.07.2010



## Principles and Characteristics of NUTS

- **Principle 1:** The NUTS regulation defines minimum and maximum population thresholds for the size of the NUTS regions:

<b>NUTS 1</b>	3 million	7 million
<b>NUTS 2</b>	800 000	3 million
<b>NUTS 3</b>	150 000	800 000

Despite the aim of ensuring that regions of comparable size all appear at the same NUTS level, each level still contains regions which differ greatly in terms of population.

- **Principle 2:** NUTS favours administrative divisions (normative criterion)  
For practical reasons the NUTS classification is based on the administrative divisions applied in the Member States that generally comprise two main regional levels. The additional third level is created by aggregating administrative units.
- **Principle 3:** NUTS favours general geographical units  
General geographical units are normally more suitable for any given indicator than geographical units specific to certain fields of activity.

EU NUTS classification uses almost exclusively the population criterion.

In the above explanatory texts of the EU NUTS classification, NUTS2 units are seen (to some extent) as "basic regions for the application of regional policies" while NUTS3 units are approached as "small regions for specific diagnoses".

Regarding NUTS2 units in the case of the Western Balkans countries which have not adopted NUTS regulation, it is obvious that these countries should define "similar to NUTS2" units which are appropriate for the application of the EU regional policies (see, among others, in: Knezevic 2010). However, they would evidently apply a political criterion: the use of existing administrative divisions as they are or the limited re-adjustment of them or the creation of new divisions depend on the political will of these countries. Evidently, they will also take into account a relevant technical criterion: the territorial features of the existing "regions".

Therefore, it is essential to examine if the existing administrative divisions in these countries which fulfill the population criterion for NUTS2, fulfill also the "application of regional policies" criterion. In order to better clarify this criterion, we should refer, even very concisely, to the concept of "development policy region".

Regions can be defined by physical, "functional", economic, social, cultural, environmental characteristics and so on; each of the many branches of territorial analysis and planning approaches regions according to its general theoretical scope. For the case of "development policy regions" in the PCC of WB most appropriate approach is to examine whether "similar NUTS2" divisions according to the NUTS populations criterion, corresponding to existing administrative divisions, comply with the "regional level" of the «spatial governance» system of each of the examined countries.

Even more, it is useful to see whether the "similar NUTS2" and "similar NUTS3" units which will be defined comply with the overall spatial configuration of the interested countries with focus on the hierarchy and the networking of the urban centres of these countries.

In next we will examine these territorial features mainly for each of the Potential CC; we will also examine this issue, relatively less, for the cases of the CC which have already adopted the EU NUTS classification.

## 1. General assessment

### 1.1 The WB and Turkey Spatial Administrative Divisions

Turkey, Croatia and FYROM have already adopted the NUTS classification.

The rest of the WB countries are at the present in the procedure of adopting it. According to the assessment using the population criterion, in the majority of these last the existing administrative divisions (regions, districts etc) could be associated to the EU NUTS definitions without considerable problems. In other words, we could create **"similar NUTS" divisions** which fulfil satisfactorily the NUTS population criterion - see in Tables 1.1 and 1.2 and in detail in section 2.

According to our methodological approach (see previously), we have examined, as possible, the administrative capacity of the NUTS and "similar NUTS" spatial administrative divisions of the WB and Turkey and, further on, the overall territorial features of these countries. As we will see in next, the results have more or less consolidated the selection of the "similar NUTS" divisions.

**Table 1.1: NUTS 1, 2, 3 regions in Croatia, FYROM and Turkey**

	<b>NUTS 1</b>	<b>NUTS 2</b>	<b>NUTS 3</b>
Croatia	Country	Regija	Counties
FYROM	Country	Country	Statistical Regions
Turkey	Regions	Sub-regions	Provinces

**Table 1.2: "Similar NUTS 1, 2, 3" regions in Albania, Bosnia and Herzegovina, Serbia, Montenegro and Kosovo\*\***

	<b>Similar to NUTS 1</b>	<b>Similar to NUTS 2</b>	<b>Similar to NUTS 3</b>
Albania	Country	(Country)	12 Prefectures
BeH	Country or: FBiH, RS, Brsko district	FBiH, RS, Brsko district	10 Cantons
Serbia	Central Serbia, Voivodina	(Central Serbia, Voivodina)	21 Districts
Montenegro *	Country	Country	Country
Kosovo* **	Country	Country	(Country)

\* See in more detail in the per country assessment

\*\* Under UN Security Council Resolution 1244

## 1.2 The data required, the “basic” data included in the ESPON Database and the “additional” data

Undoubtedly, we need a very large range of territorial data in order to support research on all kind of territorial challenges, policies and perspectives in the case of CC/PCC -as in the case of the ESPON space.

Thus necessarily we started from a limited list of themes and respective data which are considered of “primary importance for territorial analyses at NUTS3 level”: the **“required data”**.

Then we defined an even more limited set of **“basic”** data / indicators (and themes) which are used to analyse “basic” regional challenges and, at the same time, they highly comparable with the data for the ESPON space.

Finally, we characterised as **“additional data”** the data which complete the range of “required data” and we explored their availability and quality in the case of the CC/PCC.

As we have already mentioned, we paid particular attention in finding out if there exist such data at NUTS3 level at each of the CC / PCC and, also, if the existing data allow diachronic comparisons.

### ***The “required data”***

The range of themes and data / indicators which could be considered of “primary importance for territorial analyses at NUTS3 level” is evidently very large. We mainly refer to the following aspects of NUTS3 areas:

(a) Demographic and social aspects:

Population age, gender, education level etc, crude birth / death rate, population growth rate, population education level, migration, poverty, characteristics of households and dwellings etc.

(b) Economic aspects and employment:

GDP, innovation, Active population, employment and unemployment rates, division per sectors and branches etc.

(c) Land use and environmental aspects.

### ***The “basic” data included in the ESPON 2013 Database***

In line with the entire ESPON Database project, the following **“basic”** themes and corresponding datasets which contain 132 indicators have been elaborated and **integrated in the Database**.

- GDP in Euros and GDP in PPS
- Area
- Total population
- Population density

- Population age pyramid
- Population by sex
- Crude birth rate
- Crude death rate
- Migration
- Natural Growth rate
- Active population
- Unemployment

These data are coming from all available sources: **Eurostat, National Statistical Offices and other sources** and covers the years 2000 – 2007 and NUTS 0, 1, 2 and 3 or “similar NUTS” 1, 2 and 3 levels.

See for the respective **metadata** in the Database.

More specifically:

- For Croatia, FYROM and Turkey most of the data is from Eurostat, but we have, also, added data from NSO and other sources.
- For the rest CC / PCC, Eurostat provides data only at NUTS0 level –see in next; therefore, we have mainly used data from the National Statistical Organisations (NSO) and other sources.

Regarding the coverage rate of the CC / PCC per theme of the basic data, the situation varies considerably according to the country -see in Table 1.3.

## The “additional” data

(a) We give in the *Annex 2*, the Table 1 of the **existing data sources, mainly from NSO**, per CC / PCC, per group of themes and per census / survey in which the data are based, for the NUTS3 level and for a relatively long period: from 1990 to 2008.

*Usually realised censuses and specific statistical surveys concerning the ESPON Database data and indicators:*

- Population censuses, building / dwellings censuses and inventories
- Labour force surveys, household budget surveys etc.

The censuses, inventories and surveys already done in the CC / PCC are for the most comparable with those realized in the EU-27 countries.

(b) Regarding the available **“additional”** data on CC / PCC from **Eurostat**:

(1) Eurostat provides, for the “additional” themes, data for all CC / PCC at **NUTS0 level**. These data are less useful for territorial analyses. They are not included in the Database as this last does not contain respective data for the ESPON countries.

(2) Eurostat also provides a wide range of data at NUTS 1, 2, 3 levels for Croatia, FYROM and Turkey.

The data provided by Eurostat for the CC / PCC are presented in the following two Annexes:

- *Annex 3*: short presentation of the respective Eurostat data on CC per theme / sub-theme
- *Annex 4*: full description of the respective data.

(c) All CC / PCC are included in CORINE Land Cover and other land based EU programs providing useful land use and environmental data (at lower level than NUTS3 / similar NUTS3).

(d) According to the overall assessment on the availability and quality of data for the CC / PCC that we have made, we concluded that available data for Croatia, FYROM, Turkey and Serbia covers a large number of themes, years and levels and working on harmonisation is relatively easy, while for the other CC / PCC there are numerous gaps and working on harmonisation of the existing data from NSO is much more difficult – see also in the Conclusions.

(e) Since the situation in the CC / PCC varies considerably from country to country, it is necessary to present **in section 2** the results of the overall assessment that we have done regarding the availability and quality of data **per country**. We refer, specifically to:

(1) The *Eurostat data* (primarily)

(2) *The data provided by the National Statistical Organisations (NSO) of the CC / PCC*

(3) Data from a wide range of other sources: ESPON 2006 projects, ESTIA-SPOSE programme, other relevant INTERREG programmes, Wikipedia, [www.citypopulation.de](http://www.citypopulation.de) etc (see in References – Sources). The latter were mainly used to cross-check the data from Eurostat and NSO.

### ***The administrative divisions' shapefiles***

For Croatia, FYROM and Turkey there are NUTS shapefiles provided by Eurogeographics. There are not respective shapefiles (for the administrative divisions assimilated with "similar NUTS") for Albania, Bosnia and Herzegovina, Serbia, Montenegro and Kosovo (under UN Security Council Resolution 1244). Non official shapefiles for these countries have been downloaded from the Berkeley University website. Then, these shapefiles have been appropriately adjusted (by the RIATE workgroup) and included in the ESPON template.

**Table 1.3: Candidate / Potential Candidate countries data included in the ESPON Database per theme, territorial level, time series and source**

Candidate / Pot. Candidate countries	NUTS / Similar NUTS classification	GDP	AREA	Population	Population density	Age pyramid	Crude birth rate, Crude death rate	Natural growth rate	Active population, Unemployment	Migration	Population sex
<b>Albania</b>	NUTS0		2000-2006 [N]	2000-2007[E]	2000-2007 [E]	2005 [E]	2000-2007 [E]	2000-2007 [E]		2006 [E]	2000-2007 [E]
	Similar NUTS1		2000-2006 [N]	2000-2007[N]	2000-2007 [E]						
	Similar NUTS2		2000-2006 [N]	2000-2007[N]	2000-2007 [E]						
	Similar NUTS3		2000-2006 [N]	2001 & 2004 [N]	2001 & 2004 [N]						
<b>Bosnia &amp; Herzegovina</b>	NUTS0		2000-2006 [N]	2000-2007[N]	2000-2007 [E]		2000-2007 [E]	2000-2007 [E]			2000-2007 [E]
	Similar NUTS1		2000-2006 [N]	2000-2007[N]							
	Similar NUTS2		2000-2006 [N]	2000-2007[N]	2006 & 2007 [N]						
	Similar NUTS3		2000-2006 [N]	2000-2007[N]	2006 & 2007 [N]						
<b>Croatia</b>	NUTS0	2000-2006 [E]	2000-2006 [E]	2000-2007 [E]	2000-2007 [E]	2005 [E]	2000-2007 [E]	2000-2007 [E]	2002-2007 [E]	2000-2007 [E]	2000-2007 [E]
	NUTS1	2000-2006 [E]	2000-2006 [E]	2001-2007 [E]	2001-2006 [E]	2005 [E]	2002-2007 [E]	2002-2007 [E]	2007 [E]	2002-2007 [E]	2001-2007 [E]
	NUTS2	2000-2006 [E]	2000-2006 [E]	2001-2007 [E]	2001-2006 [E]	2005 [E]	2002-2007 [E]	2002-2007 [E]	2007 [E]	2002-2007 [E]	2001-2007 [E]
	NUTS3	2000-2006 [E]	2000-2006 [E]	2001-2007 [E]	2001-2006 [E]	2001 [N]	2002-2007 [E]	2002-2007 [E]		2002-2007 [E]	2001-2007 [E]
<b>FYROM</b>	NUTS0	2000-2006 [E]	2000-2006 [E]	2000-2007 [E]	2000-2007 [E]	2005 [E]	2000-2007 [E]	2000-2007 [E]		2004-2007 [E]	2000-2007 [E]
	NUTS1	2004-2006 [E]	2000-2006 [E]	2000-2007 [E]	2000-2006 [E]	2005 [E]	2004-2007 [E]	2004-2007 [E]		2004-2007 [E]	2000-2007 [E]
	NUTS2	2004-2006 [E]	2000-2006 [E]	2000-2007 [E]	2000-2006 [E]		2004-2007 [E]	2004-2007 [E]		2004-2007 [E]	2000-2007 [E]
	NUTS3	2004-2006 [E]	2000-2006 [T]	2000-2007 [E]	2000-2006 [E]		2004-2007 [E]	2004-2007 [E]		2004-2007 [E]	2000-2007 [E]
<b>Serbia</b>	NUTS0		2000-2006[N]	2000-2007[E]	2000-2007 [E]	2005 [E]	2000-2007 [E]	2000-2007 [E]			2000-2007 [E]
	Similar NUTS1		2000-2006[N]	2001-2007[N]							
	Similar NUTS2		2000-2006[N]	2001-2007[N]							
	Similar NUTS3		2000-2006[N]	2001-2007[N]	2002[N]	2002 [s]					
<b>Montenegro</b>	NUTS0		2000-2006[N]	2000-2007[N]	2000-2007 [E]	2005 [E]	2000-2007 [E]	2000-2007 [E]		2007 [E]	2000-2007 [E]
	Similar NUTS1		2000-2006[N]	2000-2007[N]	2000-2007 [E]						
	Similar NUTS2		2000-2006[N]	2000-2007[N]	2000-2007 [E]						
	Similar NUTS3		2000-2006[N]	2000-2007[N]	2000-2007 [E]						
<b>Kosovo</b> (Under UN Security Council Resolution 1244)	NUTS0		2000-2006[N]	2002-2007 [N]	2002-2007 [N]		2003-2007 [E]	2003-2007 [E]			2006 & 2007 [E]
	Similar NUTS1		2000-2006[N]	2002-2007 [N]	2002-2007 [N]						
	Similar NUTS2		2000-2006[N]	2002-2007 [N]	2002-2007 [N]						
	Similar NUTS3		2000-2006[N]	2002-2007 [N]	2002-2007 [N]						
<b>Turkey</b>	NUTS0	2000-2005 [E]	2000-2006[N]	2000-2006 [E]	2000-2007 [E]	2005 [E]	2000-2007 [E]	2000-2007 [E]	2006-2007 [E]		2000-2007 [E]
	NUTS1	2000-2001 [E]	2000-2006[N]	2000-2006 [E]	2000-2006 [E]				2006-2007 [E]		
	NUTS2	2000-2001 [E]	2000-2006[N]	2000-2006 [E]	2000-2006 [E]				2006-2007 [E]		
	NUTS3	2000-2001 [E]	2000-2006[N]	2000-2006 [E]	2000-2006 [E]						

E=Eurostat

[s] National Statistical Organisation of Serbia

See in the respective metadata files of the Database

N= National Statistical Organisation (NSO)

[T] Total land use area (no data for the total area)

## 2. Assessment per country

### Albania

#### *Spatial units' levels:*

The **total population of the country** amounted up to **3.170.000 inhabitants** in **2008** (Eurostat 2010).

Albania is divided into 12 prefectures (counties, Albanian: official qark/qarku, but often prefecture / prefektura), 37 districts and 351 municipalities.

Concerning the EU regulation for NUTS 3: 150 000 - 800 000 inhabitants; all Albania's prefectures, except two, have 150 000 - 800 000 inhabitants in 2001.

Prefectures have a Council and considerable competences which are gradually extended.

Therefore, **Albania's prefectures could be assimilated to NUTS3** – Table AL.1 and Map AL.1.

**Table AL.1: Population per prefecture 2001**

	Prefecture	Population 2001	Population NUTS criterion
1	Tiranë	597.899	150.000 – 800.000
2	Fier	382.544	
3	Elbasan	362.736	
4	Shkodër	256.473	
5	Durrës	245.179	
6	Vlorë	192.982	
7	Korçë	265.182	
8	Berat	193.020	
9	Dibër	189.854	
10	Lezhë	159.182	
11	Gjirokastër	112.831	112.831
12	Kukës	111.393	111.393
	Total Alb.	3.069.275	3.069.275

There is not an official territorial division which could be assimilated to **EU NUTS2 division**. Several scenarios are now examined in the context of the EU – Albania collaboration.

Apart from the capital: Tirana, seven other cities of the country have (roughly) more than 200.000 inhabitants: Fier, Elbasan, Shkoder, Durrës, Vlore, Korce and Berat. The eight major urban centres are capitals of prefectures; this is in line with the correspondence of the prefectures to "similar NUTS3" units. However, apart from Tirana there are not other urban centres which obviously provide services of higher level than the prefectural level.



**Population per Similar NUTS3 2001**

- 111.390 - 150.000
- 150.000 - 475.000
- 475.000 - 800.000
- > 800.000

**NUTS3 population thresholds:**  
minimum: 150.000 inhabitants  
maximum: 800.000 inhabitants

- ▭ Similar NUTS2 regions
- ▭ Similar NUTS3 regions

**Map AL.1: Albania similar NUTS2 and 3 units, Population per similar NUTS3 2001**



## **Existing data at "similar NUTS3" level (2009)**

(1) Official statistical data:

*Data at the level of prefectures ("counties") / similar NUTS 3:*

- From the *population censuses of 1989 and 2001:*

(a) Population: total, distributions: per sex and age group, per education level

(b) Active population (total, distributions: per sex), number of employed and unemployed persons, employment per primary secondary and tertiary sector.

- From the housing census of 2001

The *Labour Force Survey of 2007* refers to the national level.

Moreover, some research about *population projections 2001-2021, gender perspectives, people and work and living conditions and inequality* exist only for national level or the level of regions (north, centre except from Tirana – Durres, South and Tirana – Durres).

**See in the Table 1 in Annex 2.**

(2) Data on land uses and environment -from CLC, UMZ.

## **Albania Data included in the ESPON Database**

The data is only at NUTS0 and NUTS3 levels.

We used the data of NUTS0 level for the "similar NUTS1" and "similar NUTS2" regions (total of the country).

● GDP (Euros, PPS)	
● Area	NUTS 0 & similar NUTS1,2,3 (2000-2006 / ESTAT)
● Total population	NUTS 0, similar NUTS1,2 (2000-2007/NSO), similar NUTS3 (2001&2004/NSO)
● Population density	NUTS 0 & similar NUTS1,2 (2000-2007 / ESTAT), similar NUTS 3 (2001 & 2004 / NSO)
● Age pyramid	NUTS 0 (2005 / ESTAT)
● Pop by sex	NUTS 0 (2000-2007 / ESTAT)
● Crude birth rate	NUTS 0 (2000-2007 / ESTAT)
● Crude death rate	NUTS 0 (2000-2007 / ESTAT)
● Migration	NUTS 0 (2006 / ESTAT)
● Natural Growth rate	NUTS 0 (2000-2007 / ESTAT)
● Active population, Unemployment	

## Bosnia and Herzegovina

### ***Spatial units' levels:***

The **total population of the country** amounted up to **3.843.000 inhab.** in **2008** (Eurostat 2010).

Bosnia and Herzegovina is divided into three entities: *Federation of Bosnia and Herzegovina (FBiH), Republic of Srpska (RS), and Brčko District*, which was established in 2000 out of land from both entities<sup>3</sup>.

FBiH is divided in **10 cantons** –*Table BH1. and Map BH.1-* and 79 municipalities; Republic of Srpska has 62 municipalities; City of Brčko is a separate administrative unit - District.

It is difficult to associate the Bosnia and Herzegovina administrative units with corresponding NUTS levels, because the magnitudes of the population of the units belonging to each administrative level are dissimilar.

An additional difficulty relies on the fact that for the RS there is no census or official estimation after 2001; consequently, the estimations occurred by several sources differ significantly among each other. For FBiH there is a very recent (2007) official estimation of the population (from the FBiH's Federal Office of Statistics) that we use in the following.

FBiH (population 2007: 2.328.000), RS (population 2007 estimate: 1.439.700) and Brčko (population 2007 estimate: 68.860) could be assimilated to NUTS1 and / or NUTS2.

EU regulation population criterion for NUTS 3: 150.000-800.000 inhabitants; 6 FBiH's cantons have 227.000-496.000 inh, while 4, have 34.000 - 82.000 inhabitants in 2007. Obviously, according to this criterion, the 4 smaller cantons could be difficultly assimilated to NUTS3 units - See in Table BH.1.

The administrative power / capacity of the cantons is considerable: they have their own cantonal government, which is under the law of the Federation as a whole.

### ***The 10 cantons of FBiH could be assimilated to NUTS3.***

Specifically for the "**similar NUTS2 level**":

According to the respective population criterion: 800 000 - 3 million inhabitants, both FBiH and RS could be assimilated to NUTS2. Regarding also the administrative power / capacity criterion, these units could be assimilated to NUTS2. However, evidently, in the case of Bosnia and Herzegovina, the political criterion would be taken primarily into account, therefore even Brčko District could be assimilated to NUTS2.

Regarding the **national urban system**, apart from Sarajevo, the capital (with more than 300.000 inhabitants), which has a really primary role, there are six important regional centres: Banja Luka (in RS: 250.000 inh.) Mostar (FBiH), Tuzla (FBiH), Zenic (FBiH), Bijeljina (RS) and Prijedor (RS); each of these last has 110.000 – 140.000 inhabitants. There are also a number of secondary regional centres (two of them, situated in FBiH, have more than 50.000 inhabitants).

---

<sup>3</sup> It officially belongs to both, but is governed by neither, and functions under a decentralized system of local government.



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

Regional level: Similar NUTS 3  
Source: Eurostat, 2001  
Origin of data: Eurostat, 2010

© EuroGeographics Association for administrative boundaries

**Population per Similar NUTS3 2001**

- 35.260 - 150.000
- 150.000 - 475.000
- 475.000 - 800.000
- > 800.000

**NUTS3 population thresholds:**  
minimum: 150.000 inhabitants  
maximum: 800.000 inhabitants

- Similar NUTS2 regions
- Similar NUTS3 regions

**Map BH.1: Bosnia and Herzegovina similar NUTS2 and 3 units, Population per similar NUTS3 2001**

**Table BH.1: Official estimate of the population of FBiH cantons ("similar NUTS3") 2007**

	Surface area km <sup>2</sup>	Population, 2007 <sup>1)</sup>	Population density per km <sup>2</sup> 2007
Federacija Bosne i Hercegovine	<b>26.110,5</b>	<b>2.328.359</b>	<b>89,2</b>
Unsko-sanski kanton	4.125,0	287.878	69,8
Kanton Posavski	324,6	41.187	126,9
Tuzlanski kanton	2.649,0	496.830	187,6
Zeničko-dobojski kanton	3.343,3	401.796	120,2
Bosanskopodrinjski kanton	504,6	33.662	66,7
Srednjobosanski kanton	3.189,0	256.339	80,4
Hercegovačko-neretvanski kanton	4.401,0	227.473	51,7
Zapadno-hercegovački kanton	1.362,2	82.095	60,3
Kanton Sarajevo	1.276,9	419.030	328,2
Kanton 10	4.934,9	82.069	16,6

The major urban centres are capitals of cantons; this is in line with the correspondence of the cantons to "similar NUTS3" units.

### **Existing data at "similar NUTS3" level**

(1) Official statistical data:

*Data at the level of 3 entities:* Federation of Bosnia and Herzegovina (FBiH), Republic of Srpska (RS) and Brsko District.

- From the population census of 1991:

(a) Population: total, distributions: per sex and age group, per education level

(b) Active population (total, distributions: per sex), number of employed and unemployed persons, employment per primary secondary and tertiary sector.

- From the population official estimate 2008 (for the FBiH): population per sex age etc, active population etc – see above.

- From the Labour Force Survey, carried out in 2007: total active population and its sex distribution, number of employed and unemployed persons, employment per primary, secondary and tertiary sector.

Data on the GDP exist for the FBiH and RS– at entity level.

*Data at the level of cantons:*

Population 2008 from the population official estimate 2008 –only for the FBiH.

(2) Data on land uses and environment -from CLC, UMZ.

## Bosnia and Herzegovina Data included in the ESPON Database

• GDP (Euros, PPS)	
• Area	NUTS 0 & similar NUTS1,2,3 (2000-2006 / NSO)
• Total population	NUTS 0 & similar NUTS1,2,3 (2000-2007 / NSO)
• Population density	NUTS0 (2000-2007 / ESTAT), similar NUTS2 (2006 & 2007 / NSO), similar NUTS3 (2006 & 2007 / NSO)
• Age pyramid	
• Pop by sex	NUTS 0 (2000-2007 / ESTAT)
• Crude birth rate	NUTS 0 (2000-2007 / ESTAT)
• Crude death rate	NUTS 0 (2000-2007 / ESTAT)
• Migration	
• Natural Growth	NUTS 0 (2000-2007 / ESTAT)
• Active population, Unemployment	

We used as:

Similar NUTS1: the entire country Bosnia and Herzegovina

Similar NUTS2: the Federation of Bosnia and Herzegovina (FBiH), the Republic of Srpska (RS) and the Brčko District

Similar NUTS3: the 10 cantons of FBiH, the Republic of Srpska, and the Brčko District

# Croatia

## *Spatial units' levels*

The **total population of the country** amounted up to **4.436.000 inhab.** in **2008** (Eurostat 2010).

- Croatia has already adopted the EU NUTS (1, 2, 3) classification as follows:

**NUTS 1: Country (Hrvatska), NUTS 2: Regija (3), NUTS 3: Counties / Jupanija (21).** See in Map CR.1.

Only 11 counties had a population ranging between 150.000 and 800.000 inh, in 2001, which are the EU regulation limits for NUTS 3. The 10 remaining counties had a lower population: 54.000-142.000 inhabitants (in 2001).

Croatia includes 3 regions at **NUTS2 level**. It has not established specific administrative structures at NUTS2 level for regional development implementation.

Regarding the **national urban system**, apart from Zagreb, the capital, there are several primary and secondary regional centres; some of these last are located in the touristic coastal zone and / or they are "cities-gates".

In general terms, the capacity of the Croatian urban centres is in line with the divisions in NUTS2 and NUTS3 units.

## **Existing data at NUTS3 level (2009)**

(1) Official statistical data:

### **Data at NUTS3 level:**

- From the population censuses of 1991 and 2001:

(a) Population: total, distributions: per sex and age group, per education level

(b) Active population (total, distributions: per sex), number of employed and unemployed persons, employment per primary secondary and tertiary sector.

- From the population, *households and dwellings* census 2001 (31<sup>st</sup> March 2001).

(2) Data on land uses and environment -from CLC, UMZ.

Data at National level: Labour force survey -First Quarter of 2008.



ESPON © NTUA, ESPON Database, 2010

EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

Regional level: NUTS 3  
Source: Eurostat, 2001  
Origin of data: Eurostat, 2010  
© EuroGeographics Association for administrative boundaries

**Population per NUTS3 2001**

- 53.600 - 150.000
- 150.000 - 475.000
- 475.000 - 800.000
- > 800.000

**NUTS3 population thresholds:**  
minimum: 150.000 inhabitants  
maximum: 800.000 inhabitants

- ▭ NUTS2 regions
- ▭ NUTS3 regions

**Map CR.1: Croatia NUTS2 and 3 units, Population per NUTS3 2001**

**Table CR.1: Croatian counties (NUTS3) population in 2001**

Code	County of:	Pop. 2001
hr035	Split-Dalmatia	463.676
hr031	Primorje-Gorskiototar	305.505
hr025	Osijek-Baranja	330.506
hr033	Zadar	162.045
hr024	SlavonskiBrod-Posavina	176.765
hr028	Sisak-Moslavina	185.387
hr034	Šibenik-Knin	112.891
hr027	Karlovac	141.787
hr037	Dubrovnik-Neretva	122.870
hr021	Bjelovar-Bilogora	133.084
hr014	Varaždin	184.769
hr026	Vukovar-Sirmium	204.768
hr023	Požega-Slavonia	85.831
hr015	Koprivnica-Križevci	124.467
hr022	Virovitica-Podravina	93.389
hr016	Međimurje	118.426
hr032	Lika-Senj	53.677
hr013	Krapina-Zagorje	142.432
hr036	Istria	206.344
hro11	Grad Zagreb (City of Za-	779.145
hr012	Zagreb zupan.	309.696
Hr	REP. OF CROATIA	4.127.764

### Croatia Data included in the ESPON Database

*Data provided mainly by Eurostat*

• GDP (Euros, PPS)	NUTS 0,1,2,3 (2000-2006 / ESTAT)
• Area	NUTS 0,1,2,3 (2000-2006 / ESTAT)
• Total population	NUTS 0 (2000-2007 / ESTAT), NUTS1,2,3 (2001-2007/ESTAT)
• Population density	NUTS 0 (2000-2007 / ESTAT), NUTS1,2,3 (2001-2006 / ESTAT)
• Age pyramid	NUTS 0,1,2 (2005 / ESTAT), NUTS3 (2001 / NSO)
• Pop by sex	NUTS 0 (2000-2007 / ESTAT), NUTS1,2,3 (2001-2007 / ESTAT)
• Crude birth rate	NUTS 0 (2000-2007 / ESTAT), NUTS 1,2,3 (2002-2007/ESTAT)
• Crude death rate	NUTS 0 (2000-2007 / ESTAT), NUTS 1,2,3 (2002-2007/ESTAT)
• Migration	NUTS 0 (2000-2007 / ESTAT), NUTS1,2,3 (2002-2007 / ESTAT)
• Natural Growth	NUTS 0 (2000-2007 / ESTAT), NUTS1,2,3 (2002-2007 / ESTAT)
• Active population, Unemployment	NUTS 0 (2002-2007 / ESTAT), NUTS1,2 (2007 / ESTAT)

*We assimilated NUTS1 level to NUTS0 level.*



# FYROM

## *Spatial units' levels*

The **total population of the country** amounted up to **2.045.000 inhab.** in **2008** (Eurostat 2010).

FYROM has already adopted the EU classification of spatial units in NUTS; by level:

**NUTS 1 and NUTS 2: Country, NUTS 3: Eight (8) Statisticki Regioni / Statistical Regions** – See in the **Map FY.1.**

**Table FY.1 Population 2002 of the FYROM regions / NUTS3**

Code	Regions / NUTS3	Pop. 2002
mk008	Skopje	571.040
mk002	Eastern	203.213
mk007	Northeastern	173.814
mk005	Pelagonia	221.019
mk006	Polog	304.125
mk004	Southeastern	171.416
mk001	Vardar	133.248

In August 2004, FYROM was reorganised into 85 municipalities (10 of which comprise Greater Skopje) which could be assimilated to LAU (1) level.

This is reduced from the previous 123 municipalities established in September, 1996. Prior to this, local government was organised into 34 administrative districts (source: Wikipedia).

Seven (7) from the 8 Statistical Regions had a population ranging between 150.000 and 800.000 inh, in 2002 –which are the EU regulation limits for NUTS 3. The one remaining had a lower population: 133.000-inhabitants in 2002 (Table FY.1).

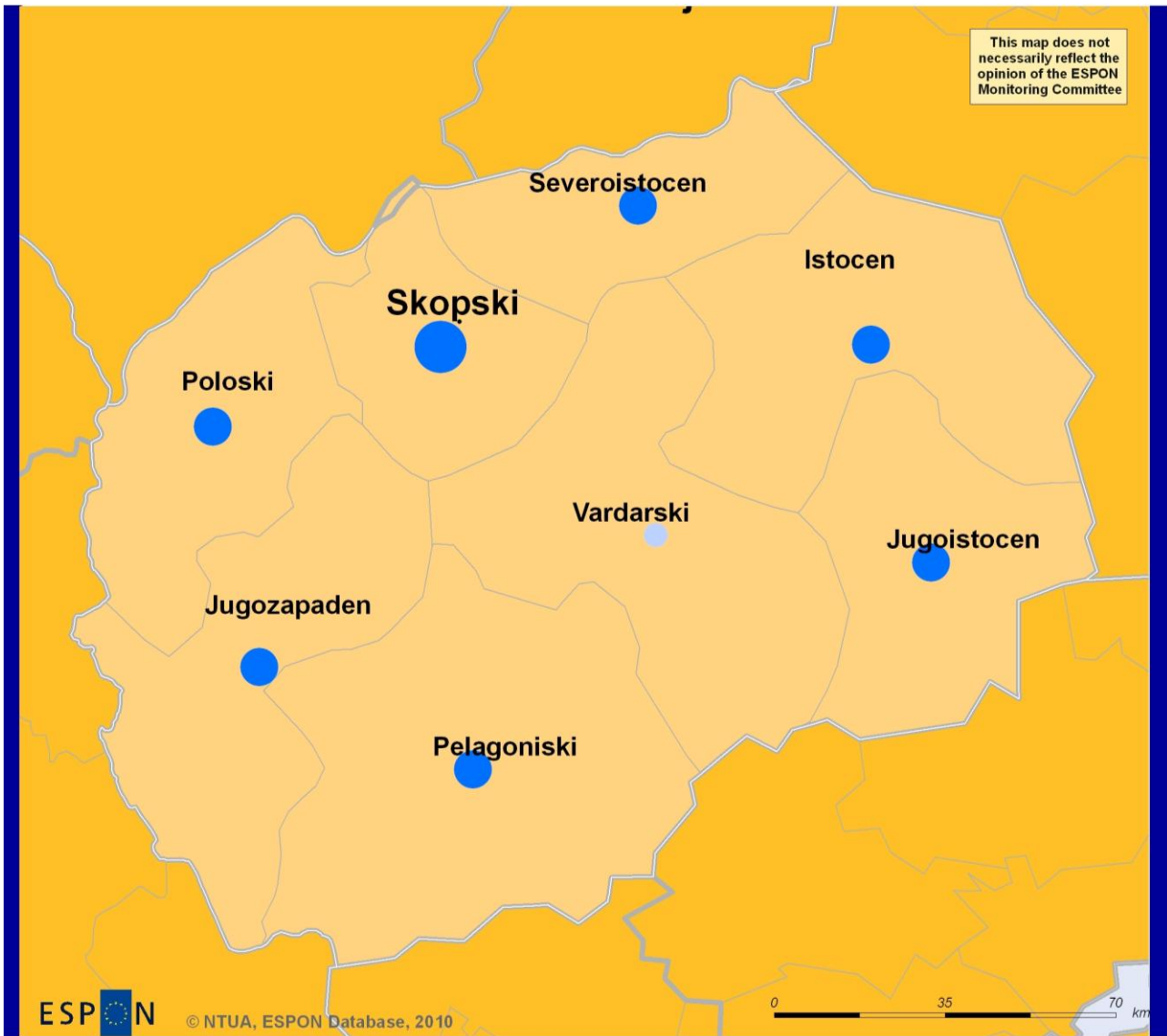
## **Existing data at NUTS3 level (2009)**

(1) Official statistical data:

Data at the level of "Statistical Regions" / NUTS 3 (by aggregation of municipalities' data):

- From the population censuses of 1991 and 2002:
  - (a) Population: total, distributions: per sex and age group, per education level
  - (b) Active population (total, distributions: per sex), number of employed and un-employed persons, employment per primary secondary and tertiary sector.
- From the population, *households and dwellings* census 2002
- Specific surveys: labour force survey etc.

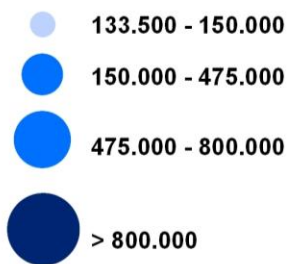
(2) Data on land uses and environment -from CLC, UMZ.



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

Regional level: NUTS 3  
Source: Eurostat, 2001  
Origin of data: Eurostat, 2010  
© EuroGeographics Association for administrative boundaries

**Population per NUTS3 2002**



**NUTS3 population thresholds:**  
minimum: 150.000 inhabitants  
maximum: 800.000 inhabitants

**NUTS3 regions**

**Map FY.1: FYROM NUTS2 and 3 units, Population per NUTS3 2001**

## **FYROM Data included in the ESPON Database**

*Data provided mainly by Eurostat*

• GDP (Euros, PPS)	NUTS 0 (2000-2006 / ESTAT), NUTS1,2,3 (2004-2006 / ESTAT)
• Area	NUTS 0,1,2 (2000-2006 / ESTAT), NUTS3 (2000-2006 / NSO [b])
• Total population	NUTS 0,1,2,3 (2000-2007 / ESTAT)
• Population density	NUTS 0 (2000-2007 / ESTAT), NUTS1,2,3 (2000-2006 / ESTAT)
• Age pyramid	NUTS 0,1 (2005 /ESTAT)
• Pop by sex	NUTS 0,1,2,3 (2000-2007 / ESTAT)
• Crude birth rate	NUTS 0 (2000-2007 / ESTAT), NUTS1,2,3(2004-2007/NSO)
• Crude death rate	NUTS 0 (2000-2007 / ESTAT), NUTS1,2,3(2004-2007/NSO)
• Migration	NUTS 0,1,2,3 (2004-2007 / ESTAT)
• Natural Growth	NUTS 0 (2000-2007 / ESTAT), NUTS1,2,3 (2004-2007 / ESTAT)
• Active population, Unemployment	

[b] Total Land Use Area

*We assimilated NUTS1 and NUTS2 levels to NUTS0 level.*

# Serbia

## *Spatial units' levels*

The **total population of the country** amounted up to **7.366.000 inhabitants** in **2008** (Eurostat 2010).

### *The "similar NUTS3" level*

Serbia is divided into two parts: the **Central Serbia** and the autonomous province of **Vojvodina** and further into **24 districts** (excluding Kosovo) plus the **City of Belgrade**.

The districts and the City of Belgrade are further divided into 157 municipalities – See in **Map SE.1** (the "similar NUTS2" divisions in the Map do not correspond to the recent official division of the Serbia territory in NUTS2 units –see in next)

The territorial organization of the Republic of Serbia is regulated by the Law on Territorial Organization, adopted in the Assembly of Serbia on 29.12.2007. Under the Law, the units of the territorial organization are: municipalities, cities and autonomous provinces.

**Districts (okruzi)** are regional centres of state authority, but have no assemblies of their own; they present purely administrative divisions, and host various state institutions such as funds, office branches and courts. Districts are not defined by the Law on Territorial Organisation, but are organised under the Government's Enactment of 29 January 1992.

Therefore:

- **Districts could be reliably assimilated to NUTS3**; 21 from the 25 districts had a population ranging between 150.000 and 800.000 inhabitants, in 2002 (EU regulation limits for NUTS 3). The four remaining had 102.000-147.000 inhab. in 2002.

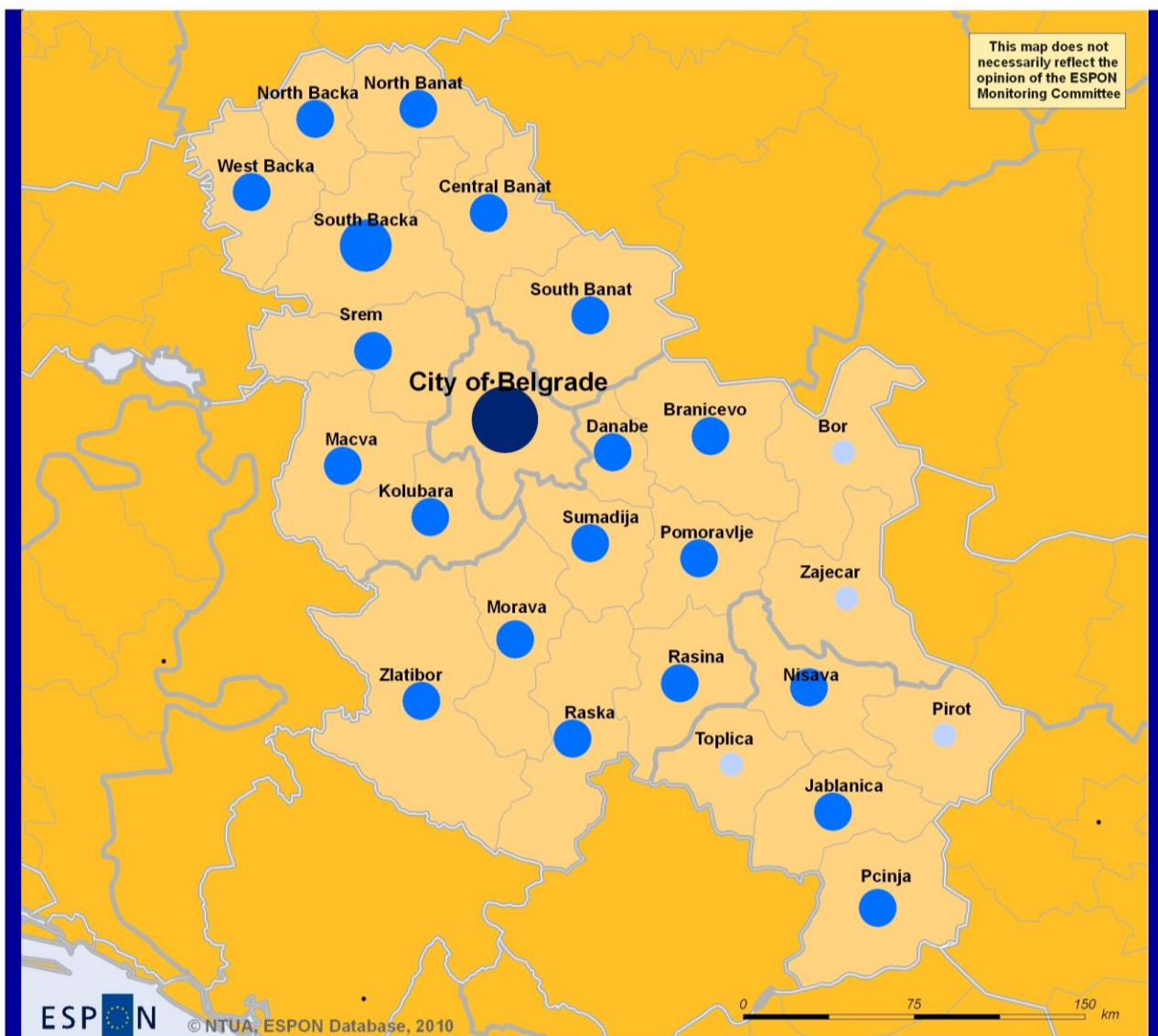
### *The "similar NUTS2" level*

According to the recent available documentation (to be used with caution)

(Wikipedia [http://en.wikipedia.org/wiki/Statistical\\_regions\\_of\\_Serbia](http://en.wikipedia.org/wiki/Statistical_regions_of_Serbia) as of 10.1.2011)

Serbia is divided into **five statistical regions in accordance to NUTS 2**, which are in turn grouped **into two higher NUTS 1 statistical units (North and South)**.

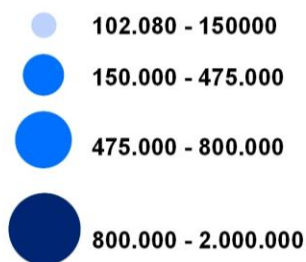
In 2009, the Serbian parliament adopted the Law on Equal Territorial Development that formed seven statistical regions on the territory of Serbia. The Law was amended on 7 April 2010, so that the number of regions was reduced to five –see in **Figure SE.2**. The Eastern Serbia region was merged with Southern Serbia and Šumadija was merged with Western Serbia.



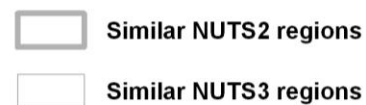
ESPON  
 EUROPEAN UNION  
 Part-financed by the European Regional Development Fund  
 INVESTING IN YOUR FUTURE

Regional level: Similar NUTS 3  
 Source: Eurostat, 2001  
 Origin of data: Eurostat, 2010  
 © EuroGeographics Association for administrative boundaries

**Population per Similar NUTS3 2001**



NUTS3 population thresholds:  
 minimum: 150.000 inhabitants  
 maximum: 800.000 inhabitants



**Map SE.1: Serbia similar NUTS2\* and NUTS3 units, Population per similar NUTS3 2001**

\* Estimation in 2009, before the reform of 2010 –see in next

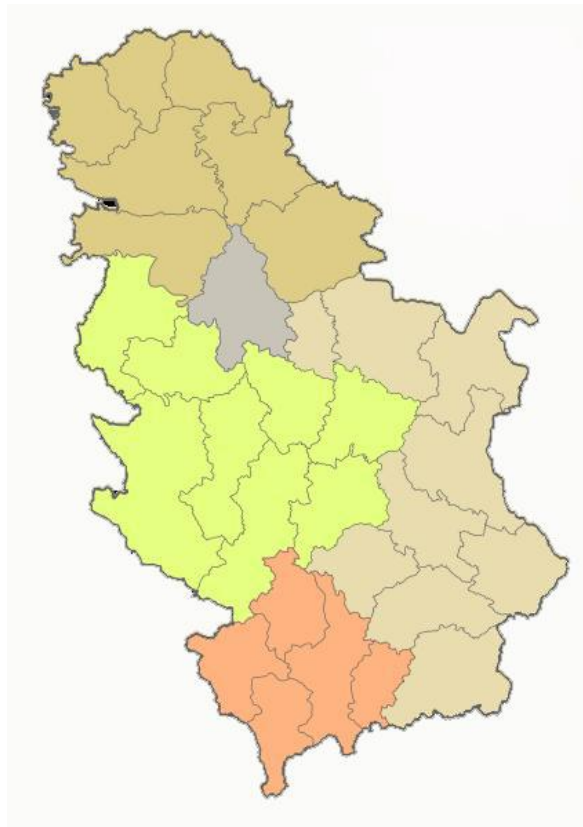
According to the above, the statistical regions and their NUTS codes are:

RS: Serbia

- RS1: Serbia - North
  - RS11: Belgrade
  - RS12: Vojvodina
- RS2: Serbia - South
  - RS21: Šumadija and Western Serbia
  - RS22: Southern and Eastern Serbia
  - RS23: Kosovo and Metohija

If we take into account the population criterion, the two Serbian provinces (plus, eventually, the City of Belgrade) could be assimilated to NUTS 2.

As we have mentioned, the statistical NUTS1 and NUTS2 regions created by the government in order to meet the NUTS criteria as well as the requirements of the EU regional policy, do not have actually a considerable administrative power; also, they are not self-governed entities. The political criterion prevailed for their creation.



**Figure SE.2: Serbia "similar NUTS2" according to the 2010 reform**

Source: Wikipedia 2010

### The *national urban system*

Vojvodina has two big cities: Novi Sad and Subotica, two other cities with 50.000-100.000 inhabitants: Zremjain and Pancevo and 4 cities with 30.000-50.000 inh.

Apart from the capital –Belgrade–, Central Serbia has two other big cities: Nis and

Kragujevac. The rest of the system of cities is balanced with around 10 cities with 50.000- 100.000 inhabitants and a considerable number of cities with 20.000-50.000 inhabitants.

In general terms, the capitals of the districts (corresponding to "similar NUTS3" units) are in most cases enough developed in order to support the development of the respective territorial units. Also, the development of the "new" NUTS2 regions could be supported by respective existing urban centres or networks of urban centres.

### **Existing data at "similar NUTS3" level (2009)**

(1) Official statistical data:

Data at the level of municipalities and districts / similar NUTS 3 (by aggregation of municipalities' data):

-From the population censuses of 1991 and 2002:

(a) Population: total, distributions: per sex and age group, per education level

(b) Active population (total, distributions: per sex), number of employed and unemployed persons, employment per primary secondary and tertiary sector.

*There also data on population (distribution per age, sex etc) from a very recent - 2006- official estimate.*

(2) Data on land uses and environment -from CLC. There are data from CLC2006 but there are not data on UMZ (2009 documentation).

Most of the data concerning censuses of the population and building, specific surveys etc are aggregated and published on the level of *municipalities (LAU1)*.

### **Serbia Data included in the ESPON Database**

• GDP (Euros, PPS)	
• Area	NUTS 0 & similar NUTS1,2,3 (2000-2006 / NSO)
• Total population	NUTS 0 (2000-2007 / ESTAT), similar NUTS1,2,3 (2001-2007/ NSO)
• Population density	NUTS 0 (2000-2007 / ESTAT), similar NUTS3 (2002 / NSO)
• Age pyramid	NUTS 0 (2005/ESTAT), similar NUTS3 (2002 / NSO)
• Pop by sex	NUTS 0 (2000-2007 / ESTAT)
• Crude birth rate	NUTS 0 (2000-2007 / ESTAT)
• Crude death rate	NUTS 0 (2000-2007 / ESTAT)
• Migration	
• Natural Growth	NUTS 0 (2000-2007 / ESTAT)
• Active population, Unemployment	

In 2009, we have assimilated the two parts of Serbia (Central Serbia and Vojvodina) to NUTS1 level as well as to NUTS2 level. As we have mentioned previously, there are more recent approaches of this issue; however, these approaches are not definitive.

# Montenegro

## *Spatial units' levels*

The **total population of Montenegro** amounted in **627.000 inhabitants** in **2008** (Eurostat 2010), therefore **the total of the country could be assimilated to NUT1, NUTS2 and NUTS3**, as according to the EU regulation limits for NUTS 3 spatial units their population should range between 150.000 and 800.000 inhabitants and only the municipality of Podgorica had more than 150.000 inhabitants (169.132) in 2003.

We should remind here that Montenegro has been recognised as candidate Country at the end of 2010.

The country is divided in 21 *municipalities* which could be assimilated to *LAU1* level.

Alternatively: only the Municipality of Podgorica (with population over 150.000) could be assimilated to NUTS3.

## **Existing data at "similar NUTS3" level (2009)**

(1) Official statistical data:

Data are available *mainly for the total of the country*

- From the population censuses of 1991 and 2003:

(a) Population: total, distributions: per sex and age group, per education level

(b) Active population (total, distributions: per sex), number of employed and unemployed persons, employment per primary, secondary and tertiary sector.

- For the education level, the available data exist only in the census of 2003.

(2) Data on land uses and environment -from CLC, UMZ.

## **Montenegro Data included in the ESPON Database**

● GDP (Euros, PPS)	
● Area	NUTS 0 & similar NUTS1,2,3 (2000-2006 / NSO)
● Total population	NUTS 0 & similar NUTS1,2,3 (2000-2007 / NSO)
● Population density	NUTS 0 & similar NUTS1,2,3 (2000-2007 / ESTAT)
● Age pyramid	NUTS0 (2005 / ESTAT)
● Pop by sex	NUTS 0 (2000-2007 / ESTAT)
● Crude birth rate	NUTS 0 (2000-2007 / ESTAT)
● Crude death rate	NUTS 0 (2000-2007 / ESTAT)
● Migration	NUTS 0 (2007 / ESTAT)
● Natural Growth	NUTS 0 (2000-2007 / ESTAT)
● Active population, Unemployment	



## Kosovo (Under UN Security Council Resolution 1244)

### *Spatial units' levels*

The **total population of Kosovo** amounted up to **2.153.000 inhabitants in 2008** (Eurostat 2010).

Republic of Kosovo is divided in **seven districts** and 30 municipalities.

According to the EU regulation limits for NUTS 3 spatial units their population should range between 150.000 and 800.000 inh; the population of more than the half of the Kosovo districts surpasses 150.000 inhabitants, therefore *the districts could be (difficultly) assimilated to NUTS 3 units*. Municipalities could be assimilated to LAU1 level.

### **Existing data at "similar NUTS3" level (2009)**

(1) Official statistical data:

Data at the level of districts / similar NUTS 3:

- From the population census of 1991 (only)

(a) Population: total, distributions: per sex and age group, per education level

(b) Active population (total, distributions: per sex), number of employed and unemployed persons, employment per primary, secondary and tertiary sector.

- No data available on GDP.

(2) There are data from CLC 2006.

Data at national level:

Labour force survey 2002, Labour Market Statistics 2007

### **Kosovo Data included in the ESPON Database**

● GDP (Euros, PPS)	
● Area	NUTS 0 & similar NUTS1,2,3 (2000-2006 / NSO)
● Total population	NUTS 0 & similar NUTS1,2,3 (2002-2007 / NSO)
● Population density	NUTS 0 & similar NUTS1,2,3 (2002-2007 / ESTAT)
● Age pyramid	
● Pop by sex	NUTS 0 (2006 & 2007 / ESTAT)
● Crude birth rate	NUTS 0 (2003-2007 / ESTAT)
● Crude death rate	NUTS 0 (2003-2007 / ESTAT)
● Migration	-
● Natural Growth	NUTS 0 (2003-2007 / ESTAT)
● Active population, Unemployment	

In the respective ESPON Database datasets we have assimilated Kosovo "similar NUTS3" with the total of the country (NUTS0).

## Turkey

The **total population of Turkey** amounted in **70.586.000 inhabitants** in **2008** (Eurostat 2010)<sup>4</sup>.

### **Spatial units' levels:**

Turkey, which has adopted the EU NUTS/LAU system, has:

- **12 NUTS1 units (Regions, BÖLGELER in Turkish),**
- **26 NUTS2 units (Sub-regions, ALT BÖLGELER in Turkish) and**
- **81 NUTS3 units (Provinces, İLLER in Turkish)** –see in the **Map TU1.**

Seventy eight (78) of these last have a population greater than 50.000 inhabitants (in 2000 and beyond).

### **Notes on the national urban system**

In almost all of the 78 NUTS3 units mentioned above, there is at least one urban region with a population greater than 50.000 inhabitants. The large majority of these urban centres have more than 100.000 inhabitants.

In general terms, the capitals of the Provinces -NUTS3 units- are in almost all of the cases enough developed in order to support the development of the respective territorial units. Also, the development of the NUTS2 regions could be supported by respective existing urban centres or networks of urban centres.

We should note that Turkey has already participated in Urban Audit since 2000 with 26 cities<sup>5</sup>. Thus there are enough data for a satisfactory sample of Turkish cities.

### **Existing data at NUTS3 level (2009)**

(1) Official statistical data:

- Data at **district level:**

From 1990 and 2001 censuses and from the *2007 Population Census which used the Address Based Population Registration System:*

Population by age group and sex, Age dependency ratio, City and village population, Sex ratio, Population density.

- Data from periodic results of households Labour Force Survey for Turkey, Urban and Rural regions (results of 1988 – 1999 terms, results of 2000- October 2007, results of November 2007 and after = Address Based Population Registration System)

(2) There are data from CLC 2006

---

<sup>4</sup> According to an official estimate (Address Based Population Registration System) the country population amounted in 70.586.260 inhabitants in December 2007.

<sup>5</sup> Ankara, Adana,, Antaya, Baikesir, Bursa, Denizi, Diyarbakir, Edirne, Erzurum, Gaziantep, Hatay, Istanbul, Izmir, Kars, Kastamonu, Kayseri, Kocaei, Konya, Maatya, Manisa, Nevsehir, Samsun, Siirt, Trabzon, Van and Zongudak.

## Turkey Data included in the ESPON Database

Data provided mainly by Eurostat

• GDP (Euros, PPS)	NUTS 0 (2000 -2005 /ESTAT), NUTS1,2,3 (2000-2001 / ESTAT)
• Area	NUTS 0,1,2,3 (2000-2006 / NSO)
• Total population	NUTS 0,1,2,3 (2000-2007 / ESTAT)
• Population density	NUTS 0 (2000 -2007 /ESTAT), NUTS1,2,3 (2000-2006 / ESTAT)
• Age pyramid	NUTS 0 (2005/ESTAT)
• Pop by sex	NUTS 0 (2000 - 2007 / ESTAT)
• Crude birth rate	NUTS 0 (2000-2007 / ESTAT)
• Crude death rate	NUTS 0 (2000-2007 / ESTAT)
• Migration	
• Natural Growth	NUTS 0 (2000-2007 / ESTAT)
• Active population, Unemployment	NUTS 0,1,2 (2006 -2007 /ESTAT)

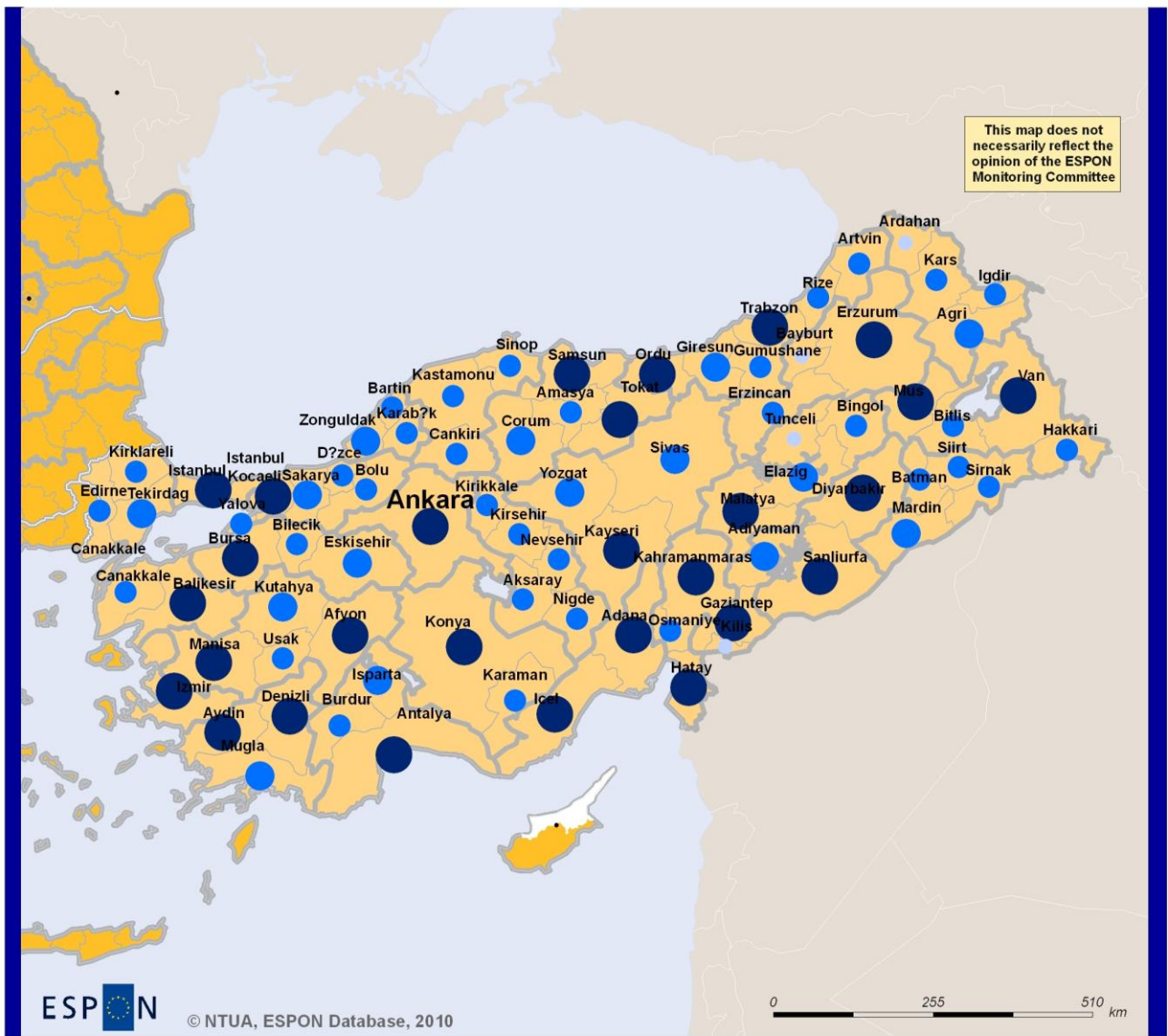
See in more detail in the respective "metadata"

We used as:

NUTS1: the 12 Regions

NUTS2: the 26 Sub-regions

NUTS3: the 81 Provinces.



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

Regional level: NUTS 3  
Source: Eurostat, 2001  
Origin of data: Eurostat, 2010  
© EuroGeographics Association for administrative boundaries

**Population per NUTS3 2001**

- 90.000 - 150000
- 150.000 - 475.000
- 475.000 - 800.000
- 800.000 - 12.000.000

**NUTS3 population thresholds:**  
minimum: 150.000 inhabitants  
maximum: 800.000 inhabitants

- NUTS2 regions
- NUTS3 regions

**Map TU.1: Turkey NUTS2 and 3 units, Population per NUTS3 2001**

### 3. Conclusions

#### Compatibility of "similar NUTS" divisions with the EU NUTS classification

Turkey, Croatia and FYROM have already adopted the NUTS classification.

For the rest WB countries, the results of the respective assessment per country, using the criteria of the population weight (formal criterion) together with the administrative capacity (informal criterion) -see for the methodology in the Introduction- ensured that the "similar NUTS" divisions used correspond almost fully with the respective divisions for the EU countries; Therefore, the "similar NUTS" could be used for the work on data without considerable problems.

The methodology developed could be further used in the control of the consistency of "similar NUTS" divisions in the Eastern Neighbouring countries (ENC) and the Southern Mediterranean Neighbouring countries (MNC).

#### Data availability at level NUTS 0

In general, it is very satisfactory for all CC / PCC; most of the data are provided by Eurostat, additional data are provided by the National Statistical Offices (NSO).

#### Data availability at NUTS2 and 3 levels

- It is in general very satisfactory for **Croatia, FYROM and Turkey**. Data are fully comparable with the EU ones as these countries have adopted the NUTS classification. Available data from Eurostat cover at NUTS2 level a wide range of topics (see in the Annexes) and at NUTS3 level: demography, economic accounts, tourism, and labour market. Some additional data for specific topics are provided by the NSO of these countries.

- It is less satisfactory for the **other Western Balkans countries**; relevant data are provided by the NSO.

*In more detail, at "similar NUTS3 level" of these countries:*

(a) For demography and labour market, it is good only for some of them while for the rest it is nearly acceptable.

(b) For the rest sections, there are important differences according to the country. Concisely, availability is more satisfactory for Serbia, much less satisfactory for the other PCC.

#### The closure of the CC / PCC "gap"

Taking into account that necessary reliable data at the appropriate NUTS level or "similar NUTS" level exist for the CC / PCC except Kosovo (under UN Security Council Resolution 1244), all *these countries should remain in the scope of the ESPON Database*; few data for Kosovo could be included at the moment in the Database.

A set of "basic" data on WB countries and Turkey is included in the ESPON 2013 Database.

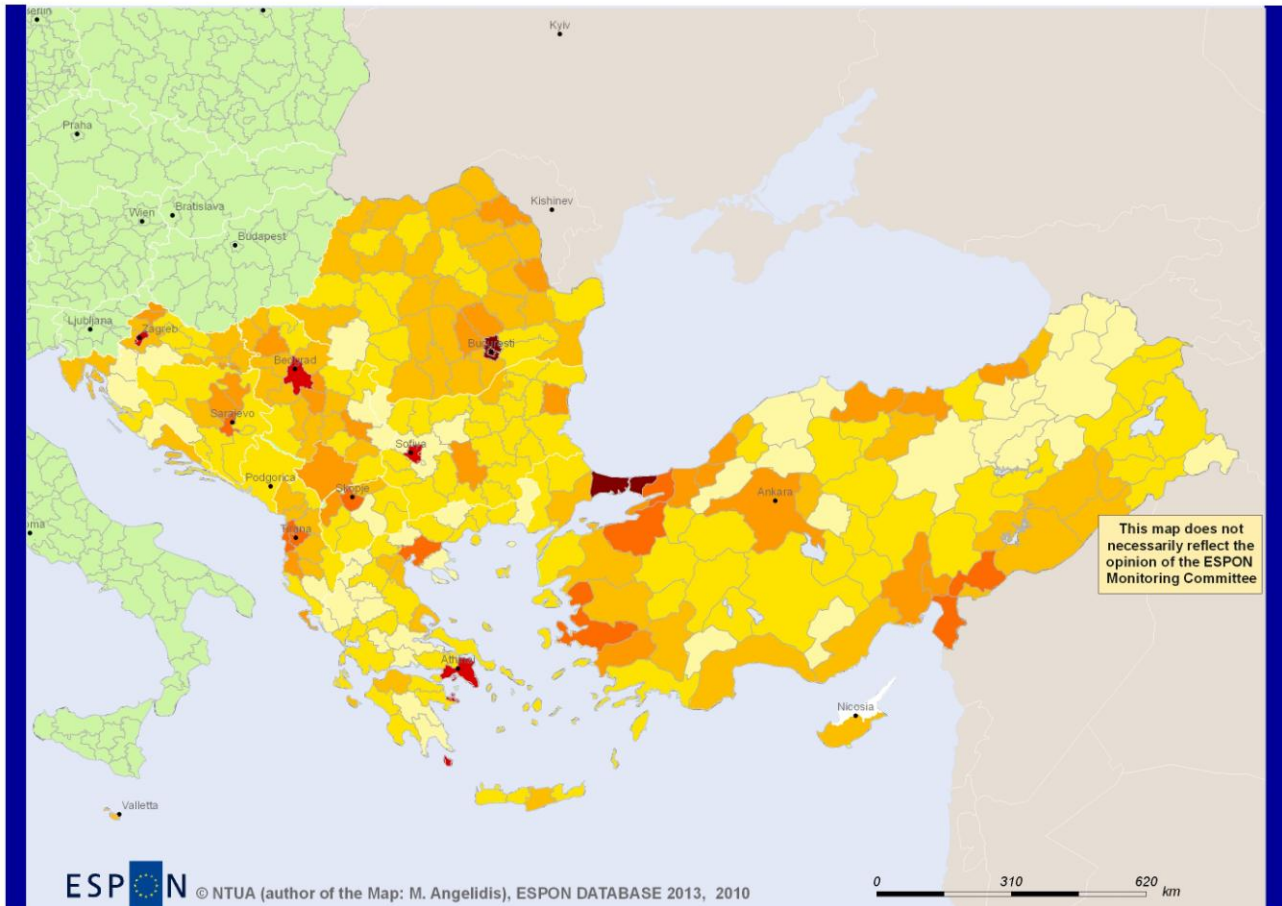
Concluding, the “gap” in the ESPON datasets and Maps corresponding to the WB countries and Turkey (CC / PCC) is closed to a considerable degree –see in the Maps Annex 1 and 2 in Annex 1 –Maps.

This step is very important because it allows the study of the territorial particularities of these countries which should be taken into account in the future Cohesion and Neighbourhood Policies of the EU.

The further development of both formal and informal collaboration of ESPON with Eurostat, DG Regio and the Statistical Institutes of the respective countries could ensure a regular bilateral flow of territorial data for these countries.

# Annex-1 Maps

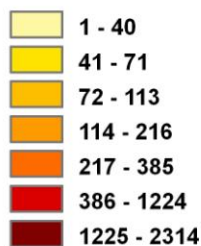
**Map Annex 1: Population density at NUTS3 level in South-eastern Europe: EU and Candidate / Potential Candidate Countries: Western Balkans and Turkey**



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

Regional level: NUTS 3, similar NUTS3  
Source: NTUA team for the elaboration of data  
Origin of data: Eurostat, National Statist. Organisations of the CC, 2010  
© EuroGeographics Association for administrative boundaries

**Population density 2006\***  
Inhabitants / Km2



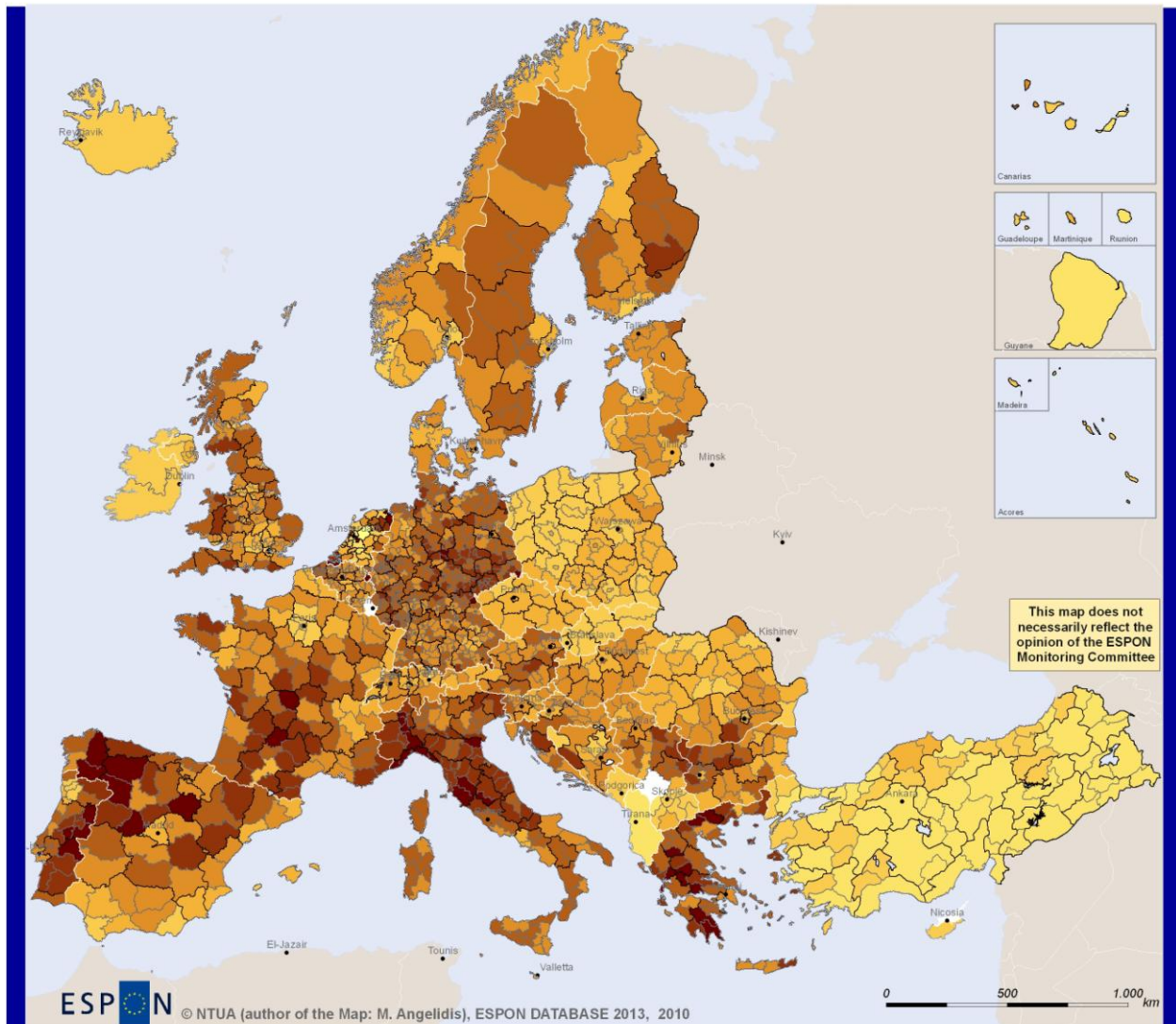
\* Data sources, reference year:

Eurostat data: density 2006  
Except: Serbia: pop. census 2002,  
Albania: pop. census 2001,  
Bosnia & Herzegovina: pop. official estimation 2007,  
Montenegro: pop. census 2003,  
Kosovo: non official estimation 2006

Geometries sources: Eurogeographics  
administrative boundaries 2006  
except: Albania, Serbia, B & H, Montenegro, Kosovo  
"similar NUTS3": other sources

**Map Annex 1: Population density at NUTS3 level in South-eastern Europe: EU and Candidate / Potential Candidate Countries: Western Balkans and Turkey**

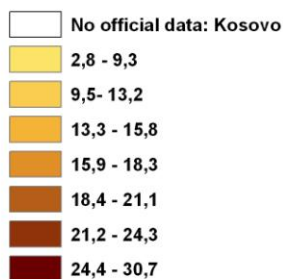
**Map Annex 2: EU and Western Balkans and Turkey\* Population 65 years and over Rate % at NUTS3 or similar NUTS3 level 2008**



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

Regional level: NUTS 3, similar NUTS3  
Source: NTUA team for the elaboration of data  
Origin of data: Eurostat, National Statist. Organisations of the CC, 2010  
© EuroGeographics Association for administrative boundaries

Population 65 years and over  
Rate % of the Total population  
at NUTS3 or similar NUTS3 level 2008\*\*



\* Candidate or Potential Candidate countries

\*\* Data sources, reference year, NUTS or similar NUTS levels:  
Eurostat data for the year 2008 at NUTS3 level (2010)  
Except:  
- Bosnia & Herzegovina: year 2008, Serbia: year 2007  
at "Similar NUTS3" level, estimations of NSO  
- Albania and Montenegro: year 2007 at country level,  
Eurostat data (estimations)  
- Germany : Eurostat NUTS3 level, year 2006  
- Netherlands: Eurostat NUTS3 level, year 2007  
- Ireland and Cyprus: Eurostat NUTS2 level, year 2005

Geometries sources: Eurogeographics administrative boundaries 2006  
except: Albania, Serbia, B & H, Montenegro, Kosovo "similar NUTS3":  
other sources

**Map Annex 2: EU and Western Balkans and Turkey Population 65 years and over Rate % at NUTS3 or similar NUTS3 level 2008**



## Annex 2 -Western Balkans and Turkey available territorial data from NSO

**Table Annex 2: Western Balkans and Turkey available (in 2009) territorial data, at NUTS3 or "similar NUTS3" level from National Statistical Organisations (NSO) (censuses, estimations etc)**

### Data, simple indicators

Data - Indicators / Country	Albania	Bosnia-Herzegov.	Croatia	FYROM	Serbia	Montenegro	Kosovo (8)	Turkey
	12 Prefectures ("counties") -"similar NUTS3"	FBiH, RS, and Brsko District (3)	Jupanija (21)	8 Statisticki Regioni / SR (4a) -"similar NUTS3"	Districts (4b) -"similar NUTS3"	Total of the country	Seven districts	81 ILLER -"similar NUTS3"
<b>(Population censuses' years 1985 – 2008)</b>	1989, 2001	1991	1991, 2001	1991, 2002	1991, 2002 (5)	1991, 2003	1991	1985, 1990, 1997 (6), 2000 (6)
<b>(Labour force surveys 1985 – 2008)</b>		2007						
<b>Demographic and social aspects</b>								
<b>Total Population</b>	1989c, 2001c	1991,1995, 2001 -2002 FBiH 2007	1991c,1995, 2001c, 2002 - 2008	1991c, 2002c	1991c, 1995, 2002c, 1998-2005a.e.	1981c,1991c, 2003c	<b>1981,1991c, 2006, 2007</b>	1990c, 2000c
<b>Population by sex: males, females</b>	2001c	1991c, 2000 - 2003, FBiH 2007	1991,1995,2001 -2008	1991c, 2002c	1991c, 1995, 1998-2007, 2002c	1981c,1991c, 2003c	1991c, 2006, 2007	1990c, 2000c
<b>Population by age group</b>	2001c	1991c, 2000- 2003 - FBiH 2007	1991c, 2001c	1991c, 2002c	1991c, 1995, 2002c, 1998-2005a.e.	1991c, 2003c	1991c, 2006, 2007	1990c, 2000c

<b>Population by sex and age group</b>	2001c	1991c	1991c, 2001c	1991c, 2002c	1991c, 1995, 1998-2005, 2002c	1991c, 2003c	1991c	1990c, 2000c
<b>Population per education level</b>	2001c	1991c	1991c, 2001c	1991c, 2002c	1991c, 1995, 1998-2005	2003-2008, 2003c	1991c	1990c, 2000c
<b>Economic aspects, Employment</b>								
<b>Total Active Population</b>	2001c	1991c, 2007 Labour force survey (lfs)	1991c, 2001c	1991c, 2002c	1991c, 1995, 2002c	1991c, 2003c	1991c	1990c, 2000c
<b>Male, Female Active Population</b>	2001c	1991c, 2007 lfs	1991c, 2001c	1991c, 2002c	1991c, 1995, 2002c	1991c, 2003c	1991c	1990c, 2000c
<b>Number of Employed persons</b>	2001c	1991c, 2007 lfs	1991c, 2001c	1991c, 2002c,	1991c, 1995, 1998-2006, 2002c, 2006 (7)	1991c, 2003c, 2004-2007	1991c	1990c, 2000c
<b>Number of unemployed persons</b>	2001c	1991c, 2007 lfs	1991c, 2001c	1991c, 2002c,	1991c, 1995, 1998-2006, 2002c, 2006(7)	1991c, 2003c, 2004-2007	1991c	2000c, 2004-2007
<b>Employment per primary, secondary, tertiary sector</b>	2001c	1991c, 2007 lfs	1991c, 2001c	1991c, 2002c,	1991c, 1995, 1998-2006, 2002c, 2006 (7)	1991c, 2003c, 2004-2007	1991c	1981-2001
<b>Gross Domestic Product (GDP) (Euros)</b>		FBiH 2005-2007		2004-2006		2000-2004	no data	1990c, 2000c

(1) a.e.=annual estimations

(2) c=census(es)

(3) Federation of Bosnia and Herzegovina (FBiH), Republic of Srpska (RS), and Brsko District

(4a) Existing results are per municipality, we can provide by aggregation results per SR

(4b) Existing results are per municipality, we can provide by aggregation results per Districts

(5) Census not carried out on the territory of Kosovo and Metohia.

(6) Turkey: 1997: Housing census only, 2000: Population census only

(7) Serbia Survey of employed per municipality 2006, we can provide by aggregation results per Districts

(8) Under UN Security Council Resolution 1244

## **Annex 3 - W. Balkans and Turkey data (2009) from Eurostat / Short presentation**

### **Albania, Bosnia and Herzegovina, FYROM, Serbia, Montenegro, Kosovo<sup>6</sup>, Turkey Data from Eurostat – 2009**

#### **(1) NUTS0 level (country), NUTS1 level**

##### **A) Key indicators on EU policy – Data for all CC – unless a different reference is made:**

1) Structural indicators: a) General Economic Background, b) Employment, c) Innovation and Research, d) Economic Reform, e) Social Cohesion, f) Environment (except Kosovo<sup>2</sup>)

##### **B) Regional statistics – Data only for Croatia, FYROM and Turkey – unless a different reference is made:**

###### 1) Regional science and technology statistics

*R&D expenditure and personnel:*

- a) Total R&D personnel by sectors of performance (employment) and region (except FYROM), b) Total intramural R&D expenditure (GERD) by sectors of performance and region (except FYROM)

*Human Resources in Science and Technology (HRST) (NUTS level 0, 1 and 2) (except FYROM):*

- a) Annual data on HRST and sub-groups, b) Annual data on HRST and sub-groups, employed, by sector of economic activity, c) Annual data on HRST and sub-groups by age, d) Annual data on HRST and sub-groups by gender

###### 2) Regional labour market statistics

*Regional economically active population - LFS series and LFS adjusted series:*

- a) Economically active population by sex and age, at NUTS level 1, (1000), b) Economically active population by sex, age and highest level of education attained, at NUTS level 1 (1000), c) Economic activity rates by sex and age, at NUTS level 1 (%), d) Economically active population by sex and age, at NUTS level 1 (1000)

*Regional employment - LFS series:*

- a) Average number of usual weekly hours of work in main job (full-time), at NUTS level 1 (hours), b) Employment by professional status, at NUTS level 1 (1000), c) Employment by full-time/part-time and sex, at NUTS level 1 (1000), d) Employment by sex, age and highest level of education attained, at NUTS level 1 (1000), e) Employment rates by sex and age, at NUTS level 1 (%), f) Employment by sex and age, at NUTS level 1 (1000)

*Regional unemployment - LFS adjusted series:*

- a) Unemployment rates by sex and age, at NUTS levels 1, 2 and 3 (%), b) Unemployment by sex and age, at NUTS level 1 (1000), c) Long-term unemployment (12 months and more), at NUTS level 1 (1000; %)

*Regional socio-demographic labour force statistics - LFS series:*

- a) Life-long learning - participation of adults aged 25-64 in education and training, at NUTS level 1 (1000), b) Population aged 15 and over by sex and age, at NUTS level 1

<sup>6</sup> Under UN Security Council Resolution 1244.

(1000), c) Population aged 15 and over by sex, age and highest level of education attained, at NUTS level 1 (1000), d) Number of households by degree of urbanisation of residence, at NUTS level 1 (1000)

**C) Economy and finance – Data for all CC – unless a different reference is made:**

1) Main Economic Indicators (except Kosovo<sup>2</sup>), 2) GDP and main aggregates, 3) Annual National Accounts – breakdowns by branches, 4) Annual National Accounts – breakdowns of final consumption expenditure, 5) Government Statistics (except Montenegro), 6) Exchange Rates and Interest Rates, 7) Monetary and other Financial Statistics, 8) Prices (except Montenegro), 9) Balance of payments

**D) Population and social conditions – Data for all CC – unless a different reference is made:**

1) Population Demography, 2) Education, 3) Labour Market, 4) Living Conditions (except Montenegro, Kosovo<sup>2</sup>)

**E) Industry, trade and services – Data for all CC – unless a different reference is made:**

1) Short-term business Statistics (except Kosovo), 2) Business demography (except Croatia, FYROM, Turkey, B n H, Montenegro, Kosovo), 3) Information Society Statistics, 4) Tourism (except Kosovo<sup>2</sup>)

**F) Agriculture, forestry and fisheries – Data for all CC – unless a different reference is made:**

1) Agriculture, 2) Forestry Statistics (except Kosovo<sup>2</sup>), 3) Fisheries (except Montenegro, Serbia, Kosovo)

**G) External trade – Data for all CC – unless a different reference is made:**

1) External Trade, 2) Trading Partners – Flows, 3) Trading Partners – Balance, 4) Trade by Commodity, 5) Terms of trade (except B n H, Montenegro, Kosovo<sup>2</sup>)

**H) Transport – Data for all CC**

**I) Environment and Energy – Data for all CC**

1) Climate change and waste, 2) Energy

**J) Science and technology – Data for all CC – unless a different reference is made:** (except Albania, B n H, Kosovo<sup>2</sup>)

## (2) NUTS2 level

**B) Regional statistics – Data only for Croatia, FYROM and Turkey – unless a different reference is made**

1) Regions

2) Regional agriculture statistics:

a) Animal populations (December) (except FYROM), b) Areas harvested, yields, production (except FYROM), c) Production of cows' milk on farms (1000 tons) (except FYROM), d) Land (except Croatia and FYROM)

3) Regional demographic statistics

*Population and area:*

- a) Population at 1st Jan. by sex and age, from 1980 to 1990, b) Population at 1st January by sex and age from 1990 onwards, c) Average population by sex and age

*Population change:*

- a) Births by age of the mother, b) Deaths by sex and age, c) Infant mortality

4) Regional economic accounts

*Gross domestic product indicators - ESA95:*

- a) Gross domestic product (GDP) at current market prices at NUTS level 2, b) Real growth rate of regional GDP at market prices - percentage change on previous year, c) Dispersion of regional GDP (%)

*Branch accounts - ESA95:*

- a) Compensation of employees at NUTS level 2, b) Gross fixed capital formation

*Household accounts:*

- a) Income of households, b) Secondary distribution of income account of households, c) Allocation of primary income account of households

5) Regional science and technology statistics

*Human Resources in Science and Technology (HRST):*

- a) Annual data on HRST and sub-groups (except FYROM)

*Employment in high technology sectors:*

- a) Annual data on employment in technology and knowledge-intensive sectors at the regional level, by gender (except FYROM)

6) Regional tourism statistics

- a) Nights spent annual data (except Turkey), b) Arrivals - annual data (except Turkey)

7) Regional labour market statistics

*Regional economically active population - LFS series and LFS adjusted series:*

- a) Economically active population by sex and age, (1000), b) Economically active population by sex, age and highest level of education attained, (1000), c) Economic activity rates by sex and age, (%), d) Economically active population by sex and age, (1000)

*Regional employment - LFS series:*

- a) Average number of usual weekly hours of work in main job (full-time), at NUTS levels 1 and 2 (hours), b) Employment by professional status, (1000), c) Employment by full-time/part-time and sex, (1000), d) Employment by sex, age and highest level of education attained, (1000), e) Employment rates by sex and age, at NUTS levels 1 and 2 (%), f) Employment by sex and age, (1000), g) Employment and commuting among NUTS level 2 regions (1000)

*Regional unemployment - LFS adjusted series:*

- a) Unemployment rates by sex and age, (%), b) Unemployment by sex and age, (1000), c) Long-term unemployment (12 months and more), (1000; %)

Regional socio-demographic labour force statistics - LFS series: a) Life-long learning - participation of adults aged 25-64 in education and training, (1000), b) Population aged 15 and over by sex and age, (1000), c) Population aged 15 and over by sex, age and highest level of education attained, (1000), d) Number of households by degree of urbanisation of residence, (1000)

### **(3) NUTS3 level**

#### **B) Regional statistics – Data only for Croatia, FYROM and Turkey – unless a different reference is made**

1) Regional demographic statistics

*Population and area:*

- a) Population density, b) Population at 1st January by sex and age from 1990 onwards, c) Annual average population by sex, d) Average population by sex and age, e) Area of the regions, f) Population at 1st January by sex and age, from 1980 to 1990

*Population change:*

- a) Births and deaths

2) Regional economic accounts

*Gross domestic product indicators - ESA95:*

- a) Dispersion of regional GDP at Nuts level 2 and 3 (%), b) Gross domestic product (GDP) at current market prices

*Branch accounts - ESA95:*

- a) Employment (in persons), b) Gross value added at basic prices

3) Regional tourism statistics

- a) Number of establishments, bedrooms and bed-places - NUTS 3 - annual data (except Turkey)

4) Regional labour market statistics

*Regional economically active population - LFS series and LFS adjusted series:*

- a) Economically active population by sex and age, (1000)

*Regional unemployment - LFS adjusted series:*

- a) Unemployment rates by sex and age, 3 (%), b) Unemployment by sex and age, (1000)

## Annex 4 – Western Balkans and Turkey data (2009) from Eurostat/ Detailed description

### Albania, B and H, FYROM, Serbia, Montenegro, Kosovo<sup>7</sup>, Turkey Data from Eurostat - 2009

#### A) Key indicators on EU policy: Structural indicators

- 1) General Economic Background: Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia, Kosovo – level NUTS 0, 1
- 2) Employment: Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia, Kosovo – level NUTS 0, 1
- 3) Innovation and Research: Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia, Kosovo – level NUTS 0, 1
- 4) Economic Reform: Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia, Kosovo – level NUTS 0, 1
- 5) Social Cohesion: Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia, Kosovo – level NUTS 0, 1
- 6) Environment: Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia (no data for: Kosovo) – level NUTS 0, 1

#### B) Regional statistics

- 1) Regions: Croatia, FYROM, Turkey (no data for: Albania, B n H, Montenegro, Serbia, Kosovo) – NUTS level 2
- 2) Regional agriculture statistics:
  - a)** Animal populations (December) (only Croatia and Turkey) NUTS level 2, **b)** Areas harvested, yields, production (only Croatia and Turkey) NUTS level 2, **c)** Land use (only Turkey) NUTS level 2, **d)** Production of cows' milk on farms (1000 tons) (only Croatia and Turkey) NUTS level 2
- 3) Regional demographic statistics
  - A) Population and area
    - a)** Population at 1st January by sex and age, from 1980 to 1990 (only Croatia, FYROM and Turkey) NUTS level 2, 3, **b)** Population at 1st January by sex and age from 1990 onwards (only Croatia, FYROM and Turkey) NUTS level 2, 3, **c)** Annual average population by sex (only Croatia, FYROM and Turkey) NUTS level 3, **d)** Average population by sex and age (only Croatia, FYROM and Turkey) NUTS level 2, 3, **e)** Area of the regions (only Croatia, FYROM and Turkey) NUTS level 3, **f)** Population density (only Croatia, FYROM and Turkey) NUTS level 3
  - B) Population change
    - a)** Births and deaths (only Croatia, FYROM and Turkey) level NUTS 3, **b)** Births by age of the mother (only Croatia, FYROM and Turkey) NUTS level 2, **c)** Deaths by sex and age (only Croatia, FYROM and Turkey) NUTS level 2, **d)** Infant mortality (only Croatia, FYROM and Turkey) NUTS level 2
- 4) Regional economic accounts
  - A) Gross domestic product indicators - ESA95
    - a)** Gross domestic product (GDP) at current market prices at NUTS level 2 (only Croatia, FYROM and Turkey), **b)** Gross domestic product (GDP) at current market prices at NUTS level 3 (only Croatia, FYROM and Turkey), **c)** Real growth rate of regional GDP at market prices at NUTS level 2 - percentage change on previous year (only Cro-

<sup>7</sup> Under UN Security Council Resolution 1244

- atia, FYROM and Turkey), **d)** Dispersion of regional GDP at Nuts level 2 and 3 (%) (only Croatia, FYROM and Turkey)
- B) Branch accounts - ESA95
- a)** Gross fixed capital formation at NUTS level 2 (only Croatia, FYROM and Turkey), **b)** Compensation of employees at NUTS level 2 (only Croatia, FYROM and Turkey), **c)** Gross value added at basic prices at NUTS level 3 (only Croatia, FYROM and Turkey), **d)** Employment (in persons) at NUTS level 3 (only Croatia, FYROM and Turkey)
- C) Household accounts - ESA95
- a)** Allocation of primary income account of households at NUTS level 2 (only Croatia, FYROM and Turkey), **b)** Secondary distribution of income account of households at NUTS level 2 (only Croatia, FYROM and Turkey), **c)** Income of households at NUTS level 2 (only Croatia, FYROM and Turkey)
- 5) Regional science and technology statistics
- A) R&D expenditure and personnel
- a)** Total intramural R&D expenditure (GERD) by sectors of performance and region (only Croatia and Turkey) NUTS level 1, **b)** Total R&D personnel by sectors of performance (employment) and region (only Croatia and Turkey) NUTS level 1
- B) Human Resources in Science and Technology (HRST)
- a)** Annual data on HRST and sub-groups (NUTS level 0, 1 and 2) (only Croatia and Turkey), **b)** Annual data on HRST and sub-groups by gender (NUTS level 0 and 1) (only Croatia and Turkey), **c)** Annual data on HRST and sub-groups by age (NUTS level 0 and 1) (only Croatia and Turkey), **d)** Annual data on HRST and sub-groups, employed, by sector of economic activity (NUTS level 0 and 1) (only Croatia and Turkey)
- C) Employment in high technology sectors (reg\_htec)
- a)** Annual data on employment in technology and knowledge-intensive sectors at the regional level, by gender (only Croatia and Turkey) NUTS level 1
- 6) Regional tourism statistics
- a)** Arrivals - NUTS 2 - annual data (only Croatia and FYROM), **b)** Nights spent - NUTS 2 - annual data (only Croatia and FYROM), **c)** Number of establishments, bedrooms and bedplaces - NUTS 3 - annual data (only Croatia and FYROM)
- 7) Regional labour market statistics
- A) Regional economically active population - LFS series and LFS adjusted series
- a)** Economically active population by sex and age, at NUTS levels 1, 2 and 3 (1000) (only Croatia, FYROM and Turkey), **b)** Economically active population by sex and age, at NUTS levels 1 and 2 (1000) (only Croatia, FYROM and Turkey)
- c)** Economic activity rates by sex and age, at NUTS levels 1 and 2 (%) (only Croatia, FYROM and Turkey), **d)** Economically active population by sex, age and highest level of education attained, at NUTS levels 1 and 2 (1000) (only Croatia, FYROM and Turkey)
- B) Regional employment - LFS series
- a)** Employment by sex and age, at NUTS levels 1 and 2 (1000) (only Croatia, FYROM and Turkey), **b)** Employment by professional status, at NUTS levels 1 and 2 (1000) (only Croatia, FYROM and Turkey), **c)** Employment by full-time/part-time and sex, at NUTS levels 1 and 2 (1000) (only Croatia, FYROM and Turkey), **d)** Employment by sex, age and highest level of education attained, at NUTS levels 1 and 2 (1000) (only Croatia, FYROM and Turkey)
- e)** Employment and commuting among NUTS level 2 regions (1000) (only Croatia, FYROM and Turkey), **f)** Employment rates by sex and age, at NUTS levels 1 and 2 (%) (only Croatia, FYROM and Turkey), **g)** Average number of usual weekly hours of work in main job (full-time), at NUTS levels 1 and 2 (hours) (only Croatia, FYROM and Turkey)
- C) Regional unemployment - LFS adjusted series
- a)** Unemployment by sex and age, at NUTS levels 1, 2 and 3 (1000) (only Croatia, FYROM and Turkey), **b)** Unemployment rates by sex and age, at NUTS levels 1, 2 and 3 (%) (only Croatia, FYROM and Turkey), **c)** Long-term unemployment (12 months and more), at NUTS levels 1 and 2 (1000; %) (only Croatia, FYROM and Turkey)
- D) Regional socio-demographic labour force statistics - LFS series



**a)** Number of households by degree of urbanisation of residence, at NUTS levels 1 and 2 (1000) (only Croatia, FYROM and Turkey), **b)** Population aged 15 and over by sex and age, at NUTS levels 1 and 2 (1000) (only Croatia, FYROM and Turkey), **c)** Population aged 15 and over by sex, age and highest level of education attained, at NUTS levels 1 and 2 (1000) (only Croatia, FYROM and Turkey), **d)** Life-long learning - participation of adults aged 25-64 in education and training, at NUTS levels 1 and 2 (1000) (only Croatia, FYROM and Turkey)

### **C) Economy and finance**

#### 1) Main Economic Indicators

Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia – NUTS level 1 (no data for: Kosovo)

#### 2) GDP and main aggregates

Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia, Kosovo – NUTS level 1

#### 3) Annual National Accounts – breakdowns by branches

Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia, Kosovo – NUTS level 1

#### 4) Annual National Accounts – breakdowns of final consumption expenditure

Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia, Kosovo – NUTS level 1

#### 5) Government Statistics

Croatia, FYROM, Turkey, Albania, B n H, Serbia, Kosovo – NUTS level 1 (no data for: Montenegro)

#### 6) Exchange Rates and Interest Rates

Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia, Kosovo – NUTS level 1

#### 7) Monetary and other Financial Statistics

Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia, Kosovo – NUTS level 1

#### 8) Prices

Croatia, FYROM, Turkey, Albania, B n H, Serbia, Kosovo – NUTS level 1 (no data for: Montenegro)

#### 9) Balance of payments

Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia, Kosovo – NUTS level 1

### **D) Population and social conditions**

#### 1) Population Demography

Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia, Kosovo – NUTS level 1

#### 2) Education

Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia, Kosovo – NUTS level 1

#### 3) Labour Market

Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia, Kosovo – NUTS level 1

#### 4) Living Conditions

Croatia, FYROM, Turkey, Albania, B n H, Serbia – NUTS level 1 (no data for: Montenegro, Kosovo)

### **E) Industry, trade and services**

#### 1) Short-term business Statistics

Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia – NUTS level 1 (no data for: Kosovo)

#### 2) Business demography

Albania, Serbia – NUTS level 1 (no data for: Croatia, FYROM, Turkey, B n H, Montenegro, Kosovo)

#### 3) Information Society Statistics

Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia, Kosovo – NUTS level 1

#### 4) Tourism

Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia – NUTS level 1 (no data for: Kosovo)

### **F) Agriculture, forestry and fisheries**

#### 1) Agriculture

Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia, Kosovo – NUTS level 1

2) Forestry Statistics

Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia – NUTS level 1 (no data for: Kosovo)

3) Fisheries

Croatia, FYROM, Turkey, Albania, B n H – NUTS level 1 (no data for: Montenegro, Serbia, Kosovo)

**G) External trade**

1) External Trade

Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia, Kosovo – NUTS level 1

2) Trading Partners – Flows

Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia, Kosovo – NUTS level 1

3) Trading Partners – Balance

Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia, Kosovo – NUTS level 1

4) Trade by Commodity

Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia, Kosovo – NUTS level 1

5) Terms of trade

Croatia, FYROM, Turkey, Albania, Serbia – NUTS level 1 (no data for: B n H, Montenegro, Kosovo)

**H) Transport**

Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia, Kosovo – NUTS level 1

**I) Environment and Energy**

1) Climate change and waste

Croatia, Turkey, Albania, Montenegro, Serbia – NUTS level 1 (no data for: FYROM, B n H, Kosovo)

2) Energy Croatia, FYROM, Turkey, Albania, B n H, Montenegro, Serbia, Kosovo – NUTS level 1

**J) Science and technology** Croatia, FYROM, Turkey, Montenegro, Serbia – NUTS level 1 (no data for: Albania, B n H, Kosovo)

## References - Data sources

### **World and EU Statistical and other data sources**

Eurostat, General and Regional Statistics / Non EU countries / *Candidate and potential candidate countries*: Regional data (for Croatia, FYROM and Turkey), other data mainly at national level.

United Nations (UN) / Statistical division (2008), *Several Tables from the UN Statistical Databases: Population and housing censuses: census dates, Population of capital cities and cities of 100,000 and more inhabitants etc.*

### **Eurostat publications, EC Regulations etc**

EC (2003) Regulation (EC) No 1059/2003 of the European Parliament and of the Council of 26 May 2003 *on the establishment of a common classification of territorial units for statistics (NUTS)* (Official Journal L 154, 21/06/2003)

EC, *Regulations (EC) No 1888/2005, No 105/2007 and No 176/2008 amending the above Regulation (EC) No 1059/2003*

Eurostat (2010), *NUTS - Nomenclature of territorial units for statistics*, [http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts\\_nomenclature/introduction](http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/introduction) as of 10.1.11.

Eurostat (2010), *Statistical regions outside the EU* [http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts\\_nomenclature/statistical\\_regions\\_outside\\_eu](http://epp.eurostat.ec.europa.eu/portal/page/portal/nuts_nomenclature/statistical_regions_outside_eu) as of 10.1.11

Eurostat / Methodologies and working papers - EC (2008), *Statistical regions for the EFTA countries and the Candidate countries*, Office for Official Publications of the EC, ISSN 1977-0375.

Eurostat / Methodologies and working papers - EC (2008), *European Regional and Urban Statistics Reference Guide*, Office for Official Publications of the EC, ISSN 1977-0375

Eurostat / Pocketbooks (2008, 2009, 2010), *Pocketbook on candidate and potential candidate countries*, Office for Official Publications of the EC.

Eurostat Leaflets, *Several leaflets on candidate and potential candidate countries (2008): economic development, population and social conditions etc*, Office for Official Publications of the EC.

Eurostat / Statistical books (2008, 2009, 2010), *Eurostat regional yearbook 2008, 2009, 2010*.

### **Official Statistical data sources for the CC**

*Several online publications on economic development, population and social conditions, dwellings, environment etc – see in detail in Chapter 2: assessment per country.*

- Albania: Albania Institute of Statistics: <http://www.instat.gov.al> .

- Bosnia and Herzegovina: Agency for statistics of Bosnia and Herzegovina: <http://www.bhas.ba>, Federation of Bosnia and Herzegovina Federal office of

- Statistics: <http://www.fzs.ba> and Republika Srpska Institute of Statistics: <http://www.rzs.rs.ba>
- Croatia: CROSTAT, Republic of Croatia – Central Bureau of Statistics: <http://www.dzs.hr/>
  - FYROM: Republic of Macedonia State Statistical Office: <http://www.stat.gov.mk/>
  - Serbia, Montenegro and Kosovo (Under UN Security Council Resolution 1244): Serbia Republic Statistical office: <http://www.statserb.sr.gov.yu/> and Serbia and Montenegro Statistical Office: <http://www.szs.sv.gov.yu/>
  - Montenegro: Statistical Office of the Republic of Montenegro – MONSTAT: <http://www.monstat.cg.yu/EngPrva.htm>
  - Kosovo (Under UN Security Council Resolution 1244): Statistical office of Kosovo: [www.ks-gov.net/ESK/](http://www.ks-gov.net/ESK/)
  - Turkey :Turkey Statistical office: <http://www.tuik.gov.tr>  
*Regional and Turkey Urban Audit statistics:*  
<http://www.tuik.gov.tr/BolgeselIstatistik/menuAction.do?dil=en>

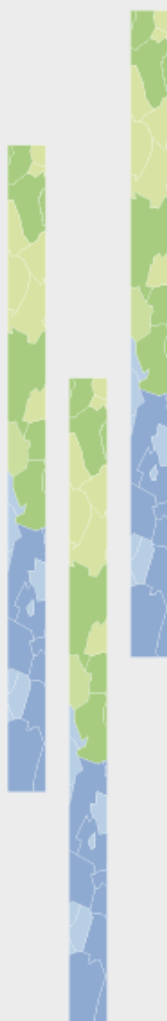
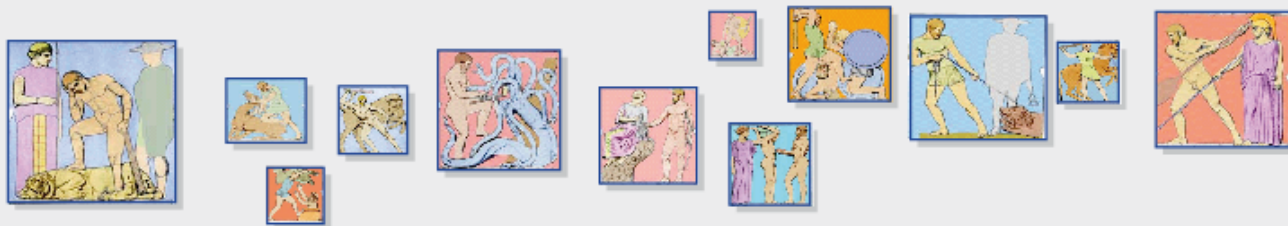
### **Other documents, research works, publications and data sources**

- Ambiente Italia Research Institute (2003): *European Common Indicators Final Project Report: Development, Refinement, Management and Evaluation of European Common Indicators Project (ECI)*, Milano
- ESPON Transnational Project Group (TPG), 2005, *project 1.1.3: Enlargement of the European Union and the wider European Perspective as regards its polycentric spatial structure*, KTH, Stockholm.
- EC (2007), *Screening report Croatia: Chapter 22 – Regional policy and coordination of structural instruments*.
- EC (2006), *EU Enlargement Strategy and Main Challenges 2006 – 2007*  
[http://ec.europa.eu/enlargement/countries/index\\_en.htm](http://ec.europa.eu/enlargement/countries/index_en.htm)
- EU / Regional Policy (2007), *State of European Cities Report: Adding value to the European Urban Audit*.
- Rakić B. – Obradović S. (2010), *Absorption problems of the EU development funds in Serbia* <http://teme.junis.ni.ac.rs/teme2-2010/teme%202-2010-07.pdf> as of 10.1.11.
- Eurogeographics, EEA, EIONET, ETCI/LUSI: *Several documents on the EU spatial information system*
- European Environment Agency / EEA (2007), *Technical report No 17/2007 CLC2006 technical guidelines*, Office for Official Publications of the EC, ISSN 1725-2237
- INSTAT and Swiss Cooperation Office Albania (2010), *Socio-demographic statistics in Albania: selected topics and future developments*, Tirana, 2010.
- Knezevic I. (2010), *Absorption Capacity of Serbia for Use of EU Funds: Practical Lessons from Slovakia*, The Pontis Foundation and the Center for Democracy Foundation, Belgrade, July 2010
- Milego R. (2007), *Report: Urban Morphological Zones 2000 version F1v0 Definition and procedural steps*, Universitat Autònoma de Barcelona / EEA.
- TPG (2004-2006), *ESTIA – SPOSE Programme: First Interim Synthetic Report (WP2) Polycentric Growth Thematic Study Final Report (WP2.2)*, National Re-

*ports of Albania, FYROM, Serbia and Montenegro etc, UEHRI / Panteion University, Athens.*

TPG (2006), *ESPON ECPs Transnational Networking activities 097/2005 Data and Indicators of Western Balkans Final Report\**, ECP Greece, Athens (\*data at NUTS0, 1 and 2 levels).

Wikipedia (2009) [www.wikipedia.org](http://www.wikipedia.org), several data on the CC: statistical territorial division, population per regions, prefectures etc, other data –until 2009.



## Local & regional data

*First investigations in Romania,  
Bulgaria, Czech Republic and  
Slovakia*

### MAIN RESULTS

- Exploring and collecting indicators at LAU 2 level is a task that must overcome three problems : the administrative changes, the chronological homogeneity of the datasets and the semantic interpretation of the indicators.
- The cumulated experience when working at LAU 1/2 level shows that populating a database becomes a learning by doing process, blocking the construction of a general algorithm.

**ESPON 2013 DATABASE**  
**MARCH 2011**



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

**26 pages**

# LIST OF AUTHORS

Octavian Groza, UAIC, TIGRIS, Iasi, Romania

Alexandru Rusu, UAIC, TIGRIS, Iasi, Romania

## **Contact**

octaviangroza@yahoo.com

tel. + 40 074 55 71 04

+ 40 232 20 14 87

# TABLE OF CONTENT

Introduction .....	3
1 The data sources - specifications .....	4
2 Choosing between 27+4 countries .....	6
3 Populating the database for Romania and Bulgaria...step by step and inch by inch.....	8
4 Building a database for the Czech Republic and Slovakia .....	12
5 Using SIRE database.....	16
6 Integrating Priority 2 projects .....	17
Conclusion .....	18
Annex : List of LAU2 indicators in Czech Republic and Slovakia .....	19
References .....	26



## Introduction

The experience accumulated during the previous ESPON contracts proves the necessity of integrating local statistical data (LAU1/LAU2 scale) in order to support more in depth analyses. Such analyses could focus on transnational thematic studies, intra-urban and urban-rural differentiations or trans-scalar approaches. Collecting and the harmonization of this data represent the mission of the TIGRIS team. After an exploring period (identifying the possible data sources, finding the appropriate structure of the database, getting familiar with the geometries or experimenting the exercise of data collection) we started to effectively collect the indicators and build the sample database for two neighboring countries.

In accordance with the proposals set out in the First Interim Report of ESPON 2013 Database Project, Tigris team had to develop a database for two neighboring countries included in the ESPON space. Dealing with this objective involved overcoming a number of problems, most of them being associated with: the harmonization of the spatial geometries, the chronological harmonization and the linguistic barriers. Also, the gap between our initial goal (to exhaustively fill in a database for two neighboring countries) and the outcome (a sample database populated with indicators available online for the Czech Republic and Slovakia at LAU1/2 scale) is mainly due to the large amount of statistical information available on the NSI web sites, that requires additional time for the processing and the integration in a coherent database. To be more explicit, the spatial information and the attribute data needed at LAU 1/2 scales is available not only on the NSI sites [e.g. the population of Slovak municipalities (LAU2) at 31.12.2008], but also from many other sources of information. Thus, building a coherent, comprehensive, comparable and functional database requires additional time and sometimes different collection methods. As a consequence, the completeness of the database was probably the first item that the TIGRIS team quit when starting the effective work.

# 1 The data sources - specifications

**The main source** of spatial information (geometries) owned for the moment by the Tigris team is the GISCO geodatabase. Two files were particularly useful: COMM\_CENS\_RG 2001 and COMM\_CENS\_2006. The two shape-files provide a base-map at LAU2 scale (polygons and center-points). As there wasn't any comparable base-map for the LAU 1 level in the mentioned database, we were determined to build up a LAU 1 map by merging the LAU2 units, according to the 2001 geometry and integrate some of the collected indicators. Using these maps was essential to our work in order to properly match the information extracted (the statistical indicators) with the available geographic coding system. However, this LAU 1 working map does not guarantee the accuracy of the resulting spatial objects, or its proper correlation with the recent extracted indicators because of the modifications in the administrative organization occurred after 2001.

**A second source** of information used in our work consists in the official lists with spatial units (LAU1/2) in each country of the European Union. The list being available on the EUROSTAT website<sup>1</sup>, it's only a matter of proper downloading in order to get an image of the administrative organization of a large part of the ESPON space. Theoretically, these lists are valid for the LAU 1/2 geometry corresponding to 2007. The quasi-chaotic evolution of this geometry at this minimal spatial scale, especially for certain countries (e.g. Romania) makes the official list proposed by EUROSTAT to be regarded with a certain dose of skepticism. Despite limitations associated with chronological inappropriateness, EUROSTAT nomenclature has been extremely useful in building the database at least for two reasons:

First, this set of lists is one of the few references which allows the appropriate integration of the LAU2 spatial frame in an hierarchically superior administrative levels (LAU2 => LAU1 => NUTS 3 => etc.). For the moment, from the perspective of indentifying the hierarchical spatial units of an LAU, a single file in the database COMM\_CENS\_2001 in GISCO equals the utility of the EUROSTAT references.

Second, the EUROSTAT classification system includes a useful coding system (national encoding, LAU labels, useful notes and remarks), which somehow permits us to connect the collected information and the indicators with the EUROSTAT references. Some countries, such as Bulgaria, are irrelevant in this respect, the coding system being very sophisticated (the LAU2 national code has its own logic; its construction does not coincide with the coding system used in the GISCO database<sup>2</sup>, although there are some "filiations" between the two systems).

---

<sup>1</sup> Finding the *bug* in the page permits also the download of information even for Bulgaria; If not, downloading Bulgaria offers Belgian information.

<sup>2</sup> According to the National Code Description included in the EUROSTAT file the "BG [Bulgarian] codes at this level consists of 5 digits. This is not a composite code. The code doesn't contain any information about the belonging of this territorial unit to any upper level of the classification. They are an inheritance from the previous Bulgarian Territorial Classification, created in the '70ies."

BULGARIAN List of LAU 1, 2 and NUTS 3, as of 01.01.2007												
NUTS level 3 - oblasti				LAU level 1 - Obshtini					LAU level 2 - Naseleni mesta			
Code	Name		ISO #	BG Code	Name		ISO #	BG Code	Name		ISO #	
NUTS3	Bulgarian	English			Bulgarian	English			Bulgarian	English		
5	BG811	VID	Видин	VIDIN	VID01	Белградчик	Belogradchik	BELOGRADČIK	03616	Белградчик	Belogradchik	BELOGRADČIK
6	BG811	VID	Видин	VIDIN	VID01	Белградчик	Belogradchik	BELOGRADČIK	05882	Боровица	Borovitsa	BOROVITSA
7	BG811	VID	Видин	VIDIN	VID01	Белградчик	Belogradchik	BELOGRADČIK	03987	Велика	Velitsa	VELITSA
8	BG811	VID	Видин	VIDIN	VID01	Белградчик	Belogradchik	BELOGRADČIK	02492	Връба	Vraba	V'RABA
9	BG811	VID	Видин	VIDIN	VID01	Белградчик	Belogradchik	BELOGRADČIK	07784	Гранитово	Granitovo	GRANITOVO
10	BG811	VID	Видин	VIDIN	VID01	Белградчик	Belogradchik	BELOGRADČIK	07765	Граничак	Granichak	GRANIČAK
11	BG811	VID	Видин	VIDIN	VID01	Белградчик	Belogradchik	BELOGRADČIK	04481	Дъбрава	Dabravka	D'BRAVKA
12	BG811	VID	Видин	VIDIN	VID01	Белградчик	Belogradchik	BELOGRADČIK	09006	Крачмир	Krachimir	KRACHIMIR
13	BG811	VID	Видин	VIDIN	VID01	Белградчик	Belogradchik	BELOGRADČIK	04517	Ошане	Oshane	OŠANE
14	BG811	VID	Видин	VIDIN	VID01	Белградчик	Belogradchik	BELOGRADČIK	08092	Праужда	Praužda	PRAUŽDA
15	BG811	VID	Видин	VIDIN	VID01	Белградчик	Belogradchik	BELOGRADČIK	05882	Проляница	Prolyanitsa	PROLYANITSA
16	BG811	VID	Видин	VIDIN	VID01	Белградчик	Belogradchik	BELOGRADČIK	01011	Рабиша	Rabiša	RABIŠA
17	BG811	VID	Видин	VIDIN	VID01	Белградчик	Belogradchik	BELOGRADČIK	02389	Райновци	Rayanovtsi	RAYANOVCI
18	BG811	VID	Видин	VIDIN	VID01	Белградчик	Belogradchik	BELOGRADČIK	05178	Салаш	Salash	SALASH
19	BG811	VID	Видин	VIDIN	VID01	Белградчик	Belogradchik	BELOGRADČIK	07400	Сливовник	Slivovnik	SLIVOVNIK
20	BG811	VID	Видин	VIDIN	VID01	Белградчик	Belogradchik	BELOGRADČIK	06865	Стакевци	Stakevtsi	STAKEVCI
21	BG811	VID	Видин	VIDIN	VID01	Белградчик	Belogradchik	BELOGRADČIK	06910	Струиндол	Struindol	STRUINDOL
22	BG811	VID	Видин	VIDIN	VID01	Белградчик	Belogradchik	BELOGRADČIK	01459	Чифлик	Chiflik	ČIFLIK
23	BG811	VID	Видин	VIDIN	VID08	Бойница	Boynitsa	BOYNITSA	08198	Бойница	Boynitsa	BOYNITSA
24	BG811	VID	Видин	VIDIN	VID08	Бойница	Boynitsa	BOYNITSA	08428	Борилонец	Borilovets	BORILOVEC
25	BG811	VID	Видин	VIDIN	VID08	Бойница	Boynitsa	BOYNITSA	07614	Градковски колиби	Gradkovski Kolibi	GRADSKOVSKI KOLIBI
26	BG811	VID	Видин	VIDIN	VID08	Бойница	Boynitsa	BOYNITSA	06049	Каниц	Kanits	KANITSA
27	BG811	VID	Видин	VIDIN	VID08	Бойница	Boynitsa	BOYNITSA	05868	Перилонец	Perilovets	PERILOVEC
28	BG811	VID	Видин	VIDIN	VID08	Бойница	Boynitsa	BOYNITSA	01039	Раброво	Rabrovo	RABROVO
29	BG811	VID	Видин	VIDIN	VID08	Бойница	Boynitsa	BOYNITSA	07183	Каловски колиби	Kalovski Kolibi	KALOVSKI KOLIBI
30	BG811	VID	Видин	VIDIN	VID08	Бойница	Boynitsa	BOYNITSA	08185	Шипкова махала	Shipkova Mahala	ŠIPKOVA MAHALA
31	BG811	VID	Видин	VIDIN	VID08	Бойница	Boynitsa	BOYNITSA	08329	Шмешци	Shishentsi	ŠIŠENCI
32	BG811	VID	Видин	VIDIN	VID06	Брегово	Bregovo	BREGOVO	02385	Балей	Baley	BALEJ
33	BG811	VID	Видин	VIDIN	VID06	Брегово	Bregovo	BREGOVO	06224	Брегово	Bregovo	BREGOVO
34	BG811	VID	Видин	VIDIN	VID06	Брегово	Bregovo	BREGOVO	02317	Връв	Vrav	VR'V
35	BG811	VID	Видин	VIDIN	VID06	Брегово	Bregovo	BREGOVO	08904	Гъзово	Gamovo	G'MIZOVO
36	BG811	VID	Видин	VIDIN	VID06	Брегово	Bregovo	BREGOVO	02068	Делейна	Delejna	DELEJNA

Figure 1: Bulgarian list of LAU 1/2 spatial units with labels in Bulgarian and English (source : EUROSTAT)

A third source of information used for the database construction is represented by the NSI websites. Obviously, the information collected from these references is not homogenous/ unequal as presentation system<sup>3</sup> (structuring, organization manner), as time-series included, as semantic relative to the indicators or as spatial dimension.

As a **PRELIMINARY CONCLUSION**: the chronological heterogeneity of our information sources constantly forced our approach to situate itself on some uncertain coordinates, dictated not only by the lack of accuracy linked with the geometries, but also by our direct interference with the inner structures of the files collected, due to some technical impossibilities related to the spatial variety of the extracted indicators.

<sup>3</sup> The file format used by the NIS sites represents one major drawback during the collection period of indicators. Some NIS (like the Slovakian one) offers free information for LAU 1 spatial units via downloadable software (AXIS), a kind of spreadsheet format which doubles the working time. The Slovak LAU 2 indicators are even more difficult to harvest because they are presented unit by unit, in *html* format (probably). The Czech Republic NIS site offers the information in *.xls* format, facilitating the collection at LAU2 scale. However, The Czech Republic NIS offers no information at LAU 1 scale.

## 2 Choosing between 27+4 countries

The selection of the countries included in our analysis was based on several criteria. First, we preferred from the start that the two countries to be located in the eastern part of the ESPON space, starting from the premise that the data collection, due the unequal experience<sup>4</sup> and the numerous readjustments imposed by the transition period could be somewhat more difficult here than in some Western states, which already managed to perfect their statistical systems, thus making it an useful experience and an easy to extrapolate one. In the meanwhile, we had to keep in mind the fact that the main difficulty in the process of extracting statistical indicators (especially in terms of chronological dimension), is linked with the search for an equilibrium between the length of the time series and the number of spatial units involved. That's why we have privileged two medium-size countries, honestly much more suitable for the statistical data collection. In the beginning of our work we have focused on Romania and Bulgaria and the rationale seemed quite logic to us.

First, the Tigris team has some experience in dealing with LAU2 databases for the two countries (e.g. Espace géographique, etc). Moreover, we have already completed a sample database for Romania and Bulgaria using LAU1 and LAU2 indicators, collected in 2007 and 2008 and some of these indicators were already chronologically harmonized. This experience is reinforced by the know-how accumulated during the elaboration of the several versions of the Atlas of Romania (the version available online is basically a LAU2 cartographic tool). All this work already undertaken for the two countries helped us in building a large and quite comprehensive database (several hundreds of indicators only for Romania) for the 2948 or 3175 LAU2 officially designated in this moment. However, this database is relatively old because of the successive administrative "micro-reforms" who multiplied the number of spatial units from 2948 in 2002 to almost 3175 in 2008. Most often, these readjustments in the elementary geometry were produced by the division of LAU2, by administrative redefinition (some rural LAU change status in urban ones) or by the modification of the existing nomenclature.

---

<sup>4</sup> We had the nice surprise to observe that the Eastern NIS sites are generally comparable with the western ones and sometime extremely innovating in their data layout or in the process of indicator selection.

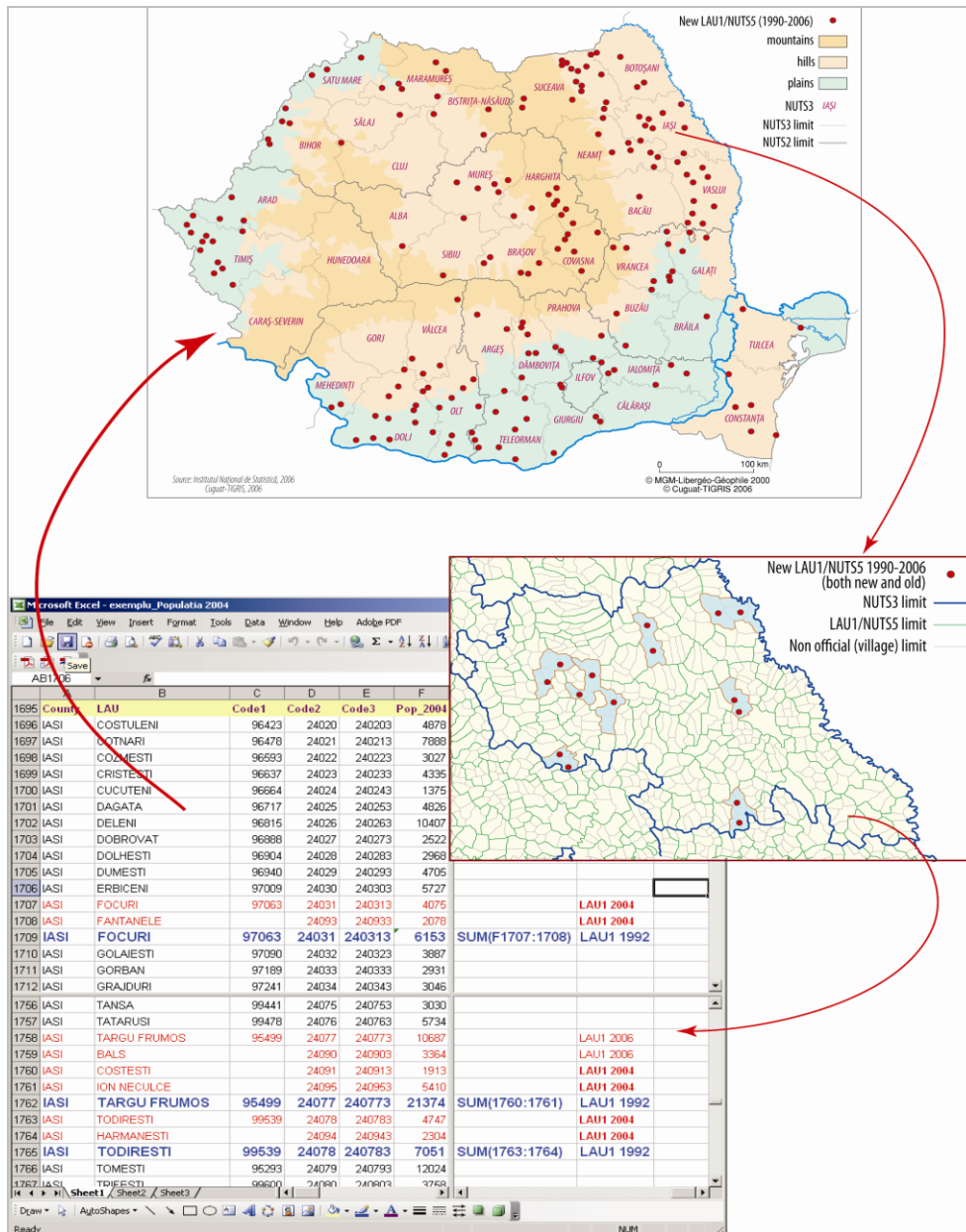


Figure 2: New LAU2<sup>5</sup> units in Romania (1990-2006)

<sup>5</sup> In the titles of the two maps one should read New Lau2/NUTS5

### 3 Populating the database for Romania and Bulgaria...step by step and inch by inch

At a normal screen resolution, the 4618 Bulgarian LAU 2 spatial units extracted from the GISCO database (COMM\_CENS\_RG\_2001) represent the equivalent of approximately 23 meters of information for only one field in the working file. The 2940 Romanian LAU2 should occupy another approximately 15 meters of information. It might seem somehow anecdotic and irrelevant information but, basically, populating a database means introducing meters of information for every indicator. One could imagine that this process is an automatic one, an easy job for post-modern geographers. Is not quite like that. Populating the database also signifies an endless verification process in order to properly match fields of information extracted from the online sources with the working files to be filled in. This matching issue represented the most time consuming aspect in the working process. However, it was also the simplest intellectual challenge in our approach.

After collecting the data from the GISCO tables and directories we have observed several inadequacies between the list of LAU 1/2 registered in this database and the lists provided by other sources (National Institutes of Statistics, TIGRIS dabase, EUROSTAT), both for Romania and Bulgaria. Bulgaria is probably the most interesting challenge in terms of rebuilding the administrative history at minimal spatial scale.

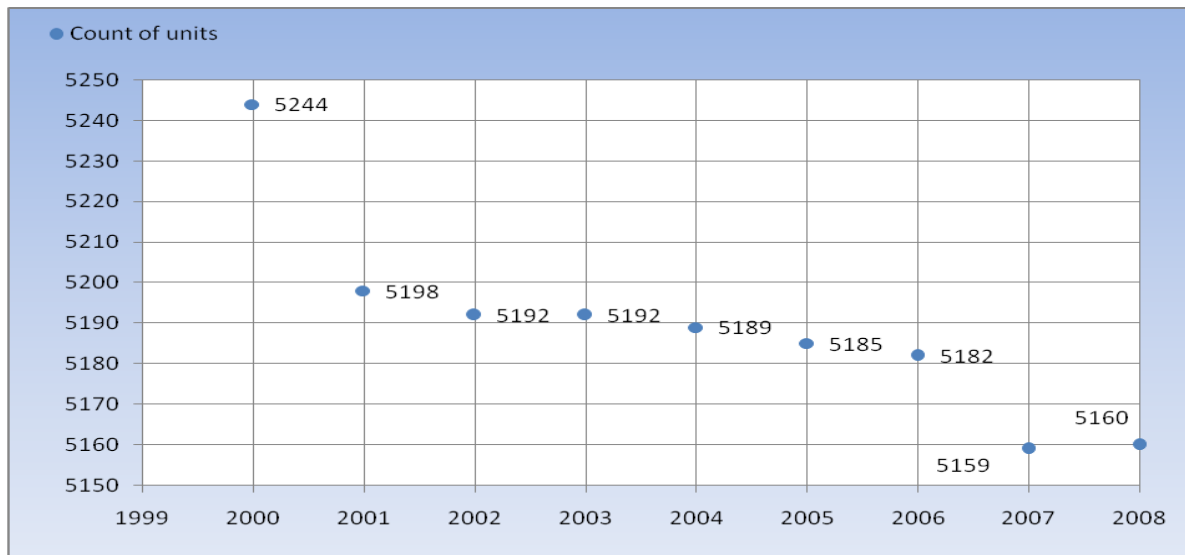
TYPES OF MODIFICATIONS	
observed for the first time	change in the list of composite units
creation	closure
creation by separation (from another populated place)	closure by new administrative-territorial structure
creation by merging	closure by merging
creation by division	closure by division
creation by new administrative-territorial structure	closure by addition
annexation to the country territory	erosion
change by new administrative-territorial structure	closure by loss of territory
change of name	restoration
change of characteristic	restoration by merging
change of administrative centre	restoration by merging
change of administrative territorial belonging	restoration by separation
separation	restoration by division
addition	change of boundaries/structure

**Table 1:** Classification of LAU 2 modifications in the administrative geometry (events recorded since 1878)

Source: NSI Bulgaria, NATIONAL REGISTER OF POPULATED PLACES

Although the Bulgarian Register of Populated Places is extremely generous in terms of information regarding the changes in the administrative geometry of the LAU2, all these references require a systematic approach which is an extremely time-consuming task. For example, 30 units of type 2 LAU were closed by addition after 2001 (and the addition term deserves a definition which was not yet found), another 3 were closed

by merging, 4 villages (LAU2 units) changed their name, one town was restored by separation, one village was restored by merging, 6 villages were created by addition, 6 new units were created by separation and unfortunately this list of modifications is not exhausted. All these territorial metamorphosis have a direct impact on the database that we are supposed to provide.



**Figure 3:** The evolution of LAU2 in Bulgaria (x axis = time)

According to this official source, the number of LAU 2 in Bulgaria constantly dropped from more than 5200 spatial units in 2000 to 5160 in 2008. However, a big number of units listed in the National Register of Populated Places was not found either in the GISCO/EUROGEOGRAPHICS reference files (599 LAU2 missing for 2001) or in the official LAU2 list presented by EUROSTAT<sup>6</sup>.

Consequently, linking geometry and database tables is impossible for the moment.

In this case, even if we have succeeded to extract one indicator for Bulgaria at this scale of analysis (population for the LAU 2 polygons between 2000 and 2008) in the absence of a proper base-map, the table is unlikely to be useful.

Similar problems have been encountered for Romania. GISCO/EuroGeographics database lists 2940 LAU2 in 2001. Comparing this source with the TIGRIS database or with some official sources (National Institute of Statistics) we found 8 LAU2 missing. If in Bulgaria<sup>7</sup> the trend shows a constant decrease in the number of LAU 2 units, in Romania the situation represents exactly the opposite. TIGRIS had several attempts to rebuild the elementary base-map of Romania but without success due to the fact that new changes in the administrative geometry are occurring almost every month. As a matter of fact, the problem is much more complicated. The decision to create a new Romanian LAU2 (generally by division) is not immediately doubled by a map with the new limits of the new born polygons. Even if we succeed to provide an updated base-map for Romania, we are not quite sure about the accuracy of these polygons.

<sup>6</sup> The two sources offer a different number of spatial units for 2007 (EUROSTAT – 5299 and NSI BULGARIA 5159). Almost 150 spatial units are to be found in list of modifications only for this chronological reference.

<sup>7</sup> According to the Bulgarian National Register of Populated Places.

The following tables synthesize the main steps and problems encountered in the development of the database for the two countries. Despite several attempts, for the moment not every problem is also accompanied by a solution.

STEP	Operation	Source	Done
1	Extracting basemap for Bulgaria (LAU2)	GISCO COMM_CENS_2001_AT	OK
2	Extracting basemap for Romania (LAU2)	GISCO COMM_CENS_2001_AT	OK
3	Merging LAU2 polygons in LAU 1 (only for Bulgaria)	GISCO COMM_CENS_2001_AT	OK
4	Creating basemap with the two countries	GISCO COMM_CENS_2001_AT	OK
5	Extracting indicators from the GISCO database	GISCO COMM_CENS_2001_AT	OK
6	Comparing LAU2 GISCO codes with other coding systems (SIRUTA for Romania and the Bulgarian NSI codes)	GISCO COMM_CENS_2001_AT, TIGRIS database, NSI databases	OK
7	Dealing with the encountered problems		OK
8	Populating the database with indicators for both countries	GISCO COMM_CENS_2001_AT, TIGRIS database, NSI databases	OK

**Table 2:** Operational steps undertaken during the database development process and data sources

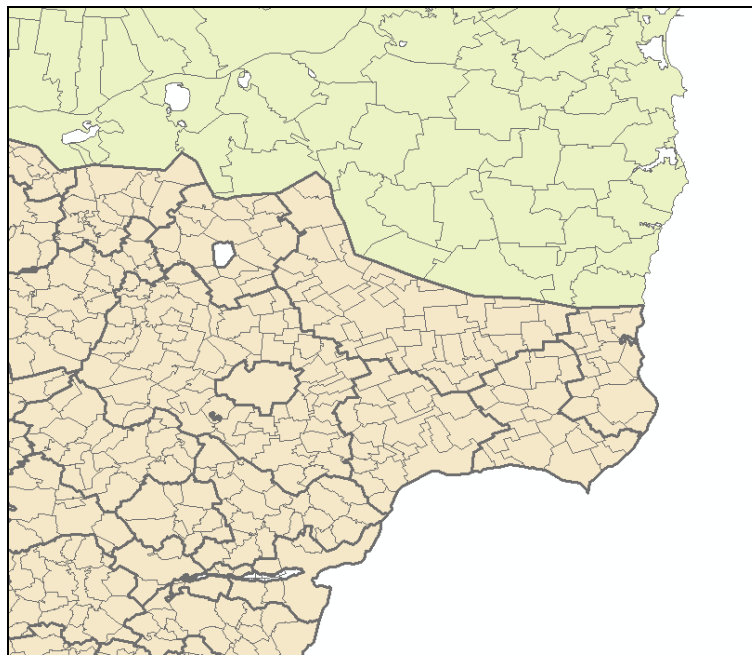
PROBLEMS	Solution
No match between the LAU2 GISCO coding system and the NSI coding system.	Inventing a new coding system.
No LAU1 label in the GISCO database (for Bulgaria) and no match between the LAU1 GISCO coding system and the NSI coding system.	Matching under Excel the labels and the codes
No match between the LAU2 geometry (GISCO) and the indicators extracted from the other sources. (599 LAU2 missing in 2001) No match between the LAU 1 geometry (GISCO) and the indicators extracted from the other sources.	Operation aborted for the moment

**Table 2:** Problems encountered in the database development process and solutions developed

Even if the issues concerning the proper linkage between the base-map and the database should be overcome, it will still be difficult to imagine a solution in order to eliminate the size differences between the LAU2 of the two countries.

The Bulgarian LAU1 has no correspondent in Romania while the Romanian LAU 2 is much bigger than the same units in Bulgaria (Fig. 4). When mapping whatever indicator, this "mass effect" should be considered. We are sure that we will encounter the same problem (linked with the surface difference) at the French-Belgian border.





**Figure no 4 – LAU2 in Romania and Bulgaria (LAU1)**

For all these reasons we have stopped working for the moment at a database for Romania and Bulgaria (technically speaking we are in standby with the history of LAU2 evolution for the both countries), even if we have somehow advanced in this problematic, and as a backup for the technical rapport and for Challenge 4, we have focused on building a database for other two countries (Czech Republic and Slovak Republic).

## 4 Building a database for the Czech Republic and Slovakia

The choice of the two countries was based on some facilities that have smoothed the collection of information and the matching exercise with the base-map extracted from the GISCO database. First, unlike Romania and Bulgaria, quite a few administrative reforms have altered the administrative geometry of LAU2 and LAU1 during the 2001-2008 period. Such mutations, but not so intense like in the Bulgarian case, are visible in Slovakia. For now, only 8 Slovak LAU2 don't find their correspondent in the GISCO tables which we use to verify the correspondence between the base-map and the database. The collection of the indicators started from the National Institutes of Statistics, in particular the 2001 Census results for the Czech Republic and the Regional database for Slovak Republic.

Despite our intention, we are not able to provide an exhaustive database for the two countries yet. In the case of the Slovak Republic, the information available at LAU 1 exceeds our possibility to collect them just in time. Anyway, a prioritization of the indicators should be considered for a proper extraction, otherwise we might populate the database with interesting but not very useful<sup>8</sup> indicators.

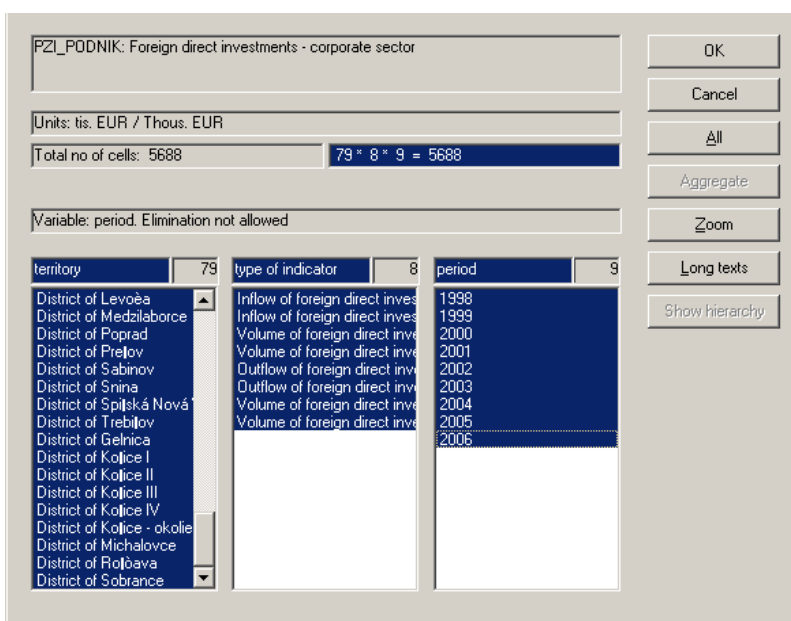


**Figure 5:** The availability of statistical indicators in the case of the Slovak Republic - Sample view

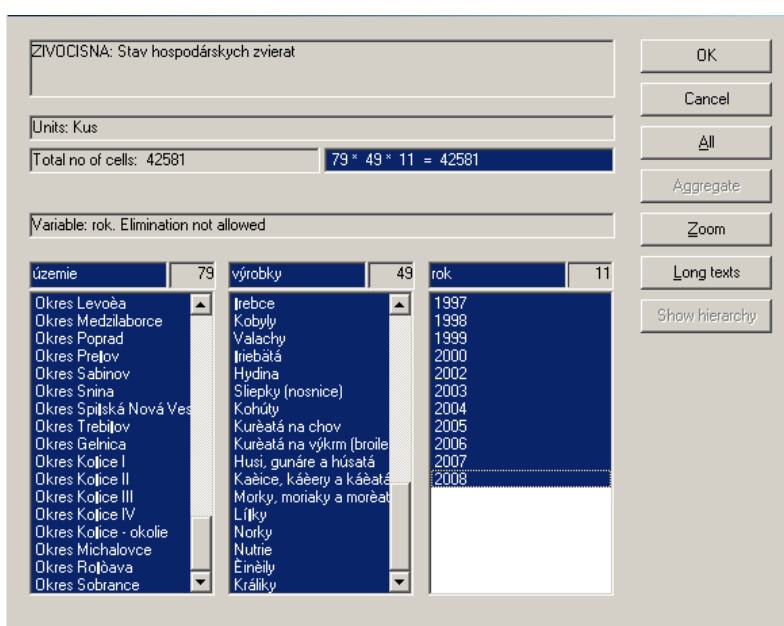
(Source: RegDat, The Regional Statistics Database hosted on the Statistical Office of The Slovak Republic website)

<sup>8</sup> As an example, we can download indicators such as the "pension's expenditures in Euro or Slovak currency between 1999 and 2008", for the Slovakian LAU1, but we cannot find the same information (the same indicator) for the Czech Republic. At a smaller scale, for the Slovak LAU2 we may download the earliest recorded mention by historical sources (e.g. Borinka (LAU2) in the District of Malacky (LAU1) was first mentioned in 1273 A.D. An exhaustive collection of the Slovakian indicators should provide even the administrative or economic central places attributes for the Slovak Lau2.

In order to integrate all this information in our data tables we were forced, (especially when collecting the indicators at LAU1 scale for Slovakia) to work with another software (Pc-Axis) allowing the visualization of the chosen variables (Fig. 6 and 7). Just to emphasize the immense data series and the sometimes overwhelming work involved: eight indicators for nine years time-series and 79 spatial units could be regarded as quite a simple case...but not as simple as downloading the agriculture indicators (Fig. 7). On the other hand, the collection of indicators for Czech Republic at LAU1 scale is not simple at all. The site of the Czech Institute of Statistics still uses the term NUTS4 instead of LAU1. Our first researches ignored this aspect. Consequently, we are not able to provide indicators for this type of administrative geometry for this country. Recently, after a routine check of the data sources, we have managed to obtain some LAU1 indicators (some demographic time-series from 1949 to 2007) and these tables will soon be ordered and integrated in the database.



**Figure 6:** Foreign direct investments in Slovakia (LAU1 – 1998-2006, Pc-Axis software view)



**Figure 7:** Stav hospodárskych zvierat by územie, výrobky and rok

(Pure Slovakian... It seems to be a file which presents indicators about the livestock according to the Google translate tool – "Status of livestock by the products and the year")

The data collection at LAU2 scale for Slovakia is also uncompleted. We have managed to include in our database 52 indicators, LAU2 by LAU2, after a long copy-paste/import data exercise LAU2 by LAU2 files (2928 multiplied by 2 files copied for each spatial unit). The 52 indicators include different information which we considered relevant at the extraction moment (economic and demographic indicators for 2007 and 2001). Generally, the other variables available for download on the site (Health Services, The Basic Characteristics of the LAU2 or the Environment Indicators) are mainly presented in text format (Boolean opposition of presence/absence). Working on Slovakia LAU2 and LAU1 indicators is a useful exercise, a training routine for the collection of information for Poland and Austria.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
NATIONAL_CODE_LAU2	LABEL_LAU2	Total population (as of Dec 31)	Population - males (as of Dec 31)	Population - females (as of Dec 31)	Population in pre-productive age - total (0 - 14)	Population in productive age - females (15 - 54)	Population in productive age - males (15 - 59)	Population in post-productive age - total (55+F, 60+M)	Number of marriages	Total increase (decrease) of population - females	Population in total	Population - males	Population - females	Population by nationality	Slovak %		
528595	Bratislava - mestská časť Staré Mesto	41255	19204	22051	4542	11423	12803	12487	29	-303	44 798	20 552	24 246		90,01		
529311	Bratislava - mestská časť Podunajské Biskupice	20717	9838	10879	2808	6294	6911	4704	14	200	19 749	9 403	10 346		82,09		
529320	Bratislava - mestská časť Ružinov	70692	31769	38923	8681	20428	20629	20954	29	205	70 004	31 439	38 565		91,65		
529338	Bratislava - mestská časť Vrakuňa	19320	9171	10149	2368	6619	6896	3439	79	70	18 366	8 786	9 600		88,24		
529346	Bratislava - mestská časť Nové Mesto	37048	18901	20147	4443	10752	11265	10589	18	-24	37 418	18 331	20 487		92,17		
529354	Bratislava - mestská časť Rača	20438	9823	10615	2352	5956	6673	5457	14	-24	20 472	9 541	10 631		93,16		
529362	Bratislava - mestská časť Vajnoch	4659	2331	2328	606	1392	1654	1007	22	86	3 826	1 899	1 929		95,69		
529401	Bratislava - mestská časť Devín	1040	527	513	150	286	355	249	85	4	884	441	443		91,97		
529371	Bratislava - mestská časť Devínska Nová Ves	15948	7791	8157	2077	5875	6094	1912	86	65	15 502	7 509	7 993		93,45		
529389	Bratislava - mestská časť Dúbravka	34405	16127	18278	4247	9697	10562	9899	23	-137	35 199	16 498	18 701		92,72		
529397	Bratislava - mestská časť Karlova Ves	33876	15893	17983	5109	11146	10979	6642	20	73	32 843	15 507	17 336		92,39		
529419	Bratislava - mestská časť Lamač	6580	2974	3606	690	1728	1938	2224	56	17	6 544	2 921	3 623		93,87		
529427	Bratislava - mestská časť Záhorská Bystrica	2852	1384	1468	411	854	941	646	21	95	2 086	1 003	1 083		96,93		
529435	Bratislava - mestská časť Čunovo	936	501	435	126	252	354	204	26	-1	911	462	449		88,83		
529443	Bratislava - mestská časť Jarovce	1296	628	668	164	412	452	268	46	20	1 199	575	624		63,8		
529460	Bratislava - mestská časť Petržalka	113443	54198	59245	11526	41514	43542	16861	60	-364	117 227	56 116	61 111		92,64		
529494	Bratislava - mestská časť Rusovce	2422	1189	1233	361	776	837	448	26	71	1 922	958	964		76,27		
507831	Borinka	557	275	282	78	148	177	154	4	6	519	252	267		95,57		
507890	Gajary	2894	1400	1494	464	878	1001	551	15	19	2 690	1 311	1 379		96,91		
507954	Jablonové	1112	539	573	170	334	369	239	7	16	1 056	510	546		97,63		
507962	Jakubov	1466	734	732	242	440	518	266	4	21	1 312	656	656		95,2		
508012	Kostolište	1132	568	564	178	340	402	212	6	30	942	476	466		98,2		
508021	Kuchyňa	1691	841	850	255	489	609	338	6	18	1 597	791	806		98,18		
543951	Vojčica	2099	1015	1084	404	615	683	397	9	-2	2 021	994	1 027		96,68		
543969	Vojka	513	258	255	125	139	170	79	5	2	434	216	218		11,98		
543977	Zatín	789	385	404	125	213	249	202	3	9	788	382	406		13,83		
543985	Zbehnov	292	149	143	47	79	106	60	0	-3	292	141	151		71,23		
543993	Zemplin	390	193	197	71	110	138	71	1	-6	399	197	202		28,32		
544001	Zemplínska Nová Ves	960	464	496	194	265	304	197	4	-5	938	438	500		98,19		
544019	Zemplínska Teplica	1500	733	767	349	452	475	224	11	4	1 384	676	708		92,7		
544027	Zemplínske Hradište	1124	524	600	140	324	346	314	8	-4	1 201	564	637		93,01		
544035	Zemplínske Jastrabie	640	327	313	109	152	226	153	4	2	643	319	324		98,44		
544043	Zemplínsky Brnč	479	231	248	93	137	156	93	2	-1	443	217	226		90,29		
2931	Indicators for 2007																
2932	Indicators collected from the Slovak Census 2001																

Figure 8: A "working file" for the Slovak LAU 2 database

The matching process between the NSI tables and the coding system used in the GISCO files for the Slovak LAU2 geometry shows that 8 new LAU2 are to be integrated in the map. These 8 LAU2 present no information recorded from the Slovak Census but they do present some indicators for 2007.

On the other hand, for the Czech Republic we have extracted 149 indicators covering a larger field of domains (from demographics to dwelling stocks and economics, table 4). As the tables and the base-maps extracted from GISCO/EuroGeographics database are chronologically correlated with the Czech Census and because no Czech LAU2 is missing for the moment, populating the database was not as complicated as was the case for Bulgaria.

Types of indicators	
1. Population by age and marital status	8. Commuters to work and schools
2. Population by age	9. Households by type
3. Population by highest educational attainment	10. Housing stock
4. Population by nationality	11. Houses by the floor number and by basic amenities
5. Population by denomination	12. Dwelling stock
6. Population by economic activity	13. Permanently occupied dwellings by legal reason of use and size of dwelling
7. Economically active population by branch of economic activity	

**Table 4:** Categories of statistical indicators for the Czech Republic

A major advantage observed during the population of the database consists in the fact that a most of the data for the Czech Republic and Slovakia comes from the Census conducted in 2001. Thus the main indicators are at least chronologically harmonized. Unfortunately, these indicators are not also semantically linked, except for the ethnic and confessional structure of the population, for the number of dwellings and for some economic variables.

Thus, as a **PRELIMINARY CONCLUSION:** the Tigris team has succeeded in creating two sample databases for 4 countries (Romania, Bulgaria, Czech Republic and Slovakia). The issue we are working to overcome now is that the indicators are not complete or harmonized yet. Several types of problems were identified, some of them having simple and/or no time consuming solutions, while some others might need a supplementary time for a more advanced analysis in order to provide effective solutions and implement them.

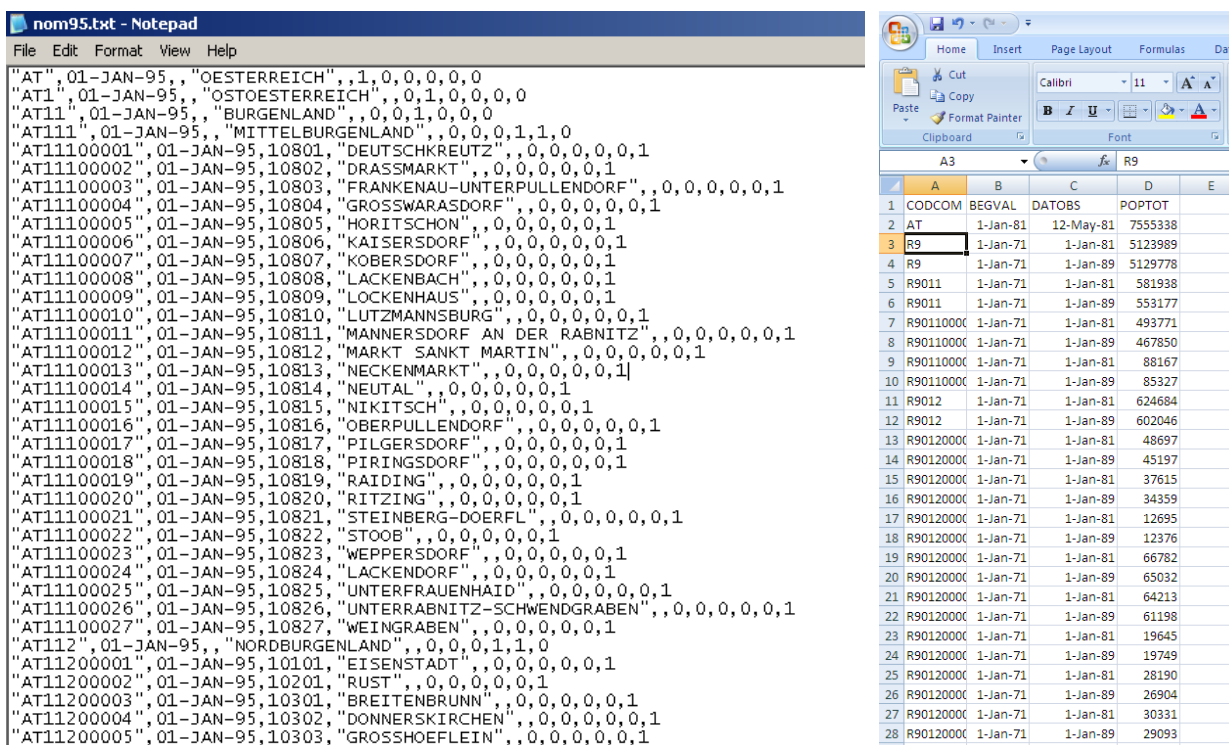
Although the focus remained on the construction of a sample database for the two countries, a part of the team has managed to gather indicators at LAU2 scale for three Scandinavian countries (Norway, Sweden and Finland) and for 2 Baltic states (Latvia and Lithuania). These 5 sets of new indicators (generally demographic and social ones) will be processed in the incoming period, in order for them to properly match with a base-map or with other data tables.

As a conclusive summary of our work, the statistical indicators collected and integrated in the database for the Czech Republic and Slovakia are presented in the table in annex. Some of the indicators are constantly repeated<sup>9</sup> (e.g. the LAU2 coding system and their names) in order to facilitate the preliminary extraction, when needed.

<sup>9</sup> This is the case only for the indicators collected for the Czech Republic.

## 5 Using SIRE database

Another element that TIGRIS team has to deal it is the recovery and transformation of indicators from SIRE database. Having a particular structure (an obsolete coding system and a spatial hierarchical structure starting from NUTS 0 to ex-NUTS 5, in the same field) the integration of information implies acquiring a specific method. The main chronological marks in SIRE are 1981 and 1991. Obviously, not all the countries in the ESPON space are present in the database and one could think that an interesting and complete exploitation of SIRE should be doubled by an investigation of datasets for recently integrated in the EU.



**Figure 9:** SIRE database before (on the left – labels and codes) and after (on the right – population in 1991) data basic integration.

The output of working with SIRE indicators is multiple. It serves for comparison between the coding systems and labels, in order to survey administrative modifications at LAU scale and it's also useful for building some chronologically based indicators between 1991 and 2001, when used in linkage with other databases.

## 6 Integrating Priority 2 projects

The integration of data obtained in Priority 2 projects represents a priority in TIGRIS work. That's why one of the deliverables was conceived as a container for this kind of information. However, a prioritization of the indicators, based on an analysis of the added value of these new indicators should also be considered as a task. If the information obtained by Priority 2 projects is too recent (2007 or 2008) it may complicate the integration when SOME not spotted administrative changes in geometry occur. A secondary problem could be linked with the eventual cartographic expression of this new information. If two finisterre are to be mapped, a proper projection will highly smooth the visual transmission.

In the next stage, the efforts of the TIGRIS team will be canalized on perfecting the database for the two countries (integrating some recent demographic indicators for Slovakia at LAU2 scale, (re)structuring/refining some data tables at LAU1 scale for the Czech Republic), on sketching a minimum administrative history for Romania and Bulgaria, finalizing the data collection for some other countries in the ESPON space.

One of the issues we are dealing with at the moment is the data validation and the elimination of the possible errors inherently occurring during the data collection and structuring process. Only after we are going to develop a system for data validation, we are going to be able to attach the metadata to our files.

For the moment, our priority still remains that of creating a proper connection between the indicators and the geometry, which could sometimes be problematic (as our experience when working for the Romania and Bulgaria database proved it).

Organizing a working plan in this context seems to depend on variables that are partially controllable by TIGRIS. In the short term our effort will focus on the elaboration of a database with indicators for at least two neighboring countries. For the midterm (December 2099) finalizing a database with indicators at LAU level would be the main priority. In the same time we shall derive a minimal history of LAU1/2 modifications. For February 2010 we had reserved the most time consuming task – recovering SIRE while populating a country by country database with a basic indicator at least.

## Conclusion

Gemeinden, Inn, Municipios, Obcine, Comune, Communes, Freguesias, Telepulesek, Ward, is the label for mostly the same geographic reality, the local level of administrative units in some countries of the ESPON space. Exploring them and collecting their basic information is a feasible and necessary task. Dealing with this task means to properly estimate the right balance between the errors in the spatial geometries, the chronological availability, the administrative changes and the sens of words behind the indicators.

The exploration of the available sources of information at LAU 2 scale (NSI, GISCO, SIRE, etc.) shows that building a database for this territorial level should overcome 3 different issues, in order to become a coherent tool. The first issue refers to the chronological heterogeneity of the indicators. Analyzing these indicators country by country, it's quite a luck to find a proper chronological match between them. This problem is underpinned with the second one, the issue of the administrative changes at local level, this last aspect heavily complicating any database populating process. The administrative changes block the construction of a general algorithm (for more than 119 000 LAU2 in ESPON space), especially when intermediate levels of territorial clip are present – the LAU 1 level. Thirdly, the semantic issue of the indicators could also become important. According to country's definition, dwelling or *others* (religion minorities, e.g.) might not have the same sens from Greece to Iceland.

However, despite TIGRIS experience, working at this scale it's learning by doing process, even if doing is pretty fuzzy in this context. The example of the database built for countries such as Slovak Republic, Czech Republic, Bulgaria or Romania shows that another aspect should be taken into account – the relevance of the indicators. The added value of different variables present in the datasets and available for extraction should be prioritize, having in mind the fact that they may largely vary because of the 3 issues already exposed.

When we try to integrate databases such as SIRE in a LAU 2 actual frame we should double the working process by an investigation of the statistical sources available for the '80 and '90 period for some countries recently integrated in the ESPON space. If not, we might obtain a proper image of the past without any link to its future. A comparable problem emerges when we integrate data form the Priority 2 projects. This time, it's not the chronological frame that worries, but the spatial one.



# Annex

## List of LAU2 indicators for the Czech Republic and Slovakia

INDICATORS	DESCRIPTION	SOURCE AND OBSERVATIONS
Iden	Basemap code	GISCO database
OBJECTID	Inner code used in ARCVIEW	GISCO database
COMM_ID	Basemap code	GISCO database
X	Dummy longitude coordinate	Automatically extracted
Y	Dummy latitude coordinate	Automatically extracted
COMM_NAME	LAU2 label	GISCO database
NAME_ASCII	LAU2 label in ASCII format	GISCO database
NAME_HTML	LAU2 label in HTML format	GISCO database
NAME_SIRE	LAU2 label in SIRE database	GISCO database
TRUE_COMM_	Dummy variable	GISCO database
CNTR_CODE	Country code	GISCO database
AREA_TOTL	Area	GISCO database
AREA_LAND	Area (only null values)	GISCO database
POPL 2001?	Population in 2001 (LAU2)	GISCO database
NSI_CODE	Code used by the National Statistical Institute	GISCO database
LAU2_CODE	LAU2 code (different from the IDEN and COMM_ID)	GISCO database
ADRG_LAU1_	LAU1 hierarchical code	GISCO database
NUTS_CODE	NUTS hierarchical code	GISCO database
DGUR_CODE	Dummy indicator ?	GISCO database
DGUR_AREA_	Area (text values)	GISCO database
DGUR_AREA	Area	GISCO database
POPL_DENS	Population's density in 2001	GISCO database
NATIONAL_CODE_LAU2	Indicator used in the matching process	Automatically extracted (no values only for Czech Republic)
LABEL_LAU2	Indicator used in the matching process	Automatically extracted (no values only for Czech Republic)
Total population (as of Dec. 31)	Total population (as of Dec. 31)	Regional Database (NSI Slovakia) Indicator valid for 2007
Population - males (as of Dec. 31)	Population - males (as of Dec. 31)	Regional Database (NSI Slovakia) Indicator valid for 2007
Population - females (as of Dec. 31)	Population - females (as of Dec. 31)	Regional Database (NSI Slovakia) Indicator valid for 2007
Population in pre-productive age - total (0 - 14)	Population in pre-productive age - total (0 - 14)	Regional Database (NSI Slovakia) Indicator valid for 2007
Population in productive age - females (15 - 54)	Population in productive age - females (15 - 54)	Regional Database (NSI Slovakia) Indicator valid for 2007
Population in productive age - males (15 - 59)	Population in productive age - males (15 - 59)	Regional Database (NSI Slovakia) Indicator valid for 2007
Population in post-productive age - total (55+F, 60+M)	Population in post-productive age - total (55+F, 60+M)	Regional Database (NSI Slovakia) Indicator valid for 2007
Number of marriages	Number of marriages	Regional Database (NSI Slovakia) Indicator valid for 2007
Number of divorces	Number of divorces	Regional Database (NSI Slovakia) Indicator valid for 2007
Number of live births total	Number of live births total	Regional Database (NSI Slovakia) Indicator valid for 2007
Number of live births males	Number of live births males	Regional Database (NSI Slovakia) Indicator valid for 2007
Number of live births females	Number of live births females	Regional Database (NSI Slovakia) Indicator valid for 2007

Number of deaths total	Number of deaths total	Regional Database (NSI Slovakia) Indicator valid for 2007
Number of deaths males	Number of deaths males	Regional Database (NSI Slovakia) Indicator valid for 2007
Number of deaths females	Number of deaths females	Regional Database (NSI Slovakia) Indicator valid for 2007
Total increase (decrease) of population - total	Total increase (decrease) of population - total	Regional Database (NSI Slovakia) Indicator valid for 2007
Total increase (decrease) of population -males	Total increase (decrease) of population -males	Regional Database (NSI Slovakia) Indicator valid for 2007
Total increase (decrease) of population - females	Total increase (decrease) of population - females	Regional Database (NSI Slovakia) Indicator valid for 2007
Population in total	Population in total	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Population - males	Population - males	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Population - females	Population - females	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Population by nationality: Slovak %	Population by nationality: Slovak %	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Hungar. %	Hungar. %	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Gipsy %	Gipsy %	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Ruthen. %	Ruthen. %	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Ukrain. %	Ukrain. %	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Czech %	Czech %	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Morav. %	Morav. %	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Siles. %	Siles. %	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
German %	German %	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Polish %	Polish %	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Population by religions: Roman-Cathol. %	Population by religions: Roman-Cathol. %	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Evangelic %	Evangelic %	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Greek-Cathol. %	Greek-Cathol. %	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Orthodox %	Orthodox %	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Cz.sl. Hussit. %	Cz.sl. Hussit. %	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
without denom. %	without denom. %	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
other %	other %	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
not specified %	not specified %	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Economically active persons - total	Economically active persons - total	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Economically active persons - males	Economically active persons - males	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Economically active persons - females	Economically active persons - females	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Employed - total	Employed - total	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)

Employed - males	Employed - males	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Employed - females	Employed - females	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Unemployed - total	Unemployed - total	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Unemployed - males	Unemployed - males	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Unemployed - females	Unemployed - females	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Houses total	Houses total	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Permanent habitational houses total	Permanent habitational houses total	Regional Database (NSI Slovakia) Indicator valid for 2001 (CENSUS)
Iden	Basemap code	GISCO database
OBJECTID	Inner code used in ARCVIEW	GISCO database
COMM_ID	Basemap code	GISCO database
LAU2 code	LAU2 code (identical to the IDEN and COMM_ID)	GISCO database
NUTS4	LAU1 hierarchical code in ancient format (NUTS)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
LAU2	LAU2 code (different from the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
NAME	Label used by the NSI Czech Republic	NSI Czech Republic - CENSUS 2001
Population, total	Population, total	NSI Czech Republic - CENSUS 2002
Economically active, total	Economically active, total	NSI Czech Republic - CENSUS 2003
Agriculture, Forestry, Water economy	Agriculture, Forestry, Water economy	NSI Czech Republic - CENSUS 2004
Industry	Industry	NSI Czech Republic - CENSUS 2005
Construction	Construction	NSI Czech Republic - CENSUS 2006
Wholesale and retail trade, Repair of motor vehicles	Wholesale and retail trade, Repair of motor vehicles	NSI Czech Republic - CENSUS 2007
Transport and Communications	Transport and Communications	NSI Czech Republic - CENSUS 2008
Public administration and Defence; Compulsory social security	Public administration and Defence; Compulsory social security	NSI Czech Republic - CENSUS 2009
Education, Health and social work, Veterinary activities	Education, Health and social work, Veterinary activities	NSI Czech Republic - CENSUS 2010
LAU2code	LAU2 code (identical to the IDEN and COMM_ID)	GISCO database
LAU1	LAU1 code	GISCO database
LAU2	LAU2 code (different from the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
name	Label used by the NSI Czech Republic	NSI Czech Republic - CENSUS 2001
Population, total	Population, total	NSI Czech Republic - CENSUS 2001
Females	Females	NSI Czech Republic - CENSUS 2001
MALESSingle	MALESSingle	NSI Czech Republic - CENSUS 2001
MALESMarried	MALESMarried	NSI Czech Republic - CENSUS 2001
MALESDivorced	MALESDivorced	NSI Czech Republic - CENSUS 2001
MALESWidowed	MALESWidowed	NSI Czech Republic - CENSUS 2001
MALESUnknown	MALESUnknown	NSI Czech Republic - CENSUS 2001
FEMALESSingle	FEMALESSingle	NSI Czech Republic - CENSUS 2001
FEMALESMarried	FEMALESMarried	NSI Czech Republic - CENSUS 2001
FEMALESDivorced	FEMALESDivorced	NSI Czech Republic - CENSUS 2001
FEMALESWidowed	FEMALESWidowed	NSI Czech Republic - CENSUS 2001
FEMALESUnknown	FEMALESUnknown	NSI Czech Republic - CENSUS 2001
LAU2codeOk	LAU2 code (identical to the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
NUTS 4	LAU1 hierarchical code in ancient format (NUTS)	Automatically extracted (NSI Czech Republic - CENSUS 2001)

Municipality code	LAU2 code (different from the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
Municipality name	Label used by the NSI Czech Republic	NSI Czech Republic - CENSUS 2001
Population, total	Population, total	NSI Czech Republic - CENSUS 2001
A.G. 0-4	A.G. 0-4	NSI Czech Republic - CENSUS 2001
A.G. 5-14	A.G. 5-14	NSI Czech Republic - CENSUS 2001
A.G. 15-19	A.G. 15-19	NSI Czech Republic - CENSUS 2001
A.G. 20-29	A.G. 20-29	NSI Czech Republic - CENSUS 2001
A.G. 30-39	A.G. 30-39	NSI Czech Republic - CENSUS 2001
A.G. 40-49	A.G. 40-49	NSI Czech Republic - CENSUS 2001
A.G. 50-59	A.G. 50-59	NSI Czech Republic - CENSUS 2001
A.G. 60-64	A.G. 60-64	NSI Czech Republic - CENSUS 2001
A.G. 5-74	A.G. 5-74	NSI Czech Republic - CENSUS 2001
A.G. 75+unknown	A.G. 75+unknown	NSI Czech Republic - CENSUS 2001
LAU2 code	LAU2 code (identical to the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
NUTS4	LAU1 hierarchical code in ancient format (NUTS)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
Municipality code	LAU2 code (different from the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
Municipality name	Label used by the NSI Czech Republic	NSI Czech Republic - CENSUS 2001
Population aged 15+	Population aged 15+	NSI Czech Republic - CENSUS 2001
Without education	Without education	NSI Czech Republic - CENSUS 2001
Basic incl. not finished	Basic incl. not finished	NSI Czech Republic - CENSUS 2001
Secondary vocational and technical without GCSE	Secondary vocational and technical without GCSE	NSI Czech Republic - CENSUS 2001
Full secondary general with GCSE	Full secondary general with GCSE	NSI Czech Republic - CENSUS 2001
Higher professional and Extension study	Higher professional and Extension study	NSI Czech Republic - CENSUS 2001
University	University	NSI Czech Republic - CENSUS 2001
Not identified	Not identified	NSI Czech Republic - CENSUS 2001
LAU2 code	LAU2 code (identical to the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
NUTS 4	LAU1 hierarchical code in ancient format (NUTS)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
Municipality code	LAU2 code (different from the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
Municipality name	Label used by the NSI Czech Republic	NSI Czech Republic - CENSUS 2001
Population, total	Population, total	NSI Czech Republic - CENSUS 2001
Czech	Czech	NSI Czech Republic - CENSUS 2001
Moravian	Moravian	NSI Czech Republic - CENSUS 2001
Silesian	Silesian	NSI Czech Republic - CENSUS 2001
Slovak	Slovak	NSI Czech Republic - CENSUS 2001
Romany	Romany	NSI Czech Republic - CENSUS 2001
Polish	Polish	NSI Czech Republic - CENSUS 2001
German	German	NSI Czech Republic - CENSUS 2001
Ukrainian	Ukrainian	NSI Czech Republic - CENSUS 2001
Vietnamese	Vietnamese	NSI Czech Republic - CENSUS 2001
LAU2 code	LAU2 code (identical to the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
NUTS4	LAU1 hierarchical code in ancient format (NUTS)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
Municipality code	LAU2 code (different from the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
Municipality name	Label used by the NSI Czech Republic	NSI Czech Republic - CENSUS 2001
Population, total	Population, total	NSI Czech Republic - CENSUS 2001
Believers	Believers	NSI Czech Republic - CENSUS 2001
Roman Catholic Church	Roman Catholic Church	NSI Czech Republic - CENSUS 2001
Czechoslovak Hussite Church	Czechoslovak Hussite Church	NSI Czech Republic - CENSUS 2001

Evangelical Church of Czech Brethren	Evangelical Church of Czech Brethren	NSI Czech Republic - CENSUS 2001
Orthodox Church	Orthodox Church	NSI Czech Republic - CENSUS 2001
Jehovah` Witnesses	Jehovah` Witnesses	NSI Czech Republic - CENSUS 2001
Undenominational	Undenominational	NSI Czech Republic - CENSUS 2001
Unknown Denomination	Unknown Denomination	NSI Czech Republic - CENSUS 2001
LAU2 code	LAU2 code (identical to the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
NUTS4	LAU1 hierarchical code in ancient format (NUTS)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
Municipality code	LAU2 code (different from the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
Municipality name	Label used by the NSI Czech Republic	NSI Czech Republic - CENSUS 2001
Population, total	Population, total	NSI Czech Republic - CENSUS 2001
Economically active (EA), total employed persons	Economically active (EA), total employed persons	NSI Czech Republic - CENSUS 2001
EA pensioners	EA pensioners	NSI Czech Republic - CENSUS 2001
women on maternity leave	women on maternity leave	NSI Czech Republic - CENSUS 2001
unemployed persons	unemployed persons	NSI Czech Republic - CENSUS 2001
Economically inactive (EI), total	Economically inactive (EI), total	NSI Czech Republic - CENSUS 2001
EI pensioners	EI pensioners	NSI Czech Republic - CENSUS 2001
Pupils,students, apprentices	Pupils,students, apprentices	NSI Czech Republic - CENSUS 2001
Economic activity not identified	Economic activity not identified	NSI Czech Republic - CENSUS 2001
LAU2 code	LAU2 code (identical to the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
NUTS 4	LAU1 hierarchical code in ancient format (NUTS)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
Municipality code	LAU2 code (different from the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
Municipality name	Label used by the NSI Czech Republic	NSI Czech Republic - CENSUS 2001
Commuters to work	Commuters to work	NSI Czech Republic - CENSUS 2001
within municipality	within municipality	NSI Czech Republic - CENSUS 2001
within district	within district	NSI Czech Republic - CENSUS 2001
within region	within region	NSI Czech Republic - CENSUS 2001
into other region	into other region	NSI Czech Republic - CENSUS 2001
Commuters to work daily out of municipality	Commuters to work daily out of municipality	NSI Czech Republic - CENSUS 2001
Pupils commuting to schools daily out of municipality	Pupils commuting to schools daily out of municipality	NSI Czech Republic - CENSUS 2001
LAU2 code	LAU2 code (identical to the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
NUTS4	LAU1 hierarchical code in ancient format (NUTS)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
Municipality code	LAU2 code (different from the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
Municipality name	Label used by the NSI Czech Republic	NSI Czech Republic - CENSUS 2001
Dwelling households, total	Dwelling households, total	NSI Czech Republic - CENSUS 2001
with 1 PV*	with 1 PV*	NSI Czech Republic - CENSUS 2001
with 2+PV*	with 2+PV*	NSI Czech Republic - CENSUS 2001
Private households	Private households	NSI Czech Republic - CENSUS 2001
with 1 census household	with 1 census household	NSI Czech Republic - CENSUS 2001
with 2 and over census household	with 2 and over census household	NSI Czech Republic - CENSUS 2001
Census households (C-H), total	Census households (C-H), total	NSI Czech Republic - CENSUS 2001
Two-parent families	Two-parent families	NSI Czech Republic - CENSUS 2001
with dependent children	with dependent children	NSI Czech Republic - CENSUS 2001
Lone-parent families	Lone-parent families	NSI Czech Republic - CENSUS 2001
with dependent children	with dependent children	NSI Czech Republic - CENSUS 2001
Non-family households	Non-family households	NSI Czech Republic - CENSUS 2001
Households of individuals	Households of individuals	NSI Czech Republic - CENSUS 2001

LAU2 code	LAU2 code (identical to the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
NUTS4	LAU1 hierarchical code in ancient format (NUTS)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
Municipality code	LAU2 code (different from the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
Municipality name	Label used by the NSI Czech Republic	NSI Czech Republic - CENSUS 2001
Houses, total	Houses, total	NSI Czech Republic - CENSUS 2001
Permanently occupied houses	Permanently occupied houses	NSI Czech Republic - CENSUS 2001
Family houses	Family houses	NSI Czech Republic - CENSUS 2001
Multi-dwelling houses	Multi-dwelling houses	NSI Czech Republic - CENSUS 2001
Houses by ownershipprivate persons	Houses by ownershipprivate persons	NSI Czech Republic - CENSUS 2001
Houses by ownershipcommunity,state	Houses by ownershipcommunity,state	NSI Czech Republic - CENSUS 2001
Houses by ownershiphousing association	Houses by ownershiphousing association	NSI Czech Republic - CENSUS 2001
Houses builtup to 1919	Houses builtup to 1919	NSI Czech Republic - CENSUS 2001
Houses built1920-1945	Houses built1920-1945	NSI Czech Republic - CENSUS 2001
Houses built1945-1980	Houses built1945-1980	NSI Czech Republic - CENSUS 2001
Houses built1981-2001	Houses built1981-2001	NSI Czech Republic - CENSUS 2001
LAU2 code	LAU2 code (identical to the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
NUTS4	LAU1 hierarchical code in ancient format (NUTS)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
Municipality code	LAU2 code (different from the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
Municipality code	Label used by the NSI Czech Republic	NSI Czech Republic - CENSUS 2001
Houses, total	Houses, total	NSI Czech Republic - CENSUS 2001
by floor (above ground)1-2	by floor (above ground)1-2	NSI Czech Republic - CENSUS 2001
by floor (above ground)3-4	by floor (above ground)3-4	NSI Czech Republic - CENSUS 2001
by floor (above ground)5+	by floor (above ground)5+	NSI Czech Republic - CENSUS 2001
Sewage: connection to the public system	Sewage: connection to the public system	NSI Czech Republic - CENSUS 2001
Water supply system	Water supply system	NSI Czech Republic - CENSUS 2001
Gas supply	Gas supply	NSI Czech Republic - CENSUS 2001
Central heating	Central heating	NSI Czech Republic - CENSUS 2001
LAU2 code	LAU2 code (identical to the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
NUTS4	LAU1 hierarchical code in ancient format (NUTS)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
Municipality code	LAU2 code (different from the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
Municipality name	Label used by the NSI Czech Republic	NSI Czech Republic - CENSUS 2001
Dwellings, total	Dwellings, total	NSI Czech Republic - CENSUS 2001
Permanently occupied dwellings	Permanently occupied dwellings	NSI Czech Republic - CENSUS 2001
Family houses	Family houses	NSI Czech Republic - CENSUS 2001
Multi-dwelling houses	Multi-dwelling houses	NSI Czech Republic - CENSUS 2001
Unoccupied dwellings in permanently occupied houses	Unoccupied dwellings in permanently occupied houses	NSI Czech Republic - CENSUS 2001
Unoccupied dwellings in unoccupied houses	Unoccupied dwellings in unoccupied houses	NSI Czech Republic - CENSUS 2001
occupied temporarily	occupied temporarily	NSI Czech Republic - CENSUS 2001
used for recreation	used for recreation	NSI Czech Republic - CENSUS 2001
LAU2 code	LAU2 code (identical to the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
NUTS4	LAU1 hierarchical code in ancient format (NUTS)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
Municipality code	LAU2 code (different from the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
Municipality name	Label used by the NSI Czech	NSI Czech Republic - CENSUS 2001

	Republic	
Permanently occupied dwellings, total	Permanently occupied dwellings, total	NSI Czech Republic - CENSUS 2001
by legal reason of use : in own house	by legal reason of use : in own house	NSI Czech Republic - CENSUS 2001
by legal reason of use : in own dwelling	by legal reason of use : in own dwelling	NSI Czech Republic - CENSUS 2001
by legal reason of use : rented	by legal reason of use : rented	NSI Czech Republic - CENSUS 2001
by legal reason of use : in dwelling of housing association	by legal reason of use : in dwelling of housing association	NSI Czech Republic - CENSUS 2001
1 living room	1 living room	NSI Czech Republic - CENSUS 2001
2 living rooms	2 living rooms	NSI Czech Republic - CENSUS 2001
3 living rooms	3 living rooms	NSI Czech Republic - CENSUS 2001
4 living rooms	4 living rooms	NSI Czech Republic - CENSUS 2001
5+ living rooms	5+ living rooms	NSI Czech Republic - CENSUS 2001
Iden	LAU2 code (identical to the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
OBJECTID	Inner code used in ARCVIEW	GISCO database
COMM_ID	Basemap code	GISCO database
LAU2code	LAU2 code (identical to the IDEN and COMM_ID)	GISCO database
NUTS4	LAU1 hierarchical code in ancient format (NUTS)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
Municipality code	LAU2 code (different from the IDEN and COMM_ID)	Automatically extracted (NSI Czech Republic - CENSUS 2001)
Municipality name	Label used by the NSI Czech Republic	NSI Czech Republic - CENSUS 2001
Permanently occupied dwellings, total	Permanently occupied dwellings, total	NSI Czech Republic - CENSUS 2001
dwellings by basic amenities : Gas supply in dwelling	dwellings by basic amenities : Gas supply in dwelling	NSI Czech Republic - CENSUS 2001
dwellings by basic amenities : Water supply in dwelling	dwellings by basic amenities : Water supply in dwelling	NSI Czech Republic - CENSUS 2001
dwellings by basic amenities : Private flush toilet	dwellings by basic amenities : Private flush toilet	NSI Czech Republic - CENSUS 2001
dwellings by basic amenities : Bathroom, shower inside dwelling	dwellings by basic amenities : Bathroom, shower inside dwelling	NSI Czech Republic - CENSUS 2001
dwellings by basic amenities : Central heating	dwellings by basic amenities : Central heating	NSI Czech Republic - CENSUS 2001
dwellings by basic amenities : Single-storey heating	dwellings by basic amenities : Single-storey heating	NSI Czech Republic - CENSUS 2001
Average number of : dwelling persons	Average number of : dwelling persons	NSI Czech Republic - CENSUS 2001
Average number of : persons per living room up to 8 m2	Average number of : persons per living room up to 8 m2	NSI Czech Republic - CENSUS 2001
Average number of: occupied living area per dwelling	Average number of: occupied living area per dwelling	NSI Czech Republic - CENSUS 2001
Average number of : occupied living area per 1 person	Average number of : occupied living area per 1 person	NSI Czech Republic - CENSUS 2001
Average number of : living rooms per dwelling	Average number of : living rooms per dwelling	NSI Czech Republic - CENSUS 2001

## References

### • *Litterature*

Korte B. G., 2001, The GIS book 5<sup>th</sup> edition, Onword Press, New York

Turcanasu G., Rusu A., 2008, Le système des villes en Bulgarie et en Roumanie. Quelles perspectives pour un polycentrisme?, Espace Geographique, no.4/2008

Groza O., 2005, Maillages administratifs officiels et identités territoriales officieuses: les échelons spatiaux de la différenciation identitaire en Roumanie, in V. Rey; T. Saint-Julien - „Territoires d'Europe, la différence en partage“, ENS- Editions, Lyon, pp.153-160,

### • *Websites*

<http://www.insse.ro/cms/rw/pages/index.en.do>, National Institute of Statistics, Romania

[http://www.nsi.bg/index\\_en.htm](http://www.nsi.bg/index_en.htm), National Statistical Institute of Bulgaria

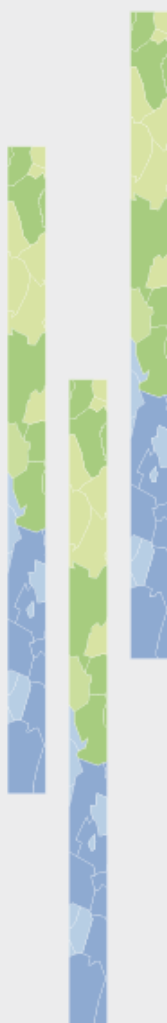
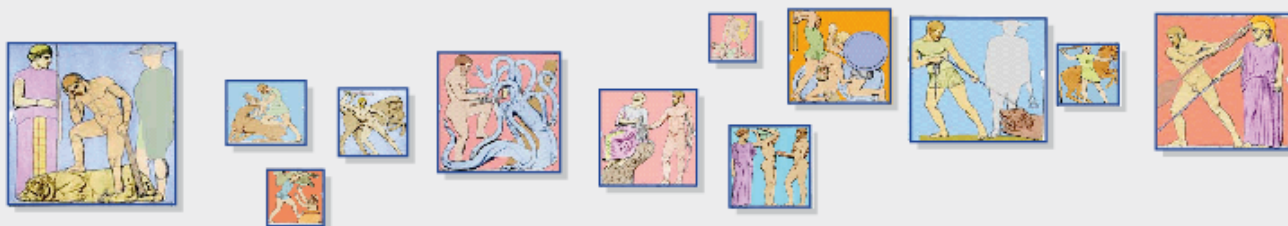
[http://px-web.statistics.sk/PXWebSlovak/index\\_en.htm](http://px-web.statistics.sk/PXWebSlovak/index_en.htm), Statistical Office of the Slovak Republic

<http://www.czso.cz/eng/redakce.nsf/i/home>, Czech Statistical Office

[http://ec.europa.eu/eurostat/ramon/nuts/lau\\_en.html](http://ec.europa.eu/eurostat/ramon/nuts/lau_en.html), Eurostat – list of LAU

[www.territorial-intelligence.eu/caenti/](http://www.territorial-intelligence.eu/caenti/), Coordination action of the European Network of Territorial Intelligence





## Local & regional data

### *Producing innovative indicators at local scale*

#### MAIN RESULTS

- Local data can be integrated from different sources or it can be produced using the LAU2 as elementary spatial patterns
- Using an appropriate number of LAU2, the data can be mobilized in specific geographical models (the estimated GDP 2006 distribution at local scale)
- The distances calculated at local scale represent an useful analysis tool which can provide new territorial indicators

**ESPON 2013 DATABASE**

**MARCH 2011**



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

**36 pages**

## **LIST OF AUTHORS**

Octavian Groza, UAIC, CUGUAT-TIGRIS, Iasi, Romania  
Alexandru Rusu, UAIC, CUGUAT-TIGRIS, Iasi, Romania

### **Contact**

[octaviangroza@yahoo.com](mailto:octaviangroza@yahoo.com)

tel. + 40 0232 20 14 88

# TABLE OF CONTENT

<b>Introduction.....</b>	<b>3</b>
<b>1 The production of basic indicators, using the LAU2 geometries and data.....</b>	<b>4</b>
1.1 The data harmonization at LAU2 scale is depending on the spatial harmonization of the geometries for the selected countries.....	4
1.2 Working with a large number of LAU2 units involves accepting the extraordinary values.....	4
1.3 Mapping the evolution(s) in connection with some explanatory factors.....	5
1.4 Explaining the demographic decline: some basic hypothesis (the spatial auto-correlation).....	6
<b>2 The production of some basic indicators, using the LAU2 geometries as a geographic objects Spatial patterns vs spatial structures.....</b>	<b>9</b>
2.1 A short introduction in the spatial patterns: points, polygons, lines and networks The LAU2 as spatial patterns	9
2.2 Using the spatial patterns as a base for the indicators construction	9
2.3 A case study on Romania and Hungary – the calculation of the LAU2 accessibility	9
<b>3 Towards more elaborate indicators and models applied to LAU2 objects The potential of interaction as a measure for the local economic performance</b>	<b>13</b>
3.1 How to measure the economic performance at LAU2 scale?.....	13
3.2 How to find relevant and harmonized data for this operation? A case study on the local aggregated turnover in 2006 – the Romanian case.....	14
3.3 How to by-pass the lack of harmonization using grid information?.....	15
3.4 Integrating disaggregated data on the LAU2 geometry frame – a problem of spatial matching.....	16
3.5 The calculation of the potential model of interaction for the local estimated GDP.....	20
<b>4 Towards more elaborate indicators and models applied to LAU2 objects The settlement hierarchy and the territorial architecture of the selected countries.....</b>	<b>24</b>
4.1 Using the distance as an indicator for territorial coherence in the Eastern Europe.....	24
4.2 The settlement’s hierarchy – how many levels function from Prague to Sulina?.....	24
4.3 How to map the distance: choropleth vs “oursins”.....	26
4.4 The local fragmentation of the territorial architecture and the hierarchical immobility	26
<b>Conclusions .....</b>	<b>28</b>
<b>Annexes .....</b>	<b>29</b>

## **Introduction**

There are four major issues to explore in this technical report. The first two are related to the construction of basic indicators, at LAU2 scale, indicators that are capable to describe and better shape specific geographical trends, such as the demographic evolution for 5 selected countries in the ESPON space. The last two parts explore the methodological problems induced by the elaboration of more complicated indicators and variables, especially in the context of a large number of spatial units. In the first part, the annexes focus on the description of a set of variables that relate the land use problematic to the local geometries. In a second part, the annexes will present some methodological aspects linked to the estimation of the local economic performance at LAU2 level, for all the ESPON space.

Intersecting specific techniques of spatial analysis with the problems raised by the management of the LAU2 geometry is another aspect present in the elaboration of this report. The questions of accessibility, the definition of spatial patterns at local level and mapping the territorial relations between these spatial patterns occupied a large part of our methodological exploration.

In a first logic, this report was conceived as a handbook of good practices and specific issues related to the LAU2 scale of analysis. However, the geographic complexity of the main studied area (Bulgaria, Czech Republic, Hungary, Romania, Slovakia) attenuates this first intention. As a matter of fact, our option for these 5 countries represents a compromise between the continuation of the first explorations in the work for this challenge and the need of a representative space for methodological experimentation (a sufficiently large number of LAU2).

## **1 The production of basic indicators, using the LAU2 geometries and data.**

### **1.1 The data harmonization at LAU2 scale is depending on the spatial harmonization of the geometries for the selected countries.**

Our intention is to describe by a map the demographic evolutions at LAU2 scale for the selected countries, between 2001 and 2006. This intention is based on the presence of harmonized data concerning the number of inhabitants for the two years. If the data is ready to use, the spatial frame (LAU2 geometries) is a major issue. If in 2001 some of the capitals were presented as a single polygon, in 2006 their administrative territory is divided (Budapest more than 20 LAU2, Bratislava more than 15, Bucarest 6 sectors). It isn't a major issue, as a matter of fact the aggregation of data is just a time problem. Other LAU2 are more concerning. Some of them have disappeared during 2001-2006, some others just appeared on the map, without necessary having a connection between the two categories. The logical way to deal with the spatial lack of harmonization is to exclude them from the analysis. This option becomes illogical when the administrative mutations in the local geometry become too important. In that case, it is recommended to apply various techniques that can finally harmonize these spatial dynamics.

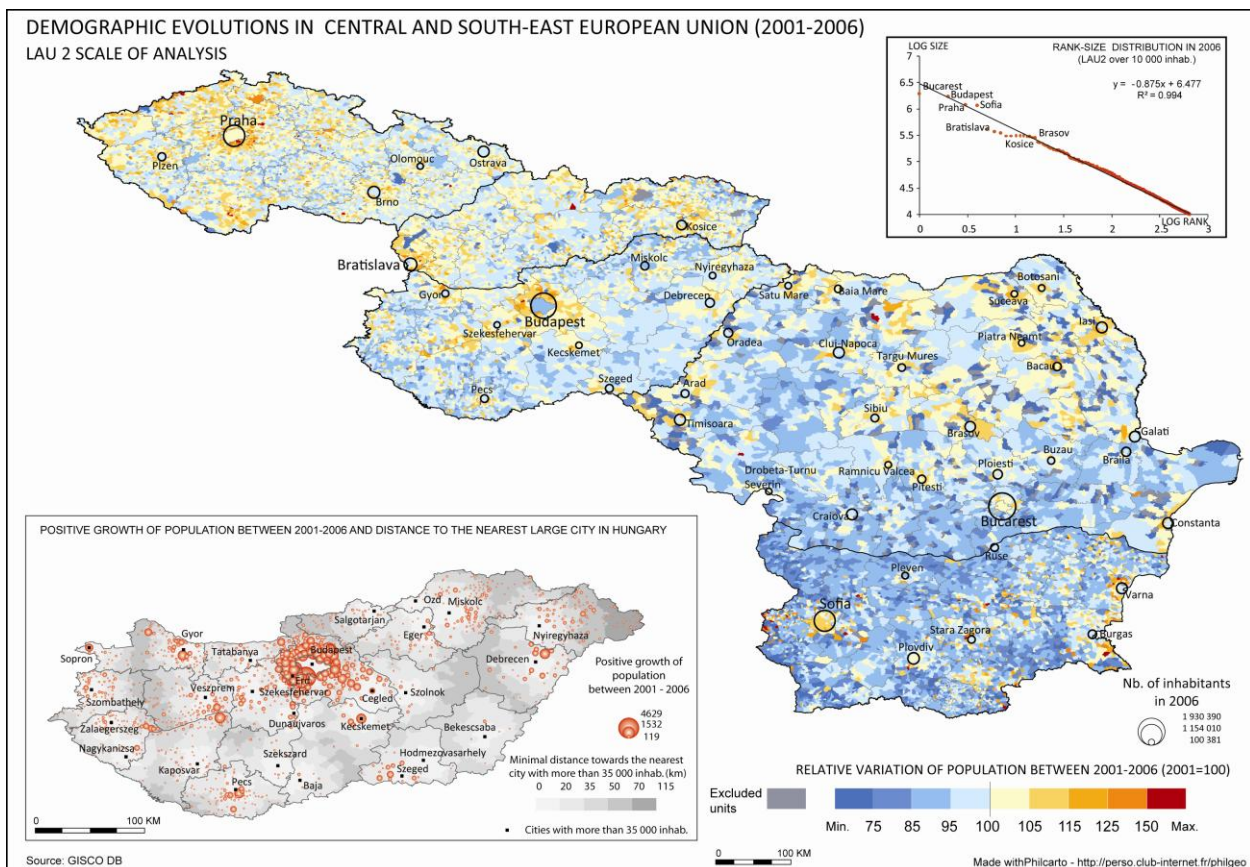
### **1.2 Working with a large number of LAU2 units involves accepting the extraordinary values.**

The issue of the extraordinary values is induced by the large number of spatial units with almost insignificant population (in Bulgaria we deal with municipalities that used to have 10 inhabitants in 2001). In this LAU2, every change in the population will transmit "considerable" evolutions when we map relative indicators, such as the variation of population for the mentioned period. The same phenomena can be observed for some LAU2 that are situated in the proximity of large cities, involved in processes of sub-urbanization. As a matter of fact, the extraordinary values are not extra-ordinary at all, if one would look at the local context in which the demographic trends are deployed. Usually, these values are consistent with the regional evolutions that shape the demographic decline or growth, the positive spatial auto-correlation being a general rule. One special case of extraordinary value appears in areas with administrative mutations. If one LAU2 suffered an administrative division/split, it will create a "fake" extraordinary value that reflects the creation of a new LAU2 rather than a demographic evolution.

### **Administrative mutations and extraordinary values**

There are many cases of administrative mutations that can influence the mapping process. One of the most common issue is the territorial division, from one LAU2 resulting 2 new spatial units. We cannot correctly estimate the demographic evolution, in this case, for obvious reasons. Sometimes is not the division of LAU that becomes a problem. The administrative union of two LAU also interferes with the data calculation, even for simple indicators like the evolution of population. We can find even more complicated cases where the

division is followed by a union, involving massing data or extraordinary values for three LAU. Occurring very often in Romania and Bulgaria, these mutations are reflecting the importance of the local dimension in the political and administrative strategies. We can observe on the map that these changes in the basic geometry are correlated with the economic territorial rhythms. Thus, in some less dynamic regions, the fragmentation indicates the stake of the public administrative finance at local scale, while in the pro-active regions we can interpret it as a trend to concentrate the local financial resources (the proximity of cities, touristic regions, rich industrial and transportation corridors).



**Figure 1** Trends of demographic evolution in 5 selected countries of the ESPON space (draft map)

### 1.3 Mapping the evolution(s) in connection with some explanatory factors.

Despite the spatial fragmentation and the lack of coherence regarding the LAU2 geometry of the 5 countries, some general trends can be easily identified. The positive demographic evolution is a function of some qualitative and quantitative transformations that reshape the role of the large cities in the territory: sub-urbanization and metropolitan development. However, defining the large city in the area of the selected countries is not an easy task. The rank size distribution for 2006 indicates that a possible superior limit should be 280 000 inhabitants (an approximation). Many cities below this limit are also involved in the suburbanization process, reflecting their key position in the national urban system and also their economic strength. Sometimes, the

demographic growth is controlled by the distance towards the nearest city, such is the case in Hungary – extremely visible for Budapest, but also for Pecs, Szeged or Debrecen. On the other side, some rural regions (with a traditional pro-natalist behavior) like Moldavia (Romania) or Haskovo-Kardzali region (Bulgaria) still conserve positive evolution. Other recent studies emphasized that this natalist behavior is in extinction. A special case is the Central-Transylvania where the final phase of the demographic transition was conjecturally interrupted by some ethnic and confessional local specificity. The demographic decline is a challenging reality for vast rural spaces in Bulgaria, Romania and Hungary, but also regionally present in Czech Republic and Slovakia. Taking into account the surface concerned by this phenomena - the cross-border Danubian region between Romania and Bulgaria or the cross-border region between Romania and Hungary, this decline could represent a policy relevant aspect regarding the demographic evolutions in the ESPON space.

#### **1.4 Explaining the demographic decline: some basic hypothesis (the spatial auto-correlation).**

The distribution of the three dimensions of the demographic evolution (stability, decline and growth) present a specific spatial pattern. As a general rule, the LAU2 characterized by decline seem to have neighbors presenting the same trend. In the proximity of the cities, the LAU2 positive evolution is also included in a locally homogeneous context. The area of relative demographic stability is also subject to similarity with the neighbors (this area is visible in the Romanian Sub-Carpathians). This effect is called spatial auto-correlation and it is largely developed and formalized in the geographic litterature. There are many ways in which we can test its existence (Geary test, Moran's I or the measure of the local dissimilarity). In our case, testing the presence of the spatial auto-correlation in the demographic evolution is a good method to estimate the relationship between the indicator and the local context. Basically, we try to estimate the size of a homogeneous region that is characterized by the same demographic trend, at local scale. That is the sens of the word "explaining" in the title of this fragment. Technically, we have to follow several steps in order to obtain the size of a homogeneous region:

- 1) Choose a method: testing the spatial auto-correlation using a GIS is a simple task. In the absence of a GIS, there is a spreadsheet method that involves the manipulation of a large table of geographic information.
- 2) Depending on the method, reflecting on the concept of neighborhood is also useful. At local scale, the administrative contiguity is problematic because we have to take into account the spatial fragmentation (e.g. administrative contiguity is problematic in Bulgaria, exaggerated in Romania, just good in Hungary). A distance bandwidth will be a better option in order to define the neighbors and the proximity effect.
- 3) Three hypothesis are now available for testing :  
**H0: the spatial auto-correlation is null.** Any two neighbor LAU are neither similar, neither different to any two LAU that are not neighborhood related.

**H1: the spatial auto-correlation is negative.** Two neighbors LAU are less similar than two distanced LAU. The local context is characterized by heterogeneity.

**H2: the spatial auto-correlation is positive.** Two neighbors LAU are rather similar compared with two distanced LAU. The local context is characterized by homogeneity.

- 4) If the test confirms the **H2**, by enlarging the neighborhood context we can obtain the size of a locally homogeneous region in relation with the demographic evolution. For more details, see <http://grasland.script.univ-paris-diderot.fr/>

In our case study, we have made an option for the GIS solution because using a spreadsheet method will involve the manipulation of 20 000 by 20 000 matrix, which is a time consuming strategy, difficult to implement. The test of spatial autocorrelation (Moran's I) confirms the **H2** hypothesis (0.145 in a range of 15 km around each LAU2 in the 5 selected countries, statistically representative for more than 20 000 LAU2) and shows that the local context and the demographic trends are related. In this case, we have proceeded to the development of the point 4 (see above). Despite our effort, we cannot provide the size of the homogeneous regions having the same demographic pattern for all the countries included in our study. Therefore, we have applied the test of spatial autocorrelation at different distance bands only for Hungary and Romania. The general trend of the Moran's I distribution for the two countries shows a decrease of the local context's effect when we extend the definition of the "local" from 5 to 50 km. For Hungary the test was started at 5 km and then incremented by 2 km. In the Romanian case, to avoid data exclusion, the spatial auto-correlation was measured at 10 km and incremented by 5 km, until the 50 km limit was reached. The analysis of the Moran's I distribution pattern shows that the local effects diminish gradually, a clear limit for regions with similar behavior being difficult to estimate.

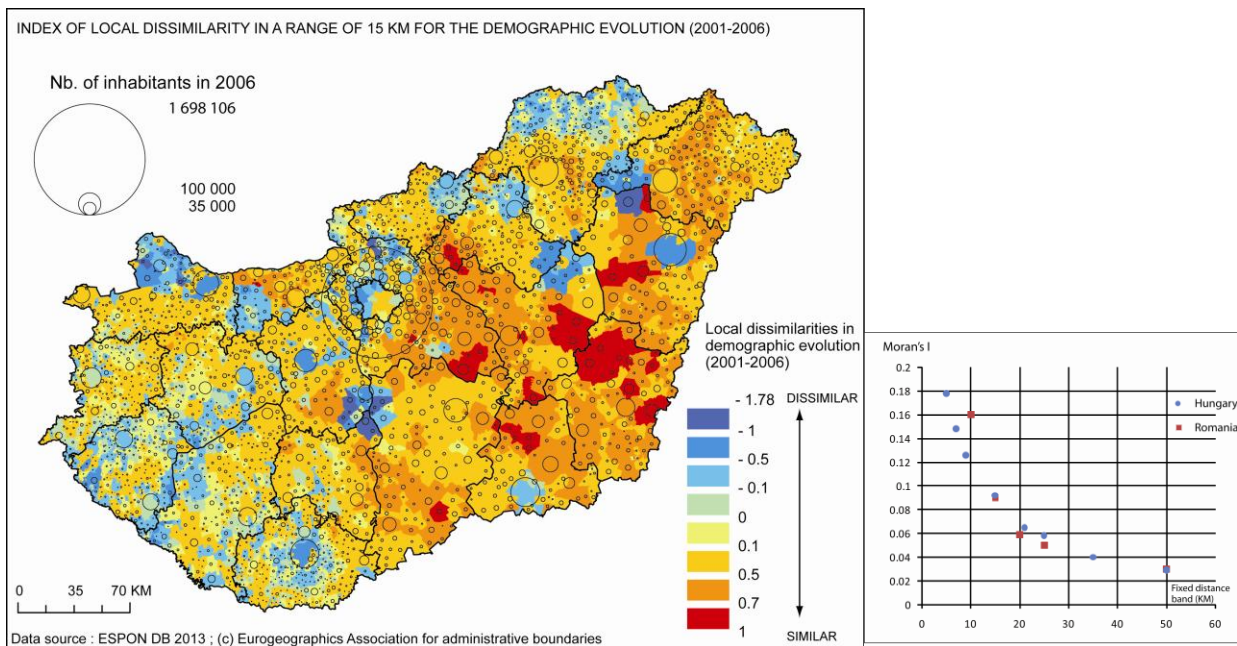
### **Hungary and Romania**

The general aspect of the Moran's I distribution indicates that the auto-correlation effect decrease almost as a power function of distance in both cases. The low values of the indicator of spatial auto-correlation define an ambiguous relation between the local context and the demographic trends. With more than 3000 spatial units for each country, even these low values are representative. A better illustration of this local effect could be emphasized by using a map of the spatial distribution of the local dissimilarities. Applied for Hungary, this kind of map is an exploration tool (not very sophisticated), showing some of the local sensitivities.

Reading a similarity/dissimilarity map is not an easy task. Rather than showing the spatial repartition of a phenomena (demographic dynamics), it shows how the spatial units are acting in relation with their local context (15 km in this case), regardless of the trend – decline, stability, growth. The positive values in the legend indicate local homogeneity, while the negative ones a different behavior compared with the neighborhood. The color's gradation shows the intensity of the homogeneity vs. heterogeneity spatial distribution concerning the demographic evolution in Hungary. In a context of general decline of



population, a large part of the Hungarian territory is behaving spatially auto-correlated. As a map description, the heterogeneous areas are situated near the large Hungarian cities and in the proximity of the “triplum confinium” of some of the NUTS 3. The homogeneous zones are occupying the central and the eastern parts of Hungary and this moderate lack of dissimilarities should also be linked to the size of the LAU2 in these regions. This map is the output of a double interrogation in a similarity matrix weighted by distances, a map that involved the manipulation of more than 3000 rows and columns. However, this method (even if time consuming) provides a good tool in order to better interpret the spatial repartition of the demographic dynamics and could be particularly useful in studies concerning the cross-border regions.



**Figure 2** Distribution of Moran's I as a distance decay function in Hungary and Romania

## **2 The production of some basic indicators, using the LAU2 geometries as a geographic objects. Spatial patterns vs. spatial structures.**

### **2.1 A short introduction in the spatial patterns: points, polygons, lines and networks. The LAU2 as spatial patterns.**

We like it or not, the geographical reality cannot be synthesized in more than 3 elementary geometries: points, lines and surfaces. Each type of spatial pattern (also called sometimes spatial structures, even if arguable concerning the epithet) involves specific methodological approaches. As an example, the point patterns can be analyzed using the weighted centroid technique, the networks by using the graph theory and the surfaces by taking into account the shape of the polygons. Reducing the LAU2 or the LAU2 information to elementary geometrical features allows us to produce some geographic indicators that can be integrated in a local database. The interest of the spatial patterns analysis is not to fill fields of information in a table (it is also a method to increase the inflation of the information), it rather touches the need to intersect or relate indicators in an explanatory process (e.g. the local economic performance as an eventual output of the accessibility).

### **2.2 Using the spatial patterns as a base for the indicators construction.**

Some of the techniques that we can use in order to approach these different spatial patterns will provide only synthetic indicators (the weighted centroid of the population or the standard distance deviation as a function of some central features, such as capitals). Producing specific information for each LAU2 situated in our 5 countries involves a different approach. Using the reticular spatial structures and the LAU2 centroids for Hungary and Romania we have explored the possibility to relate the geographical position of the spatial units to their position in a network. One of the common methods is to calculate distances in both the network and the geographical coordinate system. As a matter of fact, we have tried to confront Euclidean distances between LAU2 to the "real" distances in a network (the road network is sufficiently detailed to allow it). The obtained indicator is not quite an accessibility indicator, it rather functions as a network efficiency measure.

### **2.3 A case study on Romania and Hungary – the calculation of the LAU2 accessibility.**

The Euclidean distances in a point spatial pattern reflect relations that might occur in an isotropic and homogeneous territory. The distances within the network will measure the relations that might occur in a historically planned transportation system. A comparison between the two kinds of distances allows us to map for Hungary and Romania a derived indicator of local accessibility. In both the countries the low values appears in the plain regions or in regions with a fair road network density. When the connectivity in the network is problematic or when we take into account the mountain zones, we observe higher values of the indicator. Sometimes, these high values are linked with the absence of some essential infrastructural features: a bridge across the Danube or other main rivers, sinuous transportation corridors. The deviation to

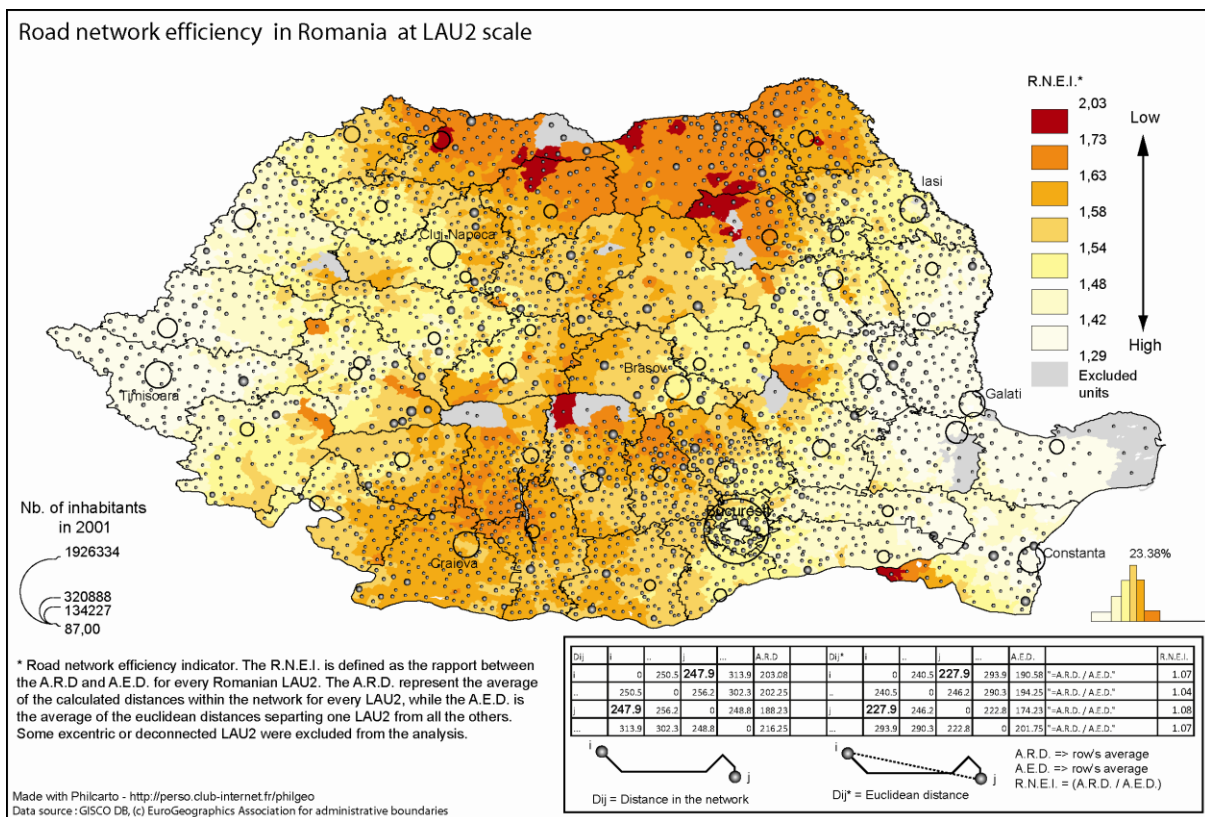
the shortest Euclidean path will have an influence on the time distances and the cost distances, being sometimes an issue in the construction of the local economic performance. It also reflects the high degree of dependence on the mentioned essential infrastructural features and the lack of alternative road segments in the transportation process. In the Romanian case the high and problematic values of the indicator are specific to some remote LAU2 from the Carpathian Mountains and in regions with an intense relief fragmentation. In Hungary we observe a concentration of the high values in the southern part, confirming the spatial discontinuity effect of the Danube. Some LAU2 were not included in the calculation process because their centroid is not in the proximity of the road network (more than 10 km.). It is the case for LAU 2 situated in the Danube Delta (where there is no road network and the access is granted by water) and in very isolated mountain regions. One critical aspect of the two maps is the lack of harmony in the classification of values in the legend. The reason for this discrepancy is the conservation of the classification method – the natural breaks (Jenks). This lack of harmony also shows that low or medium values for Romania do not have the same sense in Hungary, where they can be considered intense deviations to the shortest Euclidean path. We can also speculate and think that in countries with a larger surface, the probability to encounter high deviation of the values is also stronger. There are some steps that we have to take into consideration when working with the road network efficiency indicator:

- verifying the topology of the network (the connectivity algorithm and the presence of bridges) is a long process. In its absence bizarre situations and extraordinary values may appear.
- Choosing an appropriate system of map projection is also needed in order to avoid exaggerated deviations for the Euclidean distances.
- The conversion of the Euclidean distances in km or other distance units should be carefully supervised.

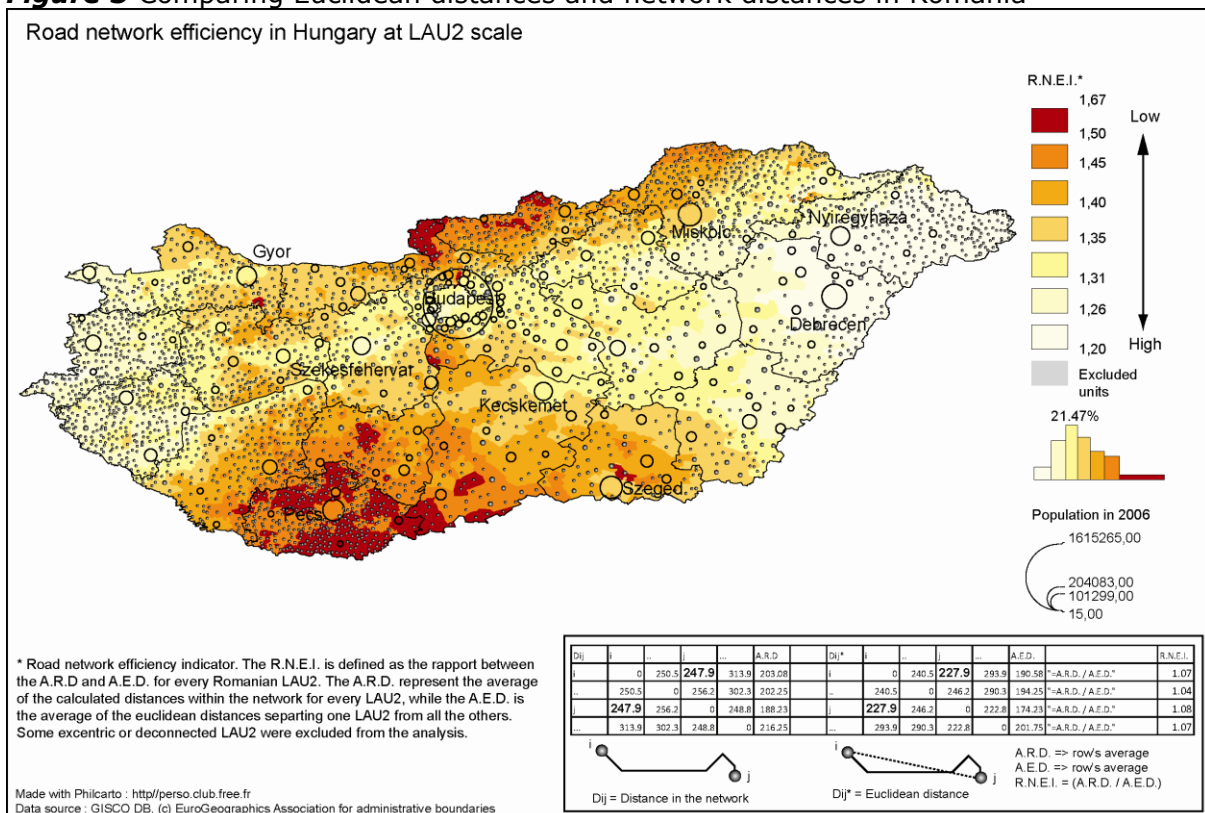
## **Accessibility at local scale**

Some of the ESPON projects already focused on the issues induced by the accessibility, but most of them touched only the NUTS3 frame. When the local dimension was taken into consideration, it was the accessibility to the various networks (roads or rails) that presented interest. One challenging intention would be to confront basic indicators of accessibility (the average distance that separates a LAU2 from all the other) in a double territorial context – the point spatial patterns and the network spatial patterns. The output of this approach reflects the relationship between the so called “natural accessibility” (the role of the natural features in the construction of the local accessibility) and the general characteristics of the road transportation network. The steps needed to complete this kind of analysis start with the construction of the Euclidean accessibility matrix. After that, the extraction of the average distance for every LAU2 is needed. A second problem is to build an origin-destination matrix in a common GIS. By summarization we will obtain the average distances within the network. The indicators will be put into a simple report and the road network efficiency measure is available for the mapping process. This efficiency refers only to the geometric characteristics of the network and it

does not take into account the qualitative aspects. If we map only one indicator, such as the average distance that will separate each LAU2 from all the other, a core-peripheries model will appear.



**Figure 3** Comparing Euclidean distances and network distances in Romania



**Figure 4** Comparing Euclidean distances and network distances in Hungary

In the absence of data, indicators like accessibility or deviation to the shortest path could become an interesting set of indicator, a tool in the exploration or description of territories at local scale. In their estimation process we mobilize the elementary spatial patterns (the points, the lines and the surfaces) and we use classical spatial analysis methods. The size of the studied region is an important aspect in the data creation. Despite the efforts, we were not able to work on a space with more than 3000 LAU2 (an approximate value) when calculating the Euclidean and the distances in the network. When we need to investigate the local accessibility for larger regions (two neighbor countries), solutions might appear if we change the working methodology and if we begin to split the table of information in data packages that don't exceed 10 million cells.

### **3 Towards more elaborate indicators and models applied to LAU2 objects. The potential of interaction as a measure for the local economic performance**

#### **3.1 How to measure the economic performance at LAU2 scale?**

The economic performance is an important dimension in the panel of official indicators defined in the Lisbon strategy. Generally it is measured using the GDP (per capita, per employed or the rhythm of its formation) as a relevant picture of the economic success. However, this success is depending on the scale that we intend to use when we map it. At NUTS level (0,2,3) the spatial pattern in the repartition of the GDP is a core-periphery/peripheries matter of distribution. At local level (LAU), we can assume that this concentration is still visible, if we would have access to data. Possibly, this local concentration would have as actors the capitals, the metropolitan areas and some privileged rural regions. At the opposite, the low or under-average performance might be associated to some **remote areas**, still in transition regions or un-adapted urban networks. This eventual "jeu d'échelle" of the investigation would finally show how the repartition of economic welfare or performance is a subject to the spatial frame in which we try to fit it. For an external neutral reader, these assumptions might sound as the foundations of a hypothesis. Data concerning the GDP at LAU2 level is generally absent for the 31 countries included in the ESPON space. In their relative absence, it is impossible to confirm the existence of the **trans-scalar** spatial processes in the distribution of the economic performance. There are two options in this case: either the use of an alternative measure for the economic performance, either to try to estimate the values for every single LAU2. We have explored the two possibilities, taking into account that the recommended solution should be a dominant strategy for an eventual research group working on the local databases. The first solution is to explore unofficial but reliable data sources that can offer indicators on the economic trends at local scale, knowing that harmonization with other countries will be a difficult task to assume. The second option is to approximate/estimate the values using **spatial analysis** techniques.

##### **Remote areas**

These territorial units could be considered only one dimension of a larger concept: the territories with specific geographical features (ESPON, 2006). However, the specificities are not equivalent to some territorial lack of assets, especially when zooming at local scale.

##### **Trans-scalar**

A geographical attribute of the spatial distributions that intersect the MAUP (modifiable area unit problem). Basically, one territorial repartition might not have the same pattern, when we observe it at different scales of analysis.

##### **Spatial analysis**

According to the modern classifications, the spatial analysis is a method used to observe and measure the spatial structures. In the GIS, especially in the mainstream of the specific jargon, the spatial analysis is a technique used to perform topologic and logic operations between the spatial features: join operations, intersections, updates of geometries etc.

### **3.2 How to find relevant and harmonized data for this operation? A case study on the local aggregated turnover in 2006 – the Romanian case.**

In Romania, but also in other countries from the ESPON space, accessing data concerning the economic performance at local scale is a challenge. During our networking activities with the official data providers, we have found two significant facts about this type of indicator: 1) the Romanian NSI doesn't necessarily gather it; 2) other official databases, such as the Ministry of Industry, will provide data aggregated by branch at NUTS scale. If GDP is not free and officially available at LAU2, we can use other reliable data providers in order to obtain information about the economic success at local scale. One way to do it is to integrate data about the economic actors that are present in a certain territory. In the Romanian case, we have downloaded and exploited a free product (a database with more than **600 000 firms** and the basic information about their economic behavior: nb. of employees in 2006, the turnover in 2006 and the foreign direct participation). In other ESPON countries, similar products might be available. This free package of data present reliability for several reasons: the data is available for consultation and validation passing by other sources (the Ministry of Finances); the territorial picture fits the expectations and the recent trends. The values concerning the firm's turnover were integrated by matching the name of the settlements with the **SIRUTA codes** (Romanian NSI official code) for more than 12 000 spatial units that compose the 3000 LAU2 from this country. It was a challenging case study because it involved a large amount of data, specific matching algorithms and it also offered a general picture of what is happening under the LAU2 level. Practically we summarized the turnover of all the economic actors that are active in a Romanian LAU2. This approach has advantages and weakness that we will discuss next.

#### **Firms**

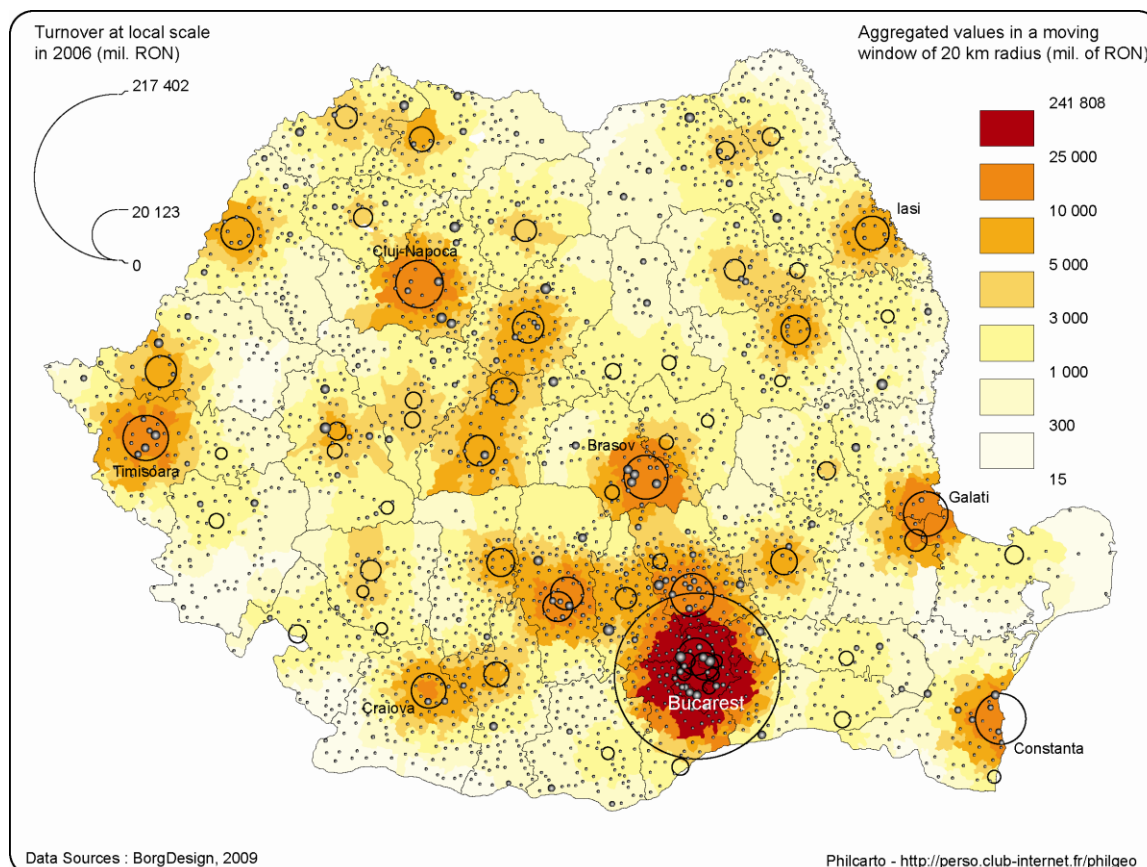
As economic actor, the firm might have different definitions from one state to another. Their number of firms is also a subject to debate, if one will take into account the variety of purposes for which firms appears in an economic system. In the Romanian case, some of the data is an average of the values for the reported year (e.g. the number of employees in 2006). For multinational firms with antennas in the local territory, the data is furnished for the LAU2 of the headquarters and not for the production of sales compartments (e.g. if the social siege is in Bucharest and the production in Iasi, the data is attached to Bucarest).

#### **SIRUTA code**

A unique identification code proposed for all the settlements in Romania. The equivalent LAU2 code is called SIRSUP (superior SIRUTA). It is sometimes a problem for the matching operations because it is different from the codes in the basemap database.

**Advantages:** a clear picture of the economic performance at local scale, especially when data is smoothed (different methods are possible). In this case, an archipelago-like territorial structure is shaped by a West-East gradient. The data can be used in order to multiply indicators: turnover per employee, local discontinuities or trans-scalar analysis.

**Disadvantages:** The data is expressed in national currency (1 RON is the approximate equivalent of 0.25 Euro) and a conversion should be made. A good knowledge of the territory should be mobilized in the map interpretation. Some of the values might look as an outlier in the territorial context – the case of Medias and Mioveni, two cities that take profit from the headquarters effect (EON GAS and Renault, two major multinational firms are located in the mentioned LAU2). The same effect could be the cause of an over-representation of Bucarest. The contextualization at an upper scale (Bulgaria and Hungary) is impossible for the moment and an eventual approach should overcome large problems of harmonization (e.g. in the case of Bulgaria, data might be available at LAU1 units).



**Figure 5** The distribution of the aggregated turnover in 2006 – Romania

In 2006, the distribution of the economic performance at LAU2 scale is an aspect of the strong metropolitan concentration. Some of the regions are better situated in this equation: the west of the country or the urban network of Transylvania. Moldavia (East) and some rural peripheral regions in the South are less visible on the map. There are several reasons for this situation: gradients of economic growth, high costs of transportation or a lack of urban economic engines. This situation is also reflected in the distribution of the welfare.

### 3.3 How to by-pass the lack of harmonization using grid information?

In ESPON DB 2013 Challenge 5 a specific methodology was developed in order to disaggregate data expressed at NUTS level in a grid (1km). The basic indicators are GDP, unemployment and active population for 2003 and 2006.



At the base of the estimation, the methodology used the distribution of population in a 1km grid and the CLC 2006 classification on the built-up area. This deliverable can be used in the analysis of the economic performance at local scale in two ways: mapping the GDP 2006 in the grid or aggregating the data in the LAU2 frame. In the second case, we will obtain an estimation of the distribution of GDP at the scale of the municipalities.

### **3.4 Integrating disaggregated data (GDP at NUTS 3 level expressed in a grid of 1km) on the LAU2 geometry frame – a problem of spatial matching.**

Despite the number of elements in the grid that have to be manipulated, the technical and methodological operations are not as complicated as one might expect. There are at least two ways to deal with this problem. The first one is to intersect every cell of the grid with the LAU2 geometry. In that case, we will obtain a large number of new polygons that contain information (GDP). Weighting the surface of the new polygons with the values for the requested indicator will approximate the share of GDP in every polygon. However, this method is time and computer resources consuming and it involves the manipulation of many spatial objects. It also has the advantage of the best precision in the estimation (to debate). A second option consists in the bypassing of the surface weight estimation and the use of the centers of every grid cell. As a matter of fact, the 1km grid cell is just a geometrical container of the information that can be reduced to one point. Intersecting the values of the point (GDP) with the LAU2 spatial frame will allow us to estimate (by summarization) the GDP at local scale. From a certain point of view, both methods are useful, it just depends on the context in which we apply them. The first one is suitable for regional research, the second for a large amount of LAU2.

#### **Data aggregation from the 1 km grid. 2<sup>nd</sup> method**

1. Don't expect to finish very soon the aggregation for all the ESPON space. First, choose your region.
  2. Create by dissolving the LAU2 limits a mask for the studied territory.
  3. Extract by this mask the grid cells that present interest.
  4. Create by extraction or selection the LAU2 map of the area. Make sure to have the same time reference as the indicator (e.g. 2006 layer map for GDP 2006 indicator)
  5. Create centers of the grid cells. You should obtain a dot/point map. Be sure to join the information to this new spatial frame (GDP or other indicators).
  6. Intersect these centers with the LAU2 map of the area. A large file containing both the codes and values of the grid and the codes and values of the LAU2 shall be created.
  7. Summarize data using the LAU2 code. A new table shall be created. This table should be saved. Else, go to Step no.1.
  8. Join this table to the LAU2 spatial frame and map the result.
- Choosing one method or another, the specific issues of spatial matching cannot be easily overcome. One major problem is to balance the degree of

generalization of the grid data representation with the geometry of the LAU2. The spatial frame for 2006 is not always perfectly overlapping the **CLC 2006** grid in which the information (GDP values) is contained. This problem is especially occurring **in the border or coastal zones**, making the estimation difficult. It is also a problem for areas with a high fragmentation of the administrative frame (the municipalities are represented as tiny polygons, intersecting 3 or 4 grid cells). The Czech Republic and Bulgaria are good examples for this last issue. Another spatial matching issue is occurring when we have to take into account the features of the natural environment. Many large LAU2 have lakes on their territory or areas with high elevation. The grid estimation method doesn't exclude them, even when their surface is limited. However, we can assume that even these areas present a contribution to the GDP creation by some economic activities taking place in relation with these zones. Sometimes the grid model penalized them (in regions with touristic vocation), sometimes not (when the natural areas present a considerable human pressure).

A zooming illustration (fig.7) of these spatial matching methodological problems is presented in the cross-border region of Dobrogea, situated on the shore of the Black Sea (Romania and Bulgaria). The map indicates that the distribution of the GDP in the grid could also be the subject of a strong territorial auto-correlation effect, induced by the NUTS0 or NUTS3 regions. On other cross-border area (Romania and Hungary or Czech Republic and Slovakia) this effect is less present in the mapping result.

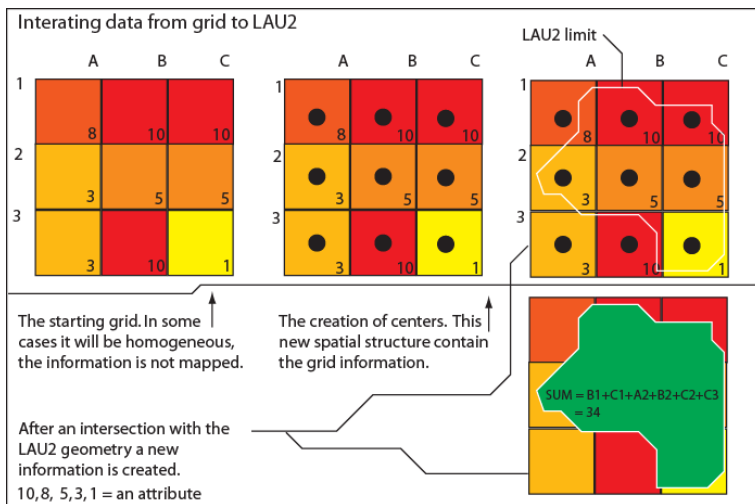
### **Methodological solutions**

In the **border or coastal zones** the data integration from grid to LAU2 is more difficult. For the border we can apply a solution based on the weighted surface of the cell intersected by the frontiers. As long as the shape of the frontier is the result of a dissolve option in a GIS, we have a good degree of shape precision. The calculated values can then be included in the file.

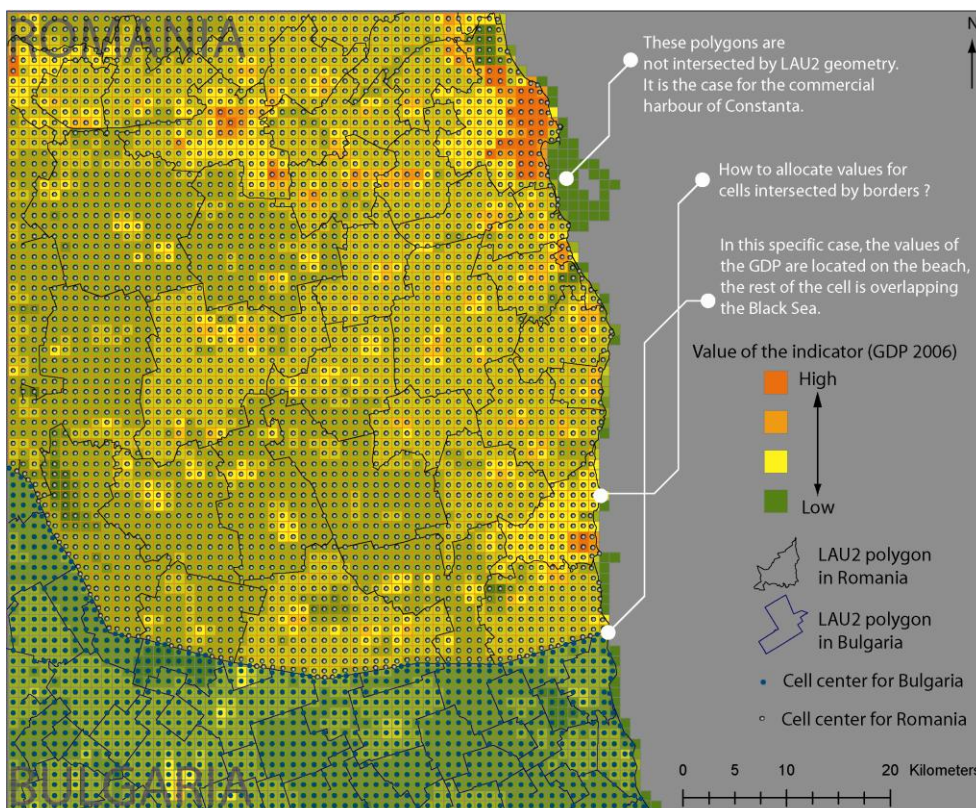
For coastal areas a solution based on buffers may be imagined. For every LAU2 that has a limit on the sea we can create a buffer of n km and calculate the values of the indicator. We are now looking for a solution able to optimize and accelerate these two methodological options.

### **Land use data integration**

Using the spatial analysis GIS methods, efforts were made in order to integrate CLC 2003 and 2006 data in the LAU2 geometries. This operation involved the intersection between the CLC 2003 classes and the LAU2 frame. The eventual output of this method would be a LAU2 typology based on the internal land use. Sensible to map projection and map scale, the measurement provides some errors. Taking into account the differences between the CLC 2003 classes and the national land use typologies, these errors are difficult to correct. Due to the large amount of time involved in this operation, the CLC data integration is available for only 5 countries.



**Figure 6** Graphic representation of grid data integration in the Lau2 geometry



**Figure 7** An illustration of the methodological issues induced by the grid data aggregation at LAU 2 scale

The main question behind this map is how far we can go with the estimation of values at local scale. The harbor of Constanta is serving the city itself, the NUTS 3 of Constanta, Bucarest and Romania. As a matter of fact, it is not an object that should be excluded from a larger spatial context. Its infrastructure is overlapping a nearby LAU2 (Agigea), a municipality almost completely integrated in the metropolitan area of Constanta. As in any port, a constellation of firms are located and they do contribute to the GDP creation. Some of them locate in Constanta, some of them in the LAU2 of Agigea. In the last case, firms are locating in order to take profit from the fiscal advantages of the free-tax zone active in the port. The both categories of economic actors use the port's infrastructure and depend on the metropolitan links created in

the recent period. Allocate the GDP from cells to one LAU or another is an attempt to perfectly map the trees and not the forest. As a compromise conclusion, the GDP is created in the area of Constanta and not in the municipalities. This area can be defined as a potential LAU1 region (neither LAU2, neither NUTS3, neither FUA, but relevant for a better integration of the local data) or it can be estimated by quantitative modeling. In Romania and some other ESPON states the LAU1 administrative level of data collection is absent. The only remaining option is to "smooth" the data by a potential of interaction model.

The integration of the grid data in the LAU2 frame was performed for 5 selected countries: Czech Republic, Slovakia, Hungary, Romania and Bulgaria. The operation involved more than 20 000 LAU2 and this considerable amount of spatial units is caused by the high degree of administrative fragmentation in 4 of the 5 states (if we also take into account the western part of Hungary). As we have integrated the 2006 values of the indicators in the 2006 LAU2 geometry, some corrections were necessary because all the capitals (except for Sofia and Prague) and some major Slovakian cities (Kosice) are divided in spatial units with LAU2 administrative competences. In the case of Bucarest, we have 6 sectors. Consequently, the data was summarized according to the most central division of the capitals, usually the sector no.1 or the Staro Mesto (the old center). A second step in the data exploration was to map the result in order to verify if extraordinary values or errors are interfering with the methodology that we used. This step works like an "expert opinion" validation of the collected or integrated data, but it is based on the visual survey and it has obvious limits. The mapping options are basically limited – symbols or choropleth design. In both cases the mass effect will be present on the map and by mass effect we understand the high distance that separates the capitals and the large cities from the rest of the urban or rural entities included in the settlement's hierarchy. A better visualization will occur if the data is smoothed (GDP in 2006 at local scale, unemployment or active population) and this better visualization is needed if one will think that we are working with the local data and not with the cities. Smoothing the data also involves a choice to be made between the methods: the average values within a rank 1 neighborhood, the summarized values in a cut-off distance neighborhood or a model based on the potential of interaction of all the LAU2 units. Some of the options take space into account in a discrete way, the last one (the model of spatial potential) in a continuous manner.

### **Building the model**

After the inclusion of the grid data in the LAU2 frame, we have prepared a model of potential of interaction that usually works when the values of the parameters are well approximated. The three variables that we had to take into account are: the interaction function, the distance decay function and the mass (the GDP in 2006 estimated for all the LAU2 in the 5 selected countries). In a classical model the interaction function and the distance decay are based on some constant values. In our model the two variables are receiving multiple

values, being weighted with the demographic rank and with the road network density in the area.

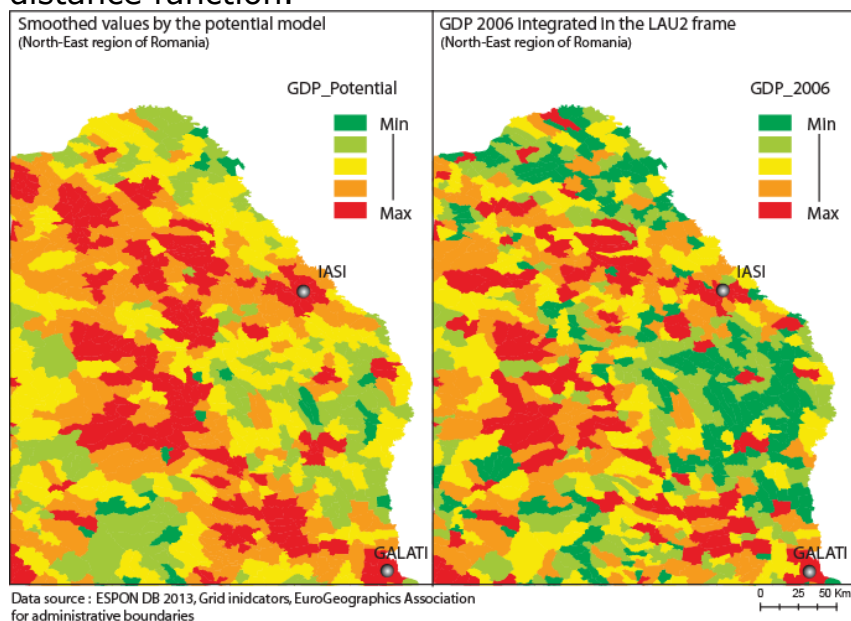


**Figure 8** A graphical formalization of the model

d1. Radius of a certain distance for a LAU2

d.2 Radius of a certain distance for a LAU2 with a larger mass.

The black and the blue lines simulate the decrease of the interaction as a distance function.



**Figure 9** Smoothed values for GDP\_2006 at LAU2 scale vs. the estimated values of GDP\_2006

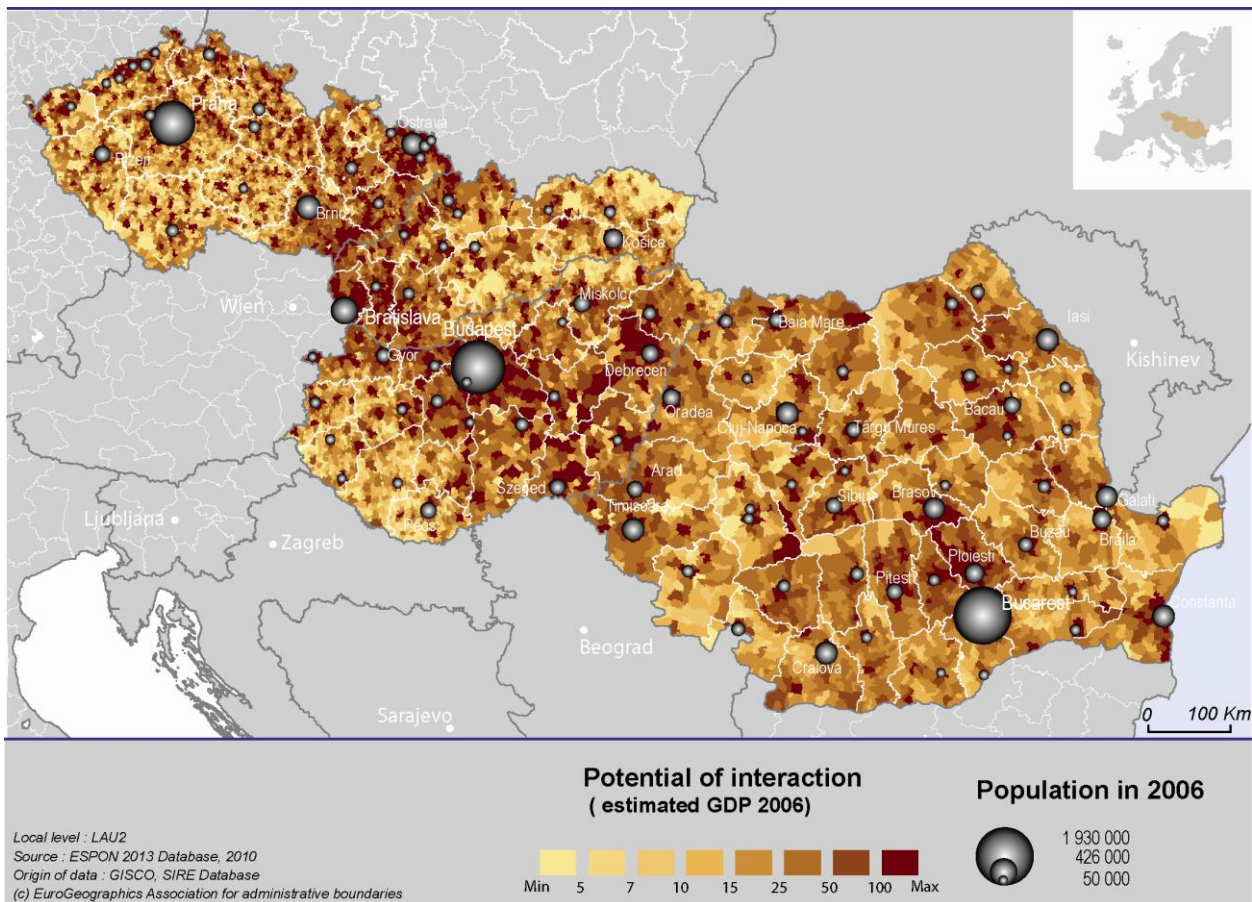
### 3.5 The calculation of the potential model of interaction for the local estimated GDP

In a classical model of potential interaction, we assume that at a "certain distance" from a spatial unit the interaction decreases at 50 % (if we use a Gaussian negative function). We also take into account the friction of the space by using a distance decay function with 2 as exponent (the canonical 2). When we apply the model we can calculate for every single spatial unit how much "interaction" will receive, if our assumptions approximate well the reality. We can better translate this into English by using an example. Let's assume that we want to calculate the potential interactions for Prague. First, we will assume that all the interactions are reduced to half at 20 km. This means that 50 % of the flows are collected in a circle with 20 km radius. The other half, obviously beyond 20 km. We will also take into account the friction induced by space (we will use 2 as value, not too much friction according to the classical models). If one will try to validate this model for Prague using the commuters flows as an empirical validation, he will observe that we were wrong in our assumptions. Prague collects 50 % of its flows from a 45 km radius, while the distance friction is only 1.2. Unfortunately, for the GDP 2006 potential of interaction we cannot validate our assumptions because we don't have the local economical flows, but our model is also supposed to work as a smoothing method for the data. If we reproduce the calculations for Brno, we will observe that the radius is about 25 km and the distance decay has 1.8 as value. One important aspect induced by the empirical "validation" is that we cannot use constant values as parameters.

Our model was built using a variable radius for the interaction function and a variable distance decay parameter. The variability was mathematically induced by using the demographic rank as a weight for the radius. The capitals received 20 km radius (in a Gaussian function) while the other cities values vary between 15 and 1 km, according to their rank. The distance decay was calculated using the road network density (high density = low friction (1.7), low density = high friction (2.3)). The road network density was estimated by intersecting the reticular spatial structure of roads with the LAU2 geometry. After this preliminary "mise en scene" of the parameters, the model was applied for more than 15 000 LAU2. Bulgaria was excluded in the absence of a network reliable file.

With this approach the smoothed values present more interest than the basic GDP disaggregated at LAU2 scale. Our intention was to eliminate the noise from the map and leave intact a territorial structure that also allows seeing the local (Figure no). We preferred to use a spreadsheet for the calculation rather than a GIS model, because the methodology can easily be reproduced by interested users. Limited by the hardware, we were forced to implement the model in packages of 15 000 LAU2 x 500 LAU2 (until we finished them all), but other strategies are possible too. A second intervention in the data was necessary, this time in order to allow the mapping process – the values were standardized using the limit of the second hierarchical class (100), letting all the superior values floating to the maximum (Budapest). There are two ways

to interpret the cartographic result. A first strategy will try to seek for the outliers and the extraordinary values and the second one involves the mobilization of the spatial structure concept.



**Figure 10** The potential of interaction for the GDP estimated at local scale in 2006

## Key findings from the map

### The economic performance is a matter of scale and mass (demography and surface).

After modeling the distribution of the GDP 2006 at local scale we can observe that the spatial pattern of this repartition is influenced by the size of the LAU2. In Hungary and Romania this phenomena is clearly visible. At the opposite, areas with high administrative fragmentation (in Slovakia or in Czech Republic) seem to be penalized even after the data smoothing. This regularity is less present in the East of the Czech Republic and in the northern region of Bucharest, two zones specialized in industrial activities.

### Long run trends are still present – the key role of the modern industrial regions

After the transition period, the industrial regions (some of them old, some of them emerged or modernized during the socialist period) seem to regain a comfortable position in the GDP hierarchy. At local scale, these industrial basins are extremely important if one will take into consideration their impact

on employment or in the welfare creation. These regions (e.g. Ostrava in Czech Republic, Győr in Hungary, Pitești-Ploiești in Romania) are there to complete the metropolitan economic nodes and their conjuncture fragility is balanced by resilience, adaptation and integration in the European economy.

**Spatial discontinuities and interfaces: when frontiers are uniting trends of economic performance.**

Transforming the frontiers in interfaces that filter the flows of persons, goods and information is a constant trend in the Eastern European countries. In some cases, these frontiers may also work as attractors for the economic activities, such is the case between Romania and Hungary or partially between Slovakia and Hungary. The real discontinuities in the local estimated GDP 2006 distribution seem to be internal, sometimes overlapping old historical limits (Moldavia and Transylvania for Romania). Maybe the new economic paradigm that installed in the transition and pre-adhesion period reactivated these old frontiers, shaping new logics of economic performance compared to the past.

**Some of the remote areas are quite well. The regions in difficulty are still precise.**

Without being a rule, some of the remote regions and areas with specific geographical features are not marginalized in the distribution of the economic performance (the coastal regions are quite dynamic, some of the mountain areas in the Carpathians present decent values of GDP due to recent turistification and re-industrialization and some of the border regions behave as economic attractors). The regions in difficulty locate (without being a clear regularity) in the “no man’s land” of the metropolitan and urban polarization, making us to assume that the economic performance could be a distance decay function towards the nearest city.



## **4 Towards more elaborate indicators and models applied to LAU2 objects. The settlement hierarchy and the territorial architecture of the selected countries.**

### **4.1 Using the distance as an indicator for territorial coherence in the Eastern Europe**

Distance is still an important component that shapes the local space in the selected countries (it filters the flows, it may explain how the economic performance is distributed or how the urban network is functioning). Working with the distances at LAU2 scale is complicated if we take into account the large number of spatial units. However, not all the distances are relevant or interesting in an eventual study. If we want to compare how the estimated GDP at local scale in 2006 is distributed in relation with distance to the nearest city, we will work with a reasonable quantity of elements in the matrix (about 15000 LAU2 x almost 100 spatial units). In this case we can proceed to the calculation of distances using the regional road network. If no GIS instrument is available, the Euclidean distances can also be estimated using the classical model. Both methodologies involve a number of compulsory steps to check, in order to obtain the "oursin" map that we seek for.

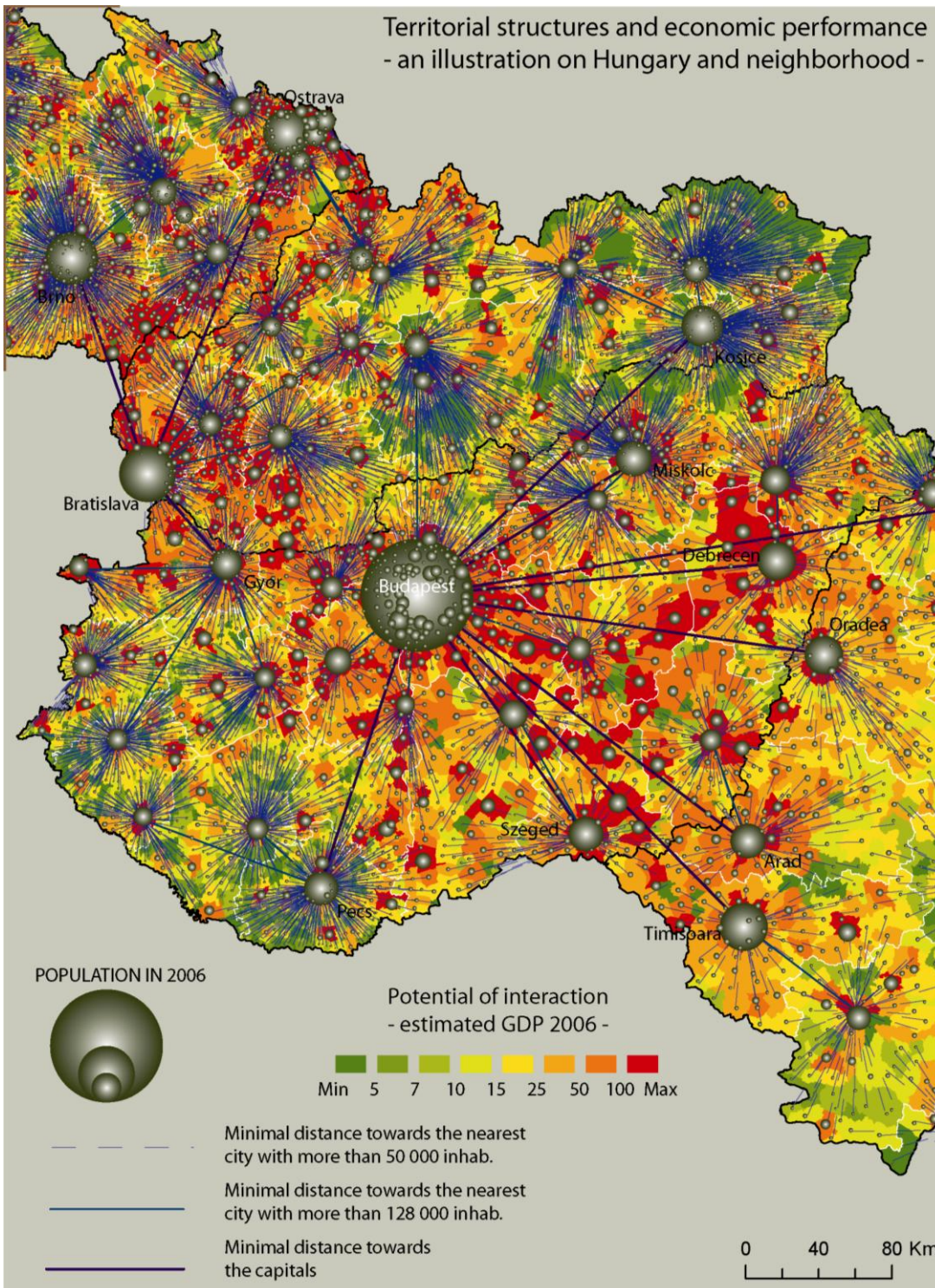
### **4.2 The settlement's hierarchy – how many levels function from Prague to Sulina?**

As our intention is to put in relation LAU2 s and cities in the selected countries (Czech Republic, Slovakia, Hungary and Romania) and as the definitions of the cities are heterogeneous, we have assumed that the starting point should be the choice of a demographic cut-off in the urban hierarchy. The analysis of the rank-size distribution for 2006 population shows that 3 hierarchical levels are clearly visible in the region: the capitals, the so called "large cities" (over 128 000 inhab.) and the medium sized cities (over 50 000 inhab.). As an alternative, we could use the distinction between FUA and MEGA, qualitatively complicating the model. These three demographic levels were put into relation with the LAU2 using the distance in the road network as an indicator for some potential and theoretical urban influence area. If there is any association between the economic performance and the distance decay based on cities, the mapping process and the interpretation should show it. We also could assume that this relation will work better in a homogeneous space, like Hungary. That's the reason for focusing our cartographic processing and argumentation on this country. Of course, if we could make the map readable.

#### **Steps to obtain a link map**

1. Every map comes from software. From the point of view of the software, not every sum of lines is a network. Manipulation is needed in order to obtain nodes and links in the network.
2. In the case of an oursin map, not all the points (LAU2, cities, features) shall be involved. The eventual user should think that what he will obtain will be an origin vs. destination asymmetrical matrix.

3. The origin vs. destination matrix will be provided in a vector format which is sometimes difficult to manage.
4. In order to obtain the minimal distance that separates a LAU2 unit from the cities (our case) an interrogation in the matrix is necessary.
5. After the interrogation is applied, a new vector matrix is obtained and it will be saved. Else, go to step no.1.
6. The data obtained from the calculation offers two mapping options: links or choropleth. The links can be used to emphasize the shape of a theoretical zone of urban influence, the choropleth its limits.



**Figure 11** Distances and economic performance in Hungary and neighborhood

### **4.3 How to map the distance: choropleth vs. "oursins".**

If one will use the minimal distances as an indicator for the territorial or administrative fragmentation at the local scale, the mapping method will be an issue to take into consideration. Using the LAU2 polygons for the cartographic product will produce a "tropical fish" like map. At the opposite, the link map will provide an agglomerated picture of the territorial structures, emphasizing the shapes rather than the limits of the theoretical zones of influence. The main advantage of the second map type (links or "oursin") resides in the opportunities that offer to hierarchically imbricate different distance levels. (e.g. LAU2 vs. nearest city with more than 50 000 inhab., nearest city with more than 50 000 inhab. with the nearest city with more than 128 000 inhab., the nearest city with more than 128 000 inhab. with the capitals). In a choropleth map only one kind of distance could be mapped, excepting the case where we use a cluster analysis. If the LAU2 used as destinations are too dense, there is a risk to make the map unreadable (which is the case for Slovakia or Czech Republic, in our illustration).

### **4.4 The local fragmentation of the territorial architecture and the hierarchical immobility.**

The relation between the economic performance and the distance towards the nearest medium (50 000 inhab.) or large city (128 000 inhab.) seems to obey to a U shaped like function rather than a classical power decay function (based on the observations made on Hungary and neighborhood). This output has two explanations: the christallerian pattern in the city distribution or errors induced by the aggregation method (GDP for 2006 in a grid format to GDP 2006 in LAU2 frame, smoothed by a potential method).

In many local cases, the shape of the theoretical urban influence area is largely overlapping the NUTS 3 limits. It is not the NUTS3 limit that should be put into question, but the limited number of destinations used by the model.

The national borders are irrelevant in the design of the capital's theoretical areas of influence. Some western cities in Romania are closer to Budapest, than Bucharest. Their local hinterland too, inducing some sensible questions regarding the equilibrium between national and trans-national public planning policies.

The context does matter and by context we can imagine the role that would play MEGA such as Vienna in the context of the Eastern and Central Europe. How the potential interaction of the GDP 2006 (locally estimated) would be reshaped, if Vienna were on the map?

In the eventuality of a recalibration of the hierarchical demographic levels (using 27 000 inhab. instead of 50 000), the relation between economic performance and distance towards the nearest city will take (maybe) another form. However, at the top of the hierarchy, the situation will present few changes and Oradea will still be closer to Budapest than to Bucharest. The hierarchical immobility will still function, it is the local that will present interesting dynamics.

The corridors of welfare are complicating the gradients and the core(s)-peripheral spatial patterns present on the map. Some of these corridors reply the major transportation network, some others the linear proximity to economic engines or consecrated MEGA. The disconnections in relation with the metropolitan nodes sometimes give them the attribute of fragility or territorial tunnel effect.

## Conclusions

These conclusions are organized as a set of relevant key findings concerning the problematic of the LAU2 scale of analysis. As the main studied area is composed by 5 countries situated in a particular geographical context, these conclusions cannot be extrapolated to other states.

At the horizon of the year 2006, the demographic decline affects an area of the size of a medium ESPON country. The process presents spatial homogeneity and has chances to become a trans-scalar and cross-border issue, relevant for policy decision.

The demographic growth is possible, even in a context of turbulent economic performance. It almost concerns only the large metropolitan areas and some regions with specific geographical features.

The elementary spatial patterns (points, polygons, networks) can be mobilized in the production of relevant indicators for the local territories. Basic techniques of spatial analysis can be implemented, even when we deal with a large number of LAU2. It is the case for the accessibility at local scale.

The economic performance at local level is a matter of scale and mass (demography and surface). As an hypothesis, it also might be linked to some patterns of spatial organization (territorial auto-correlation or the effect of territorial belonging).

The modern industrial regions still play a role in the distribution of economic performance in the Eastern Europe, together with the large metropolitan areas.

The frontiers are not always behaving as spatial discontinuities. In some contexts, the border is uniting trends of economic performance. Such is the case for the frontier between Romania and Hungary.

Some of the remote areas are quite well, despite their lack of accessibility or their under-average territorial integration. The regions in difficulty are still to be precised, but using a combination of indicators derived for the local scale (economic performance, land use patterns, accessibility and theoretical connectiveness to the metropolitan areas) will allow the development of a methodology.

## Annexes

### Land use data integration in LAU2 geometry

In some contexts, working with a large amount of data at local level is counter-productive for two reasons: data manipulation is becoming an extremely time-consuming process, even for simple operations and the level of map visibility might be compromised by the number of polygons. That's why an option was made to produce indicators only for a limited number of countries (Bulgaria, Czech Republic, Hungary and Slovakia), indicators that were previously spatially and chronologically harmonized. This space covers an interesting area from Central Europe to the Balkans, allowing us to test different methodologies (data collection, indicator creation and harmonization) for sufficiently large dataset (almost 25 000 LAU2). In many cases, the indicators provided by the NSI were available only for a country or two, not allowing us a complete collection of data in order to cover the mentioned space. In this case, looking for indicators that are covering the 5 countries might present much more interest and added value compared to five different indicators which will cover separately each of the 5 countries.

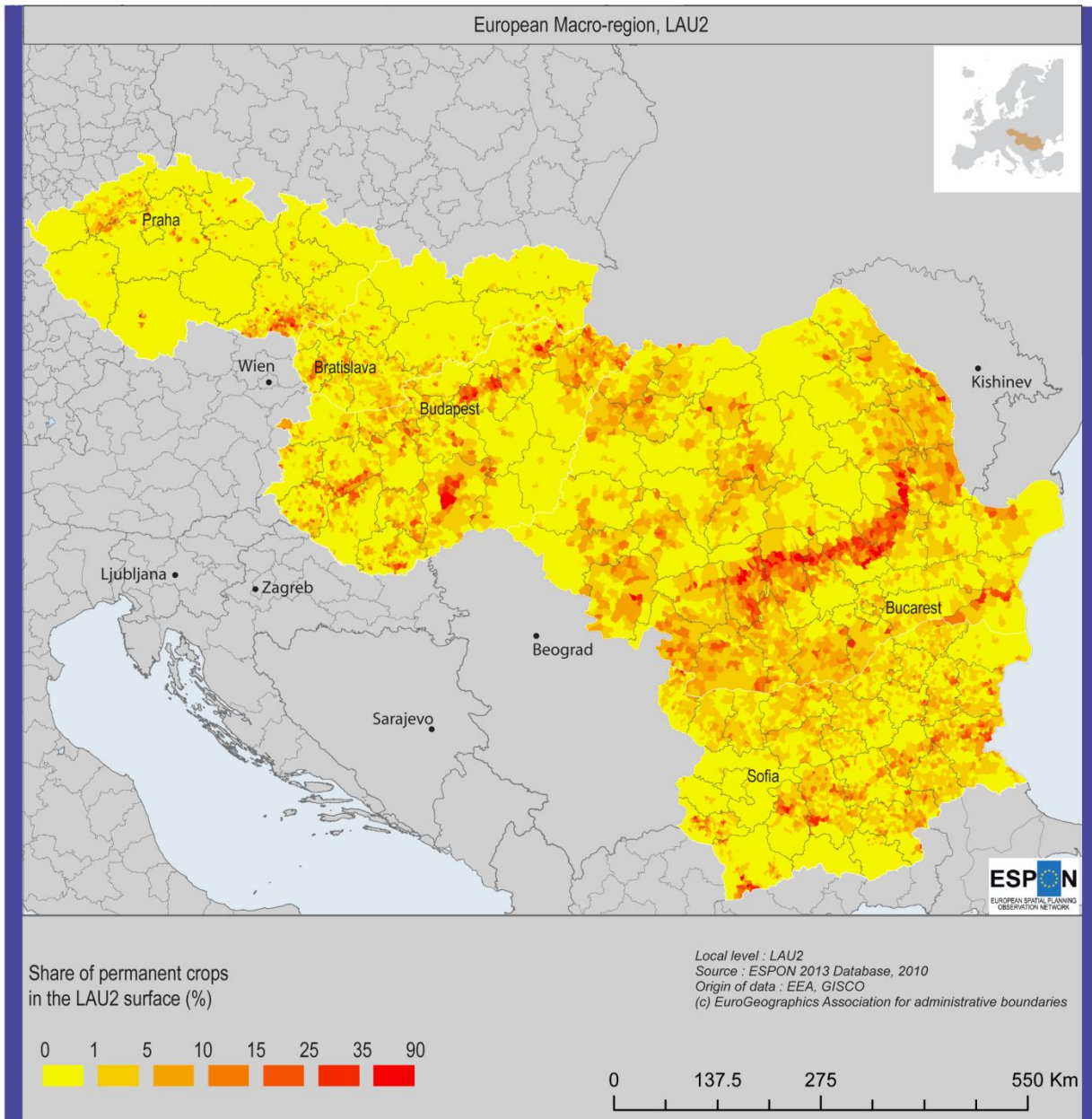
One exploratory dimension in our work was to integrate data about the land use at local level, both as absolute and relative values. Taking into account some of data collection problems that we have already mentioned, it looked naturally to explore the possibility to integrate data from the European Environment Agency (the CLC 2006 vector data) in the LAU2 geometry. The methodological steps are not extremely complicated but (again) time consuming:

- 1) Data download – 44 layers of information were extracted (<http://www.eea.europa.eu/data-and-maps/data/clc-2006-vector-data-version>).
- 2) Data extraction for the 5 countries chosen for the test (GIS operation: extract by a mask/clip formed by the concerned region).
- 3) Geometry re-projection in order to obtain surfaces for LAU2 and the land use layers.
- 4) Intersection of two layers (LAU2 geometry vs. land use layer – e.g. LAU2 2006 geometry intersected with the green urban areas). The operation was implemented 35 times for each land use layer.
- 5) For the new objects, the surface was calculated and summarized for every land use category, according to the LAU2 2006 frame.

The results of this operation were systemized in compliance with the CLC2006 classification, as follow:

- 1) Surfaces in square meters for every land use category of rank 3. Relative share of a land use category as percentage in the LAU2 surface. (e.g. C221 represents the vineyards surface in the LAU2 and the CREL221 the surface of vineyards as % in the LAU2 geometry).
- 2) Surfaces in square meters for every land use category of rank 2. Relative share of a land use category as percentage in the LAU2 surface.

- (e.g. C22 represents the permanent crops surface in the LAU2 and the CREL22 the surface of the permanent crops as % in the LAU2 geometry).
- 3) Surfaces in square meters for every land use category of rank 1. Relative share of a land use category as percentage in the LAU2 surface. (e.g. C2 represents the agricultural areas in the LAU2 and the CREL2 the surface of the agricultural as % in the LAU2 geometry).



**Figure 12** Share of the permanent crops as percentage in the LAU2 surface in 2006.

The table is populated with 70 indicators for the first type of land use (rank 3), 30 indicators for the second one and 10 for the rank 1. Some categories are not represented for obvious geographical reasons (olive crops, glaciers or intertidal flats). Concerning the utility and the data quality, there are some aspects to emphasize and discuss. Eventual errors in the data table might occur from the combination of several factors: the CLC methodology of classification, the degree of generalization for the LAU2 polygons and the projection of data in the system of coordinates. Consequently, the values of

the land use categories at LAU2 2006 level should be considered as approximations of the reality.

Despite the data quality (which might be questionable), there is a double interest related to these indicators and to this methodology. From a theoretical point of view, having access to a general picture of the land use offers some opportunities to create new indicators (e.g. location coefficients for the artificial spaces), to provide or to refine spatial typologies or to better define territories with specific geographical features. If we take the example of the permanent crops (vineyards and fruit trees), the map will allow us to better understand how patterns of "stable" agriculture are organized in this European macro-region. Being quite and exigent form of agriculture because it demands specialized labor force and superior agronomic skills, the permanent crops may play a key role in the economic development for some privileged rural areas (Sub-Carpathians in Romania or Tokay region in Hungary). In the same time, a high degree of specialization might become problematic in contexts of economic instability or in confrontation with natural hazards. This kind of indicator, combined with other datasets available at local scale (accessibility, economic performance), can be used in order to better shape territories with specific economic vocation, taking thus into account two policy relevant aspects: sustainability and vulnerability.

### **The estimation of the local economic performance**

Using a regular grid of 1 km, in the ESPON DB 2013 projects, one of the teams (UAB) successfully disaggregated indicators from NUTS3 level to this new geometry (GDP 2006 and 2003, Unemployment 2006 and 2003 and Active population 2006 and 2003). The output of this methodology served us to better explore two problems:

- a) How to re-aggregate data in an intermediary geometry (LAU2 scale)?
- b) How to build a homogeneous indicator that covers all the ESPON space?

The first issue was already presented in this report and we have observed that the optimal way to map and use the new indicators is to pass by a smoothing process. The smoothing was already implemented for 4 countries (CZ, SK, HU, RO), in the case of the GDP 2006 indicator. A second attempt was recently made to estimate the local economic performance (based on the GDP 2006 values in a regular grid) for all the countries in the ESPON space (with three exceptions – CH, NO, IS).

The working methodology for this estimation follows several main steps:

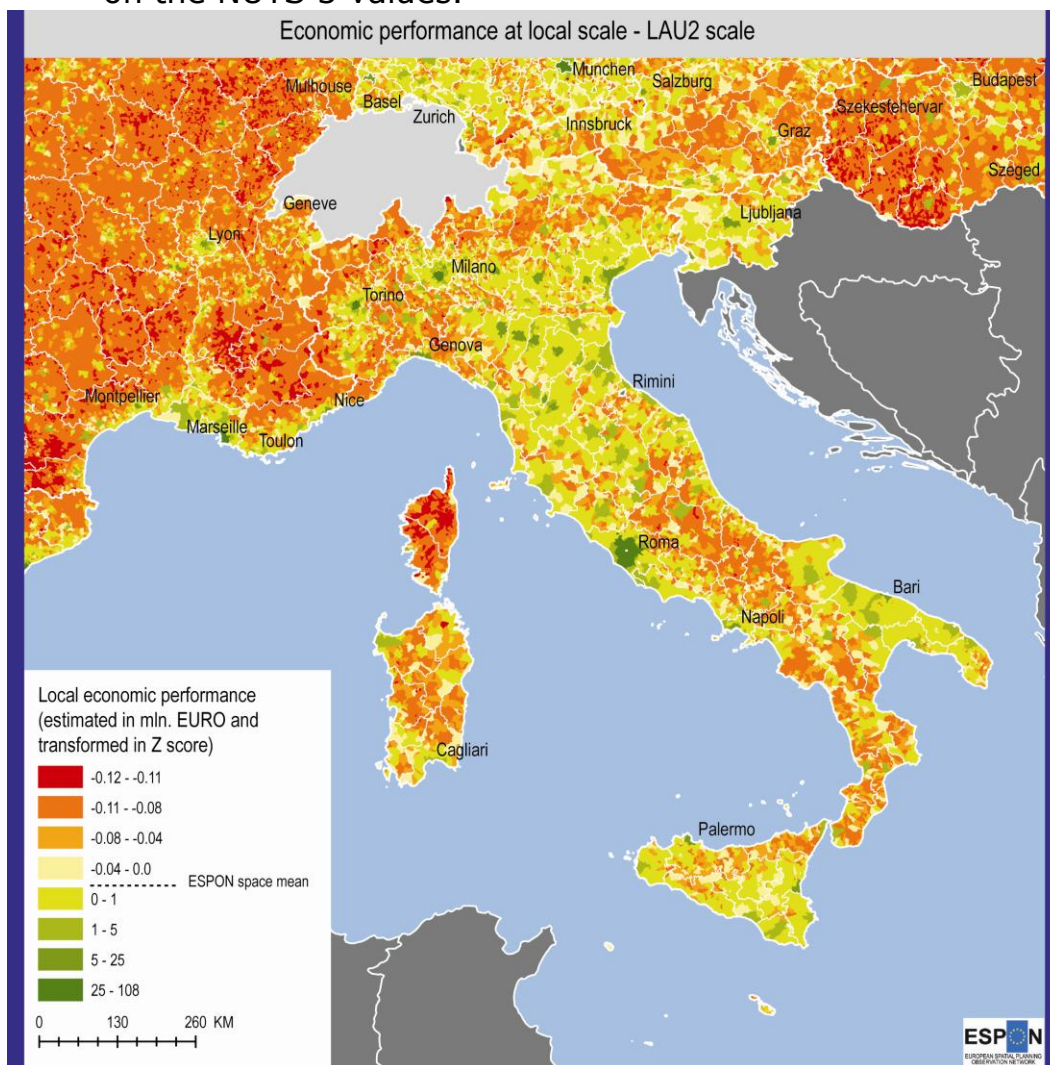
- 1) The grid with the GDP 2006 estimated indicator was split in 31 new features, using the country boundaries as a clip.
- 2) The new features were intersected with the LAU2 geometry for each state (31 operations)
- 3) The area of each cell in the new grids was calculated in order to better estimate the value of GDP 2006 that must be allocated to the LAU2
- 4) The data was summarized using the LAU2 code and merged in a new table that covers the ESPON space.

In this point, several cartographic explorations were made, trying to check if the local economic performance (estimated) is veridical.



Based on the cartographic explorations, there are some issues to signal:

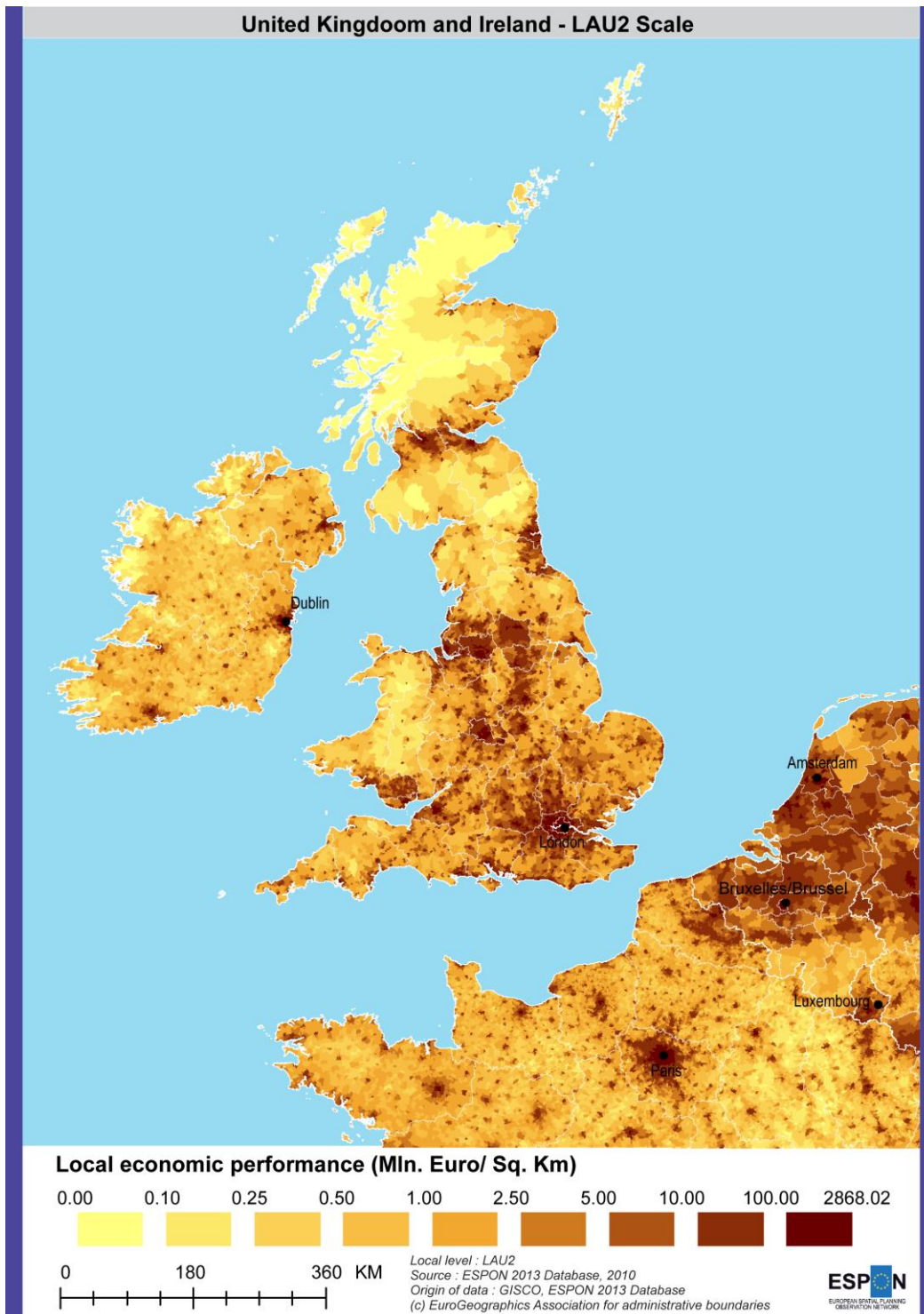
- 1) The local economic performance estimated for every LAU2 (expressed in millions of EURO for 2006) is strongly depending on the LAU2 area. Theoretically, a bigger LAU2 will be richer than a small one. As an example, this aspect penalizes the French "communes".
- 2) The transformation of data (a Z score was built to better shape the areas of local economic performance) don't reduce the geometrical effect. However, it privileges the urban areas and the engines of growth.
- 3) A strong territorial auto-correlation effect is interfering with the map process. This effect is impossible to be reduced, if we take into account the fact that the grid data estimation provided by the UAB team is based on the NUTS 3 values.



**Figure 13 Local economic performance in 2006 - Italy**

Having these three issues in mind, we have derived a new indicator of local economic performance – the LAU2 values being expressed as a density, in millions of Euro/sq. km. These new values reduce the territorial auto-correlation effect and link the estimated local economic performance to the distribution of population at LAU2 scale (Fig. 14). Probably, a smoothing method applied to these values will compensate all the issues described and will provide a much more interesting indicator. There are some technical limitations to this smoothing method, limitation which are induced by the large number of features that are present in the table and on the map.

Another aspect to signal here is related to the division of the large metropolitan areas in LAU2 polygons (such as Paris or London). Building a top of the economic performance will present some surprises, from this point of view.



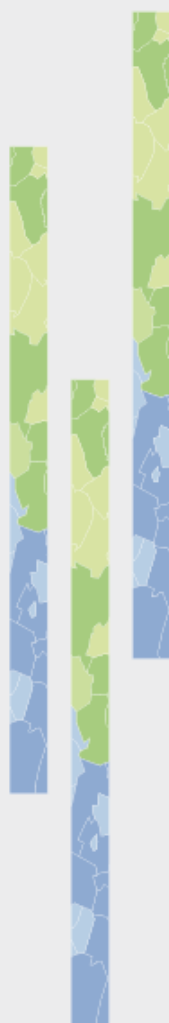
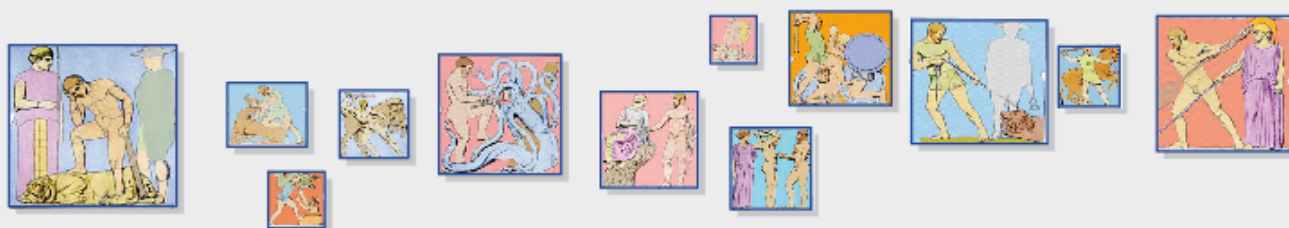
**Figure 14 Local economic performance in 2006 – UK and Ireland**

## List of indicators – rank 3 land use categories at LAU2 scale

Indicator	CLC 2006 - LABEL3	CLC 2006 - LABEL1
C111	Continuous urban fabric	Artificial surfaces
C112	Discontinuous urban fabric	Artificial surfaces
C121	Industrial or commercial units	Artificial surfaces
C122	Road and rail networks and associated land	Artificial surfaces
C123	Port areas	Artificial surfaces
C124	Airports	Artificial surfaces
C131	Mineral extraction sites	Artificial surfaces
C132	Dump sites	Artificial surfaces
C133	Construction sites	Artificial surfaces
C141	Green urban areas	Artificial surfaces
C142	Sport and leisure facilities	Artificial surfaces
C211	Non-irrigated arable land	Agricultural areas
C213	Rice fields	Agricultural areas
C221	Vineyards	Agricultural areas
C222	Fruit trees and berry plantations	Agricultural areas
C231	Pastures	Agricultural areas
C242	Complex cultivation patterns	Agricultural areas
C243	Land principally occupied by agriculture, with significant areas of natural vegetation	Agricultural areas
C311	Broad-leaved forest	Forest and semi natural areas
C312	Coniferous forest	Forest and semi natural areas
C313	Mixed forest	Forest and semi natural areas
C321	Natural grasslands	Forest and semi natural areas
C322	Moors and heathland	Forest and semi natural areas
C324	Transitional woodland-shrub	Forest and semi natural areas
C331	Beaches, dunes, sands	Forest and semi natural areas
C332	Bare rocks	Forest and semi natural areas
C333	Sparsely vegetated areas	Forest and semi natural areas
C411	Inland marshes	Wetlands
C412	Peat bogs	Wetlands
C421	Salt marshes	Wetlands
C422	Salines	Wetlands
C511	Water courses	Water bodies
C512	Water bodies	Water bodies
C521	Coastal lagoons	Water bodies
C523	Sea and ocean	Water bodies
CREL111	Continuous urban fabric (share in% at LAU2 level)	Artificial surfaces
CREL112	Discontinuous urban fabric (share in% at LAU2 level)	Artificial surfaces

CREL121	Industrial or commercial units (share in% at LAU2 level)	Artificial surfaces
CREL122	Road and rail networks and associated land (share in% at LAU2 level)	Artificial surfaces
CREL123	Port areas (share in% at LAU2 level)	Artificial surfaces
CREL124	Airports (share in% at LAU2 level)	Artificial surfaces
CREL131	Mineral extraction sites (share in% at LAU2 level)	Artificial surfaces
CREL132	Dump sites (share in% at LAU2 level)	Artificial surfaces
CREL133	Construction sites (share in% at LAU2 level)	Artificial surfaces
CREL141	Green urban areas (share in% at LAU2 level)	Artificial surfaces
CREL142	Sport and leisure facilities (share in% at LAU2 level)	Artificial surfaces
CREL211	Non-irrigated arable land (share in% at LAU2 level)	Agricultural areas
CREL213	Rice fields (share in% at LAU2 level)	Agricultural areas
CREL221	Vineyards (share in% at LAU2 level)	Agricultural areas
CREL222	Fruit trees and berry plantations (share in% at LAU2 level)	Agricultural areas
CREL231	Pastures (share in% at LAU2 level)	Agricultural areas
CREL242	Complex cultivation patterns (share in% at LAU2 level)	Agricultural areas
CREL243	Land principally occupied by agriculture, with significant areas of natural vegetation (share in% at LAU2 level)	Agricultural areas
CREL311	Broad-leaved forest (share in% at LAU2 level)	Forest and semi natural areas
CREL312	Coniferous forest (share in% at LAU2 level)	Forest and semi natural areas
CREL313	Mixed forest (share in% at LAU2 level)	Forest and semi natural areas
CREL321	Natural grasslands (share in% at LAU2 level)	Forest and semi natural areas
CREL322	Moors and heathland (share in% at LAU2 level)	Forest and semi natural areas
CREL324	Transitional woodland-shrub (share in% at LAU2 level)	Forest and semi natural areas
CREL331	Beaches, dunes, sands (share in% at LAU2 level)	Forest and semi natural areas
CREL332	Bare rocks (share in% at LAU2 level)	Forest and semi natural areas
CREL333	Sparsely vegetated areas (share in% at LAU2 level)	Forest and semi natural areas
CREL411	Inland marshes (share in% at LAU2 level)	Wetlands
CREL412	Peat bogs (share in% at LAU2 level)	Wetlands
CREL421	Salt marshes (share in% at LAU2 level)	Wetlands
CREL422	Salines (share in% at LAU2 level)	Wetlands
CREL511	Water courses (share in% at LAU2 level)	Water bodies
CREL512	Water bodies (share in% at LAU2 level)	Water bodies
CREL521	Coastal lagoons (share in% at LAU2 level)	Water bodies
CREL523	Sea and ocean (share in% at LAU2 level)	Water bodies

**(Source of labels : CLC 2006 legend ; <http://www.eea.europa.eu/data-and-maps/data/clc-2006-vector-data-version/corine-land-cover-2006-classes>)**



# Naming UMZ: a database now operational for urban studies

## MAIN RESULTS

- From physical zones to urban settlements. Currently, the UMZ (CLC 2000, EEA) are not described by a name. This attribute is yet essential for creating a semantic link to the territory.
- Automatic algorithm. According to the way an UMZ overlaps reference units (LAU or other), a new method attributes automatically one or several names.
- Validation. A final check is done automatically, by comparing results to other European database names. Some particular cases are corrected by expertise.
- A first thematic insight. An exploration of the main features is proposed in the last part. New results are given concerning the European city size distribution, the general and regional density patterns, and the main characteristics of international UMZ.

**ESPON 2013 DATABASE**  
**MARCH 2011**





# LIST OF AUTHORS

Anne Bretagnolle, University Paris 1, UMR Géographie-cités

Marianne Guérois, University Paris 7, UMR Géographie-cités

Guilhain Averlant, UMR Géographie-cités

Hélène Mathian, C.N.R.S., UMR Géographie-cités

François Delisle, UMR Géographie-cités

Liliane Lizzi, C.N.R.S., UMR Géographie-cités

Timothée Giraud, UMR Géographie-cités, UMS 2414 Riate

## **Contact**

anne.bretagnolle@parisgeo.cnrs.fr

tel. + 33 1 01 40 46 40 00

# TABLE OF CONTENT

<b>LIST OF AUTHORS .....</b>	<b>3</b>
<b>1 STAKES AND MATTER .....</b>	<b>5</b>
1.1 PRESENTATION OF UMZ.....	5
1.2 FROM PHYSICAL ZONES TO URBAN SETTLEMENTS ?.....	6
1.3 A NEW VERSION OF UMZ DATA BASE.....	6
<b>2 NAMING METHODOLOGY .....</b>	<b>7</b>
2.1 AUTOMATIC ALGORITHMS .....	7
2.1.1 General presentation.....	7
2.1.2 Algorithm steps and illustrations.....	7
2.1.2.1 Geometrical and statistical sources .....	8
2.1.2.2 Computation steps.....	8
2.1.2.3 A particular case: different UMZ with identical names .....	11
2.2 AUTOMATION OF PROCESSING CHAIN .....	14
2.2.1 The need for an automated process .....	14
2.2.2 Process description and software solution .....	14
2.3 EXPERTISE AND VALIDATION .....	16
2.3.1 Selecting relevant reference units.....	16
2.3.1.1 LAU1 instead of LAU2.....	18
2.3.1.2 NUTS instead of LAU .....	18
2.3.1.3 National settlement areas.....	19
2.3.2 Countries without population density grid .....	20
2.3.3 Validation process .....	20
2.3.3.1 Sources.....	20
2.3.3.2 Typology of errors .....	21
2.3.3.3 Solutions proposed .....	21
2.4 RESULTS : TYPOLOGY OF NAMING SITUATIONS .....	22
<b>3 A THEMATIC INSIGHT INTO EUROPEAN CITIES.....</b>	<b>25</b>
3.1 URBAN HIERARCHY AND CITY-SIZE DISTRIBUTION .....	25
3.2 DENSITY PATTERNS .....	26
3.3 INTERNATIONAL UMZ .....	29
<b>4 CONCLUSION.....</b>	<b>33</b>



# 1 Stakes and matter

*Urban Morphological Zones* have been created in 2004 by the European Environment Agency. This data base forms a perspective for the future, for three main reasons: it is constructed using highly automated methods, it is regularly updated (two dates are now available for UMZ perimeters, 1990 and 2000 and the 2006 version will be soon available) and it is fully documented.

This database has however not been widely used to date in urban studies, mainly because it is not operational: the objects are simply spots or patches, without names, and hence without semantic links with the territory. They only constitute a set of geometrical objects, and not of geographical objects. The general aim of this Technical Report is to describe the automatic methods and expertises that have been used for naming UMZ and getting them usable for a first exploration of the European urban settlements.

## 1.1 Presentation of UMZ

UMZ have been created in order to analyze "the extent of urban land-take in Europe, where sprawl happens and how it is shaped" (*EEA activities*, <http://www.eea.europa.eu/themes/urban/eea-activities>). An UMZ can be described as "a set of urban zones, defined from land cover classes contributing to the urban tissue and function", forming a continuous built-up area (i.e. laying less than 200 m. apart)<sup>1</sup>.

Since September 2009, the geographical coverage of the UMZ 2000 database is the following one:

- the 27 countries of the European Union
- 5 countries in the Balkan region (Albania, Bosnia-Herzegovina, Kosovo, Macedonia and Serbia)
- Norway, Lichtenstein and Island<sup>2</sup>.

The UMZ dataset can be downloaded freely on EEA website<sup>3</sup>. Different attributes are available:

- Identification code (not the same than for UMZ 1990)
- Population (estimated from JRC's Population density grid, see Javier Gallego, *Joint Research Center*)<sup>4</sup>.
- Area and perimeter

---

<sup>1</sup> Urban Morphological Zones 2000 Version F1v0. Definition and procedural steps, Roger Milego, February 2007, <http://dataservice.eea.europa.eu/dataservice/metadetails.asp?id=995>.

<sup>2</sup> CLC2006 should also cover Switzerland.

<sup>3</sup> <http://www.eea.europa.eu/data-and-maps/figures/urban-morphological-zones-umz-2000>

<sup>4</sup> For further details, see Downscaling population density in the European Union with a land cover map and a point survey, <http://www.eea.europa.eu/data-and-maps/data/population-density-disaggregated-with-corine-land-cover-2000-2>.

## 1.2 From physical zones to urban settlements?

Going from physical zones to urban settlements is not a trivial operation. For example, overlaying on a GIS the UMZ and Google Earth or LAU 2 names is not sufficient: if we make a zoom on Berlin surroundings, it will be easy to put the "Berlin" name on the UMZ whose centroid is the closest to the historic center, but several difficult questions remain:

- Should other close UMZ inside the same LAU 2 receive also the name "Berlin" or another one (for instance, more local names given by Google Earth)?
- What are the reference units for choosing the right names? In most of the cases, city names fit with LAU 2 names (for instance in France, Germany or Belgium, as the eponym name of city fits with the central municipality). But in some other cases, like in Portugal, Greece or Denmark, city names fit with LAU1 names. And it is even more complicated in United Kingdom or Ireland, where city names don't fit with one only administrative level but with other administrative entities.
- How can we manage the case of polycentric cities, like industrial or littoral conurbations? Should they receive several names, for instance when the population is well distributed among the different cores, or just one name?
- What about large cities cases, which are extending now at the scale of one NUTS 3 rather than LAU 1 or 2? Should we give them the name of the eponymous LAU 2 or the name of the region that currently fits with their spatial coverage?
- How ensuring a quick update of UMZ names, facing the evolution of perimeters (corrections or new dates), the evolution of population density grid (JRC), or the need to apply the methods to smaller objects (in the current data base, names are given only to UMZ larger than 10 000 inhabitants, i.e. less than 50 % of the total number of UMZ)?

The answers given to these different questions are discussed and fully described and illustrated in the following sections of the Technical Report.

## 1.3 A new version of UMZ data base

Different adjustments have been made to UMZ database in order to facilitate its use by ESPON partners. UMZ larger than 10 000 inhabitants have been considered (a total of 4437 UMZ).

- Updated Population: using automatic methods, we have updated the population of all the UMZ with the last version (v.5) of the Population density grid built in 2007 by *Joint Research Center*<sup>5</sup>. The scale used for this grid is 100x100 meters.
- New indicators: Name(s), Centroid<sup>6</sup>, Density (inh./km<sup>2</sup>), Country<sup>7</sup>, International code (number of countries crossed by the UMZ), International index (% of population not living in the main country).

---

<sup>5</sup> Gallego J., 2007; Downscaling population density in the European Union with a land cover map and a point survey, <http://dataservice.eea.europa.eu/dataservice>.

## 2 Naming methodology

### 2.1 Automatic algorithms

#### 2.1.1 General presentation

The methodology that has been chosen is largely inspired by the one used by French Census Board (INSEE) to give names to French urban areas (*unités urbaines*)<sup>8</sup>. Rules and criteria have been elaborated to differentiate three types of spatial configurations resulting from the overlap of UMZ data base, Population density grid and the reference units data base (i.e. the data base that has been selected for giving the names, for example LAU 2) (Figure 1).

In the first situation, the major part of the UMZ population (more than 50%) is located inside one reference unit<sup>9</sup>. The urban settlement extends rather clearly around one morphological centre, and receives one name.

In the second and third situations, no reference unit concentrates more than 50 % of the UMZ population: we retain therefore the unit that has the major contribution as the main one, then we examine the other reference units that largely contribute to the UMZ population. If they represent more than 50% of the main reference unit contribution, we retain them and the UMZ is considered as "UMZ with several cores" (Situation 2). If not, we keep only the main reference unit for naming UMZ. It is then considered as "UMZ with a weak core" (situation 3).

- *Situation 1 : UMZ with a strong core* (it receives one name)
- *Situation 2: UMZ with several cores* (it receives several names)
- *Situation 3: UMZ with a weak core* (it receives one name)

#### 2.1.2 Algorithm steps and illustrations

In order to simplify the presentation of this sub-section, the selected reference source for city names is LAU 2.

---

<sup>6</sup> The centroid is the centre of gravity computed as the average of the coordinates of all the UMZ's vertices.

<sup>7</sup> If the UMZ overlays more than one country, it is associated to the country which includes the largest part of the UMZ population (main country)

<sup>8</sup> « Composition communale des unités urbaines, Population et délimitation 1999, Nomenclatures et codes » ; INSEE, mars 1999.

<sup>9</sup> We have retained, like INSEE, the minimal threshold of 50% inhabitants, which gives rather goods results (see validation section below).

The methodology can be presented as a succession of steps or algorithms. Each step involves automatic calculations.

### 2.1.2.1 Geometrical and statistical sources

Three different types of objects are overlaid:

- UMZ 2000<sup>10</sup>
- Population density grid from JRC (version 5)<sup>11</sup>
- Local administrative units (LAU 2, EuroBoundaryMap 2006 v2.0 from EuroGeographics, validity: 2006).

### 2.1.2.2 Computation steps

We compute the population intersecting LAU and UMZ and we observe the maximal value for each LAU related to one UMZ.

Let  $L$  describe the LAU and  $u$  the UMZ. After the intersection, let  $L(u)$  be the part of the LAU  $L$  intersecting the UMZ  $u$ , and  $P_{L(u)}$  be the population of this part, when  $P_u$  is the population of the whole UMZ.

$$P_u = \sum \{ P_{L(u)}, L \text{ intersecting } u \}$$

Thus for each UMZ  $u$ , the series of  $\{ P_{L(u)}, L \text{ intersecting } u \}$  is considered, ranked by decreasing order, and let then  $L_i(u)$  be the  $i^{\text{th}}$  part in this ordered series.

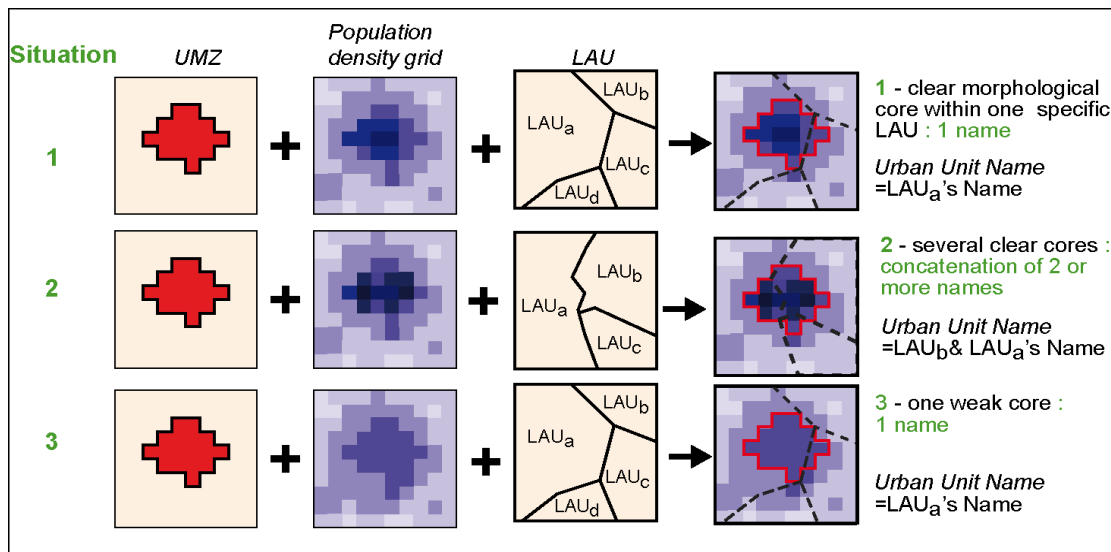
For a given UMZ  $u$ , three different situations can occur (Figure 1).

---

<sup>10</sup> Latest version given by the European Topic Center on Land Use and Spatial Information (ETCLUSI) in June 2010, which should be available in the future EEA dataserver. Official distribution: <http://www.eea.europa.eu/data-and-maps/figures/urban-morphological-zones-umz-2000>.

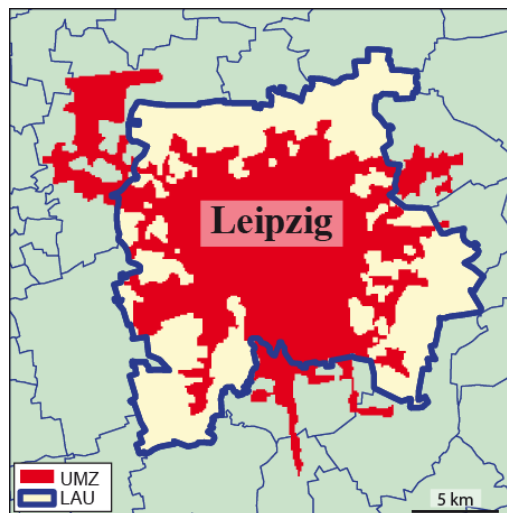
<sup>11</sup> <http://www.eea.europa.eu/data-and-maps/data/population-density-disaggregated-with-corine-land-cover-2000-2/population-density-grid-geotiff-format>.

**Figure 1 : Naming methodology (Situation 1, 2 and 3)**



**SITUATION 1:** The largest population of the LAUs intersections is more than 50% of the UMZ's population. We have an UMZ with one strong core, clearly organized around one center. The UMZ is named with the name of this LAU<sub>a</sub> (Figure 1). This is the case of Leipzig example (Figure 2).

**Figure 2 : Leipzig (Germany), an UMZ with one strong core (Situation 1)**



UMZ population: 536 552 inh.  
UMZ population in Leipzig LAU 2: 483 285 inh.

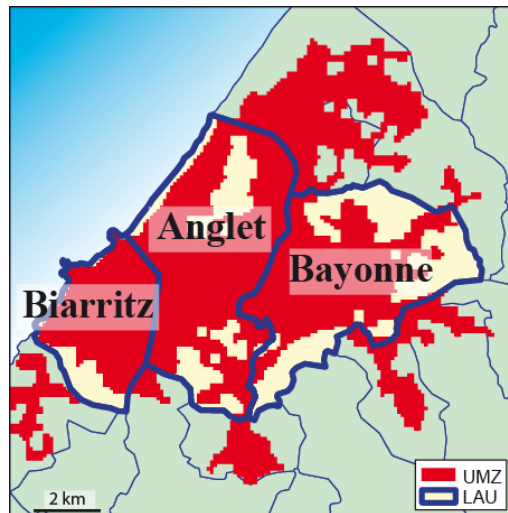
Sources: LAU 2 (EuroBoundaryMap 2006, v2.0) from EuroGeographics, UMZ2000 from European Environment Agency, Population density Grid v.5 from Joint Research Center.

**SITUATION 2 and 3:** There is not a unique main core as defined above, thus the larger part is retained as the reference, and the other parts are considered successively, in decreasing order of population, as long as their populations exceed 50% of the first part population.

$$\text{Secondary units} = \{L_j(u) / P_{L_j(u)} \geq 0.5 * P_{L1(u)}\}$$

**Situation 2** : one or several secondary units' population represent more than 50% of the population of the largest part. We retain the name of the concerned secondary units, and the final name of the UMZ is a compounded name. The order of the names is not alphabetical but follows the decreasing order of population contributions to UMZ. This is the case of Bayonne-Anglet-Biarritz (Figure 3).

**Figure 3 : Bayonne-Anglet-Biarritz (France), an UMZ with several cores (Situation 2)**

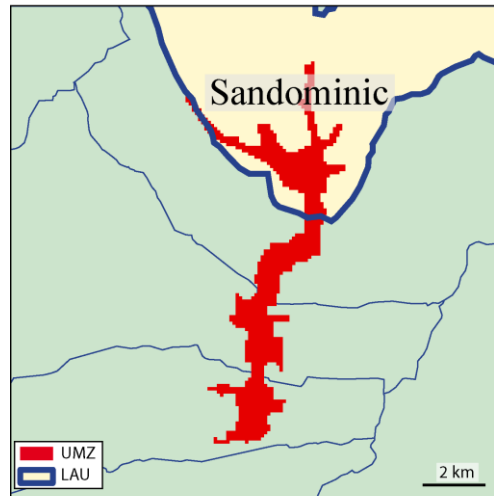


UMZ population: 128 554 inh.  
 Bayonne LAU 2 population inside UMZ: 39 708 inh.  
 Anglet LAU 2 population inside UMZ: 35 185 inh.  
 Biarritz LAU 2 population inside UMZ: 30 156 inh.  
 Other LAU 2 population inside UMZ < 12 000 inh.

Sources: LAU 2 (EuroBoundaryMap 2006, v2.0) from EuroGeographics, UMZ2000 from European Environment Agency, Population density Grid v.5 from Joint Research Center.

**Situation 3**: no secondary unit's population represents more than 50% of the larger part. We retain finally only the name of the main LAU unit, fitting again with a "one core" context (one morphological core, but less strong than in Situation 1) (Figure 4).

**Figure 4: Sandominic (Romania), an UMZ with one weak core (situation 3)**



UMZ population: 10 678 inh.  
 Sandominic LAU 2 population inside UMZ: 4 893 (46% of total UMZ inhabitants)  
 Other LAU 2 population inside UMZ < 2446 inh.

Sources: LAU 2 (EuroBoundaryMap 2006, v2.0) from EuroGeographics, UMZ2000 from European Environment Agency, Population density Grid v.5 from Joint Research Center.

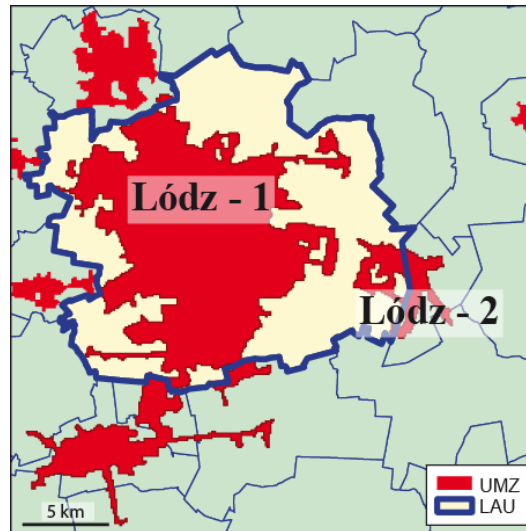
The different steps of the algorithm can be summarized by:

IF  $P_{L_1(u)} \geq 0.5 * P_u$  THEN Name( $u$ )=Name( $L_1(u)$ )  
 ELSE IF Secondary units= $\{L_j(u) / P_{L_j(u)} \geq 0.5 * P_{L_1(u)}\} \neq \emptyset$   
 THEN Name( $u$ )=Name( $L_1(u)$ )+ $\{Name(L_j(u)), j / P_{L_j(u)} \geq 0.5 * P_{L_1(u)}\}$   
 ELSE Name( $u$ )=Name( $L_1(u)$ )

### 2.1.2.3 A particular case: different UMZ with identical names

In the large majority of cases, each UMZ could be associated to a unique name (situations 1 and 3) or to a unique combination of names (situation 2). However, in about 10% of the cases, several UMZ share the same administrative unit. It is for instance what happens in the case of the city of Łódź in Poland (Figure 5): the most populated parts of two different UMZ fall into the same LAU2, so that they both receive exactly the same name. In order to maintain the attribution of distinct identifiers for UMZ, we add a number after the name, according to the decreasing size of UMZ populations (Łódź - 1 and Łódź - 2).

**Figure 5 : Łódź (Poland), two UMZ with the same name**



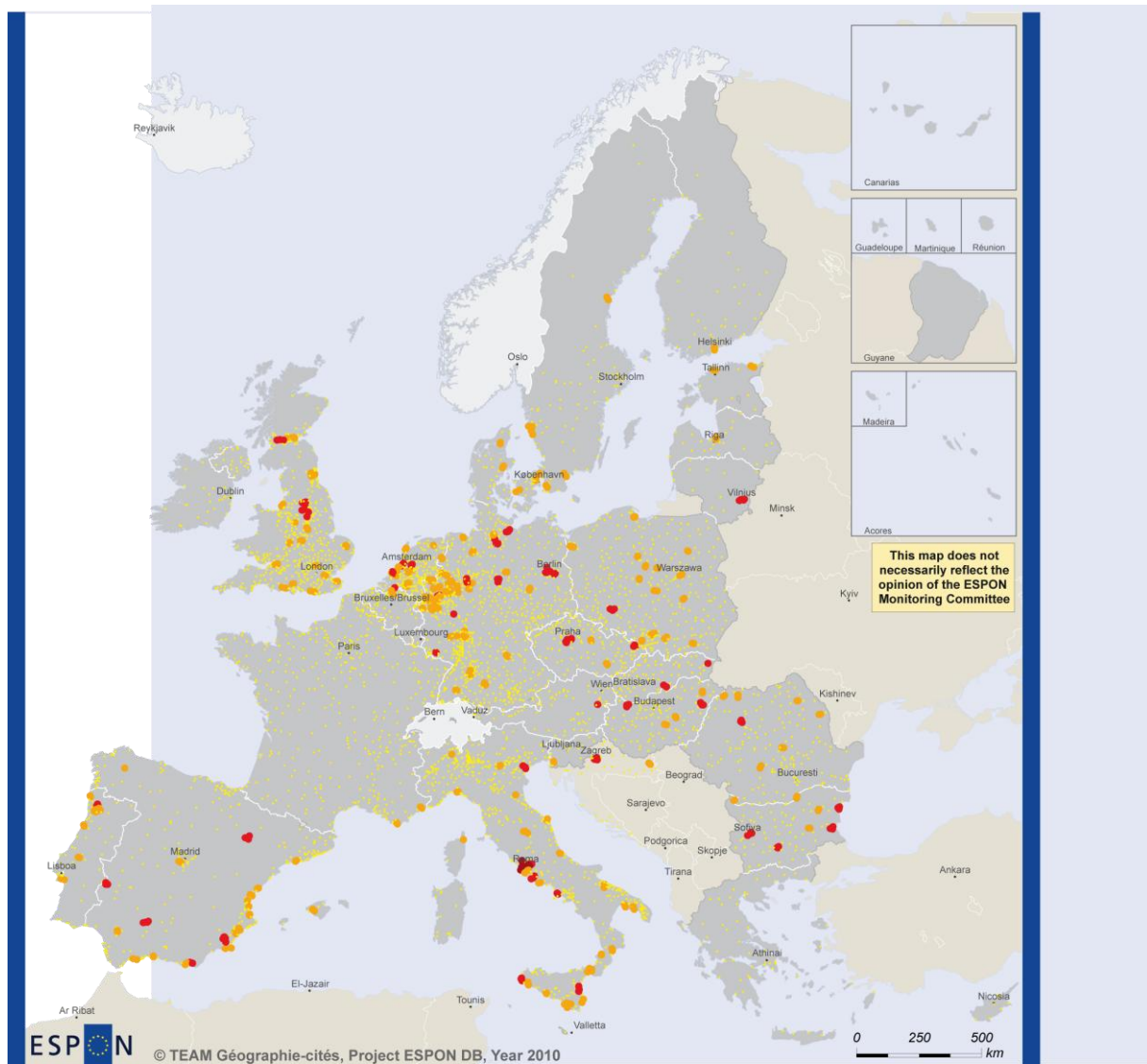
UMZ Łódź - 1: 822545 inh.  
UMZ Łódź - 2: 43894 inh.

Sources: LAU 2 (EuroBoundaryMap 2006, v2.0) from EuroGeographics, UMZ2000 from European Environment Agency, Population density Grid v.5 from Joint Research Center.

Figure 6 displays the location of those cases all over Europe. They appear to be quite well distributed from one country to another, even if some regions concentrate a large number of cases (like in the Rhine-Ruhr Valley or in the Netherlands) and even if some countries do not host any of them (like in France, where administrative units are particularly small).



**Figure 6 : UMZ 2000 with identical names**



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

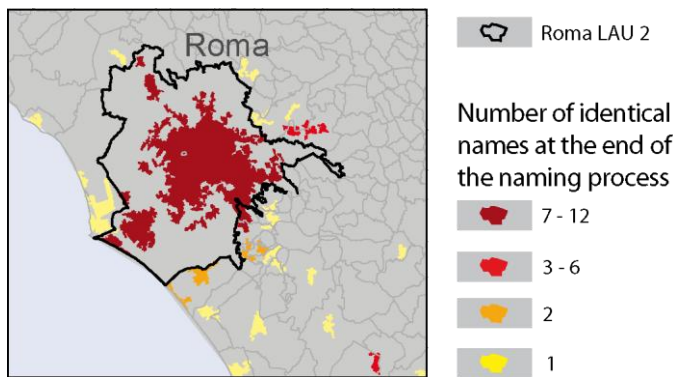
Regional level: NUTS 0  
Source: ESPON DB, year 2010  
Origin of data: The European Environment Agency (UMZ 2000 V.2), Joint Research Center (Density Grid V.5), LAU2 (2006, V.2) and national sources (see figure 9)  
© EuroGeographics Association for administrative boundaries

Number of identical names at the end of the naming process

- 7 - 12
- 3 - 6
- 2
- 1
- no UMZ data

In most of the cases, only 2 identical names result from the naming process, and in a few cases we obtain 3 to 6 repetitions of the same name. The last class in the map (more than 6 repetitions) is only illustrated by Roma: 12 UMZ share this name! This is due to the very large size of the Roma LAU 2 (Figure 7), where local units have probably been merged into a unique metropolitan level.

**Figure 7 : The case of Roma: 12 UMZ sharing the same name**



Sources: LAU 2 (EuroBoundaryMap 2006, v2.0) from EuroGeographics, UMZ2000 from European Environment Agency, Population density Grid v.5 from Joint Research Center.

## 2.2 Automation of processing chain

### 2.2.1 The need for an automated process

The automation of the rules defined for naming UMZ is necessary for three main reasons:

- The inputs represent a huge mass of data emanating from different files which requires automatic support instead of manual process inside a GIS:

  - UMZ: 4 437 UMZ over 10 000 inhabitants

  - Euroboundary, LAU2: 106 452 administrative units

  - Population Grid: More than 2 billion pixels in the density grid

- The calibration of the naming method supposes to conduct different tests which are useful to choose the right administrative level of reference in some specific cases (see below, 2.3.1).

- The databases contents are constantly evolving and it is essential to be reactive to these changes. Automation allows quick updating with new versions of sources or methods (EEA and JRC; for example, there are at least two different versions of UMZ2000) or new dates (as regards for instance to the integration of the future UMZ 2006, 2010...).

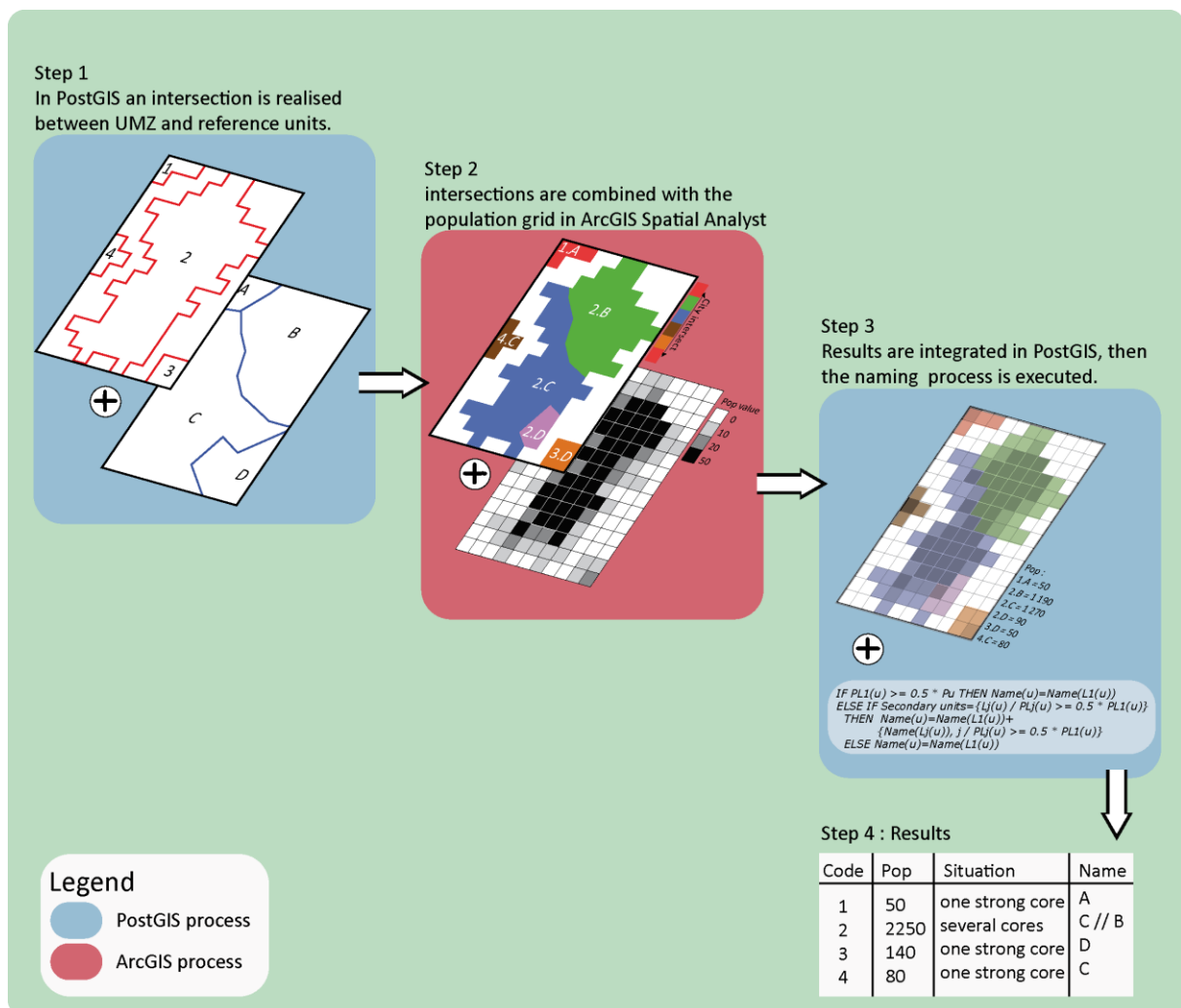
### 2.2.2 Process description and software solution

From a technical point of view, the automated implementation of the naming algorithm is based on three different steps which have been executed through

PostGIS and ArcGIS softwares, thanks to an integrated processing chain (Figure 8)<sup>12</sup>:

- Step 1: the geometrical intersections between the UMZ and the reference units are created through PostGIS
- Step 2: the population of the resulting intersections is calculated with ArcGIS Spatial Analyst. To a recent date indeed, PostGIS could not allow manipulating any raster data and the program had to use the raster solutions of the spatial analyst add-on of ESRI®. A Python language program using the Geospatial Data Abstraction Library (GDAL) has thus been developed to interface the two softwares.
- Step 3: the population computed is retrieved and integrated into PostGIS, in which the algorithm of naming is implemented

**Figure 8 : The different steps of the processing chain**



This program can process all the data and all the steps at once, which prevents from errors and duplicates. Eventually the automatic naming for the whole Europe could be realized in about one hour.

<sup>12</sup> Ultimately, this program should only rely on open sources technologies. The statistical processes can be now implemented through a PostgreSQL database with the help of the PostGIS add-on, which allows processing data with a geometry and realizing spatial requests.

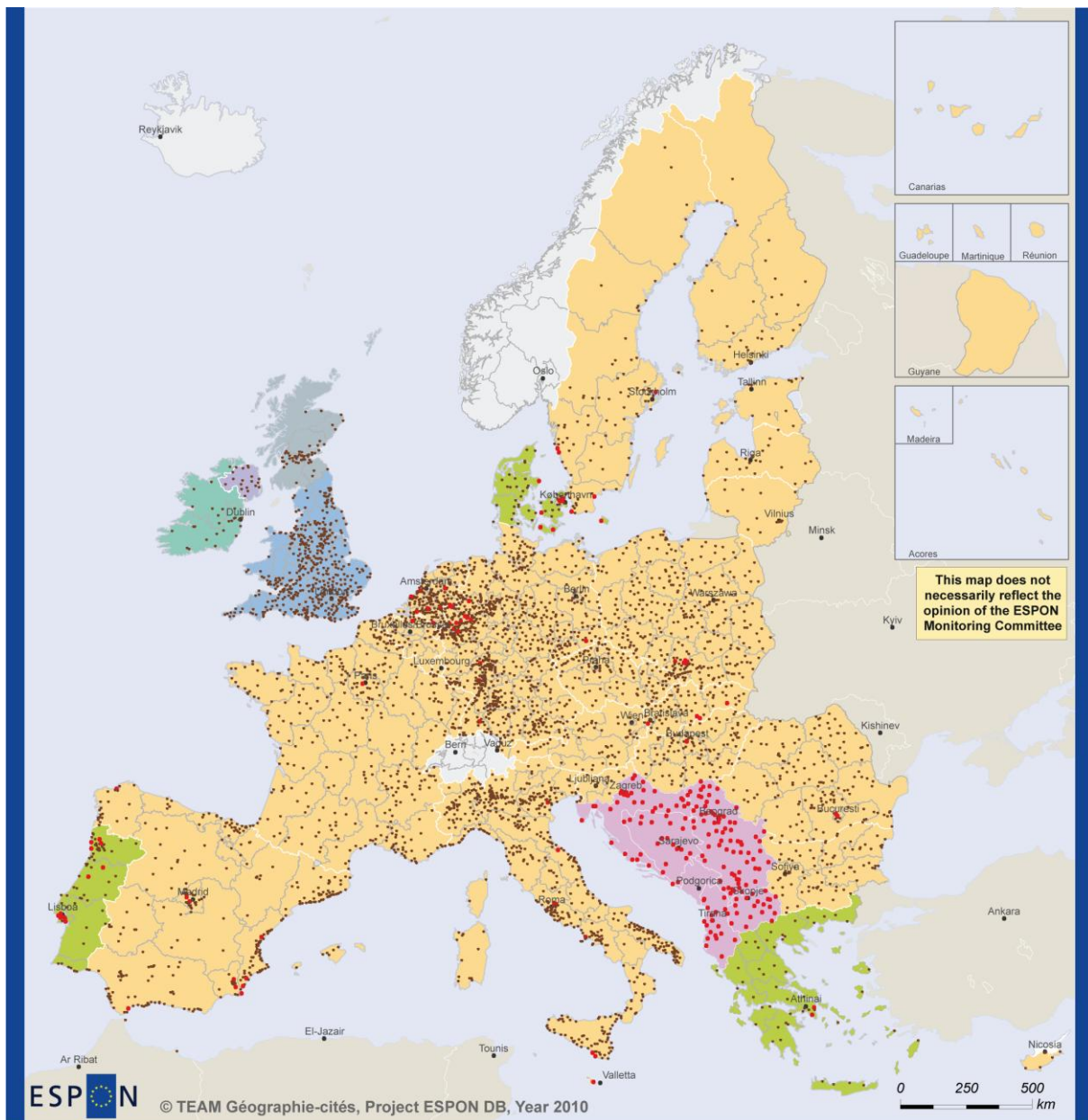
## **2.3 Expertise and validation**

The validation of the naming method results from an expertise based on a comparison with the relevant national data base for city names (LAU 2 in majority, but also LAU1 or national settlement areas). Final results are then systematically matched to other sources (Eurostat, Geopolis) for validation. The last particular cases are checked manually, using Google Earth.

### **2.3.1 Selecting relevant reference units**

The identification of relevant units of reference for choosing UMZ names deals with a critical issue: is there a semantic level more suitable than another for naming towns and cities? At first sight, LAU 2 seems to be the most accurate level and the most usual reference at European scale. This level is relevant in the large majority of cases, but it cannot be used for the whole countries: the correlation between cities usual names and administrative levels names depends indeed on the history of administrative divisions and on the way the status of city was formerly given. It may thus vary in some countries or experiment local variations within some countries. The name "Leipzig" fits for instance with LAU 2 level whereas the name "Dublin" fits with LAU 1 level and the name "Paris" with NUTS 3. An expertise was thus necessary to select the best reference unit for each country (Figure 9).

**Figure 9 : Relevant reference units for naming UMZ**



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

Regional level: NUTS 0  
Source: ESPON DB, year 2010  
Origin of data: The European Environment Agency (UMZ 2000 V.2),  
Office for National Statistics (England & Wales),  
Northern Ireland Statistics & Research Agency, General Register Office for Scotland.  
© EuroGeographics Association for administrative boundaries

**City names**

- naming from manual process
- naming from automatic

**Sources**

**Homogeneous sources and data grid population**

- SIRE (Lau2 level)
- SIRE (Lau1 level)

**No data grid population**

- www.citypopulation.de

**National sources and data grid population**

- Urban areas (England and Wales)
- Settlement Development Limits (Northern Ireland)
- Census town (Ireland Republic)
- Settlements (Scotland)

no UMZ data

### **2.3.1.1 LAU1 instead of LAU2**

In some countries, the relevant administrative level appeared to be rather LAU1 than LAU2. We have used the LAU 1 version of EuroBoundaryMap 2006 v2.0 from EuroGeographics (validity: 2006).

#### *Portugal*

In Portugal, the status of a city was formerly given by decree and most of the cities corresponded to LAU 1 capital cities (*capitais de distrito*). This legacy is still present, in the sense that the current names of the LAU 2 have no relation with the names of the cities. We have then chosen LAU 1 (*concelhos-municipios*) for naming UMZ.

#### *Denmark*

In Denmark, the LAU 2 level corresponds to a parish level, whose names do not fit with the real names of cities. The most accurate level for naming UMZ is the *Kommuner* level (LAU 1).

#### *Greece*

The same issue occurs for Greece where the LAU2 level has no relation with the city name usually used. The LAU 1 level (*Demoi* and *Kointites*) has thus been chosen.

### **2.3.1.2 NUTS instead of LAU**

NUTS level has been selected for some capital cities or other particular cases.

- Paris, Bucharest, and Budapest: the LAU 2 fits with sub-city districts (called "arrondissement" or "sector"), so that NUTS 3 level has been used in the algorithm.

- London: the name "London" is not represented at LAU 2 level (and the algorithm gives a "UMZ with several cores", with several hundred of names) neither at LAU 1 level (28 names obtained). At NUTS 3 level, the names are like "Inner London West" etc., at NUTS 2 level "Inner London" and "Outer London". The best administrative level fitting with the name "London" and with the spatial extent of the UMZ is the NUTS 1.

- Brussels: there is one LAU 2 called Brussels but it is a very little one compared to the present extent of the city, so that the name of the LAU 2 is not retained by the automatic process (the final name of the "UMZ with several cores" would be Antwerpen-Gent). Thus we have chosen the NUTS 3 level ("Arr. de Bruxelles-Capitale / Arr. van Brussel-Hoofdstad"). The definitive name resulting from the algorithm is Brussel-Antwerpen-Gent.

- Valetta (Malta): there is just one administrative level below the national one (a LAU 2 level), and the eponym LAU 2 is too small to emerge from the automatic algorithm in the final name of the UMZ. We have then attributed the name Valetta.

- In Slovakia: Bratislava and Košice are divided in several districts at LAU 2 and LAU 1 levels. The best level for naming is NUTS 3 ("Bratislava region" and

“Košice region”) but it is very large compared to the UMZ spatial extents. Here again, we have attributed the names Bratislava and Košice to the UMZ.

### 2.3.1.3 National settlement areas

In some other countries, neither LAU2 nor LAU1 appeared suitable for naming and national data bases have been used as a reference.

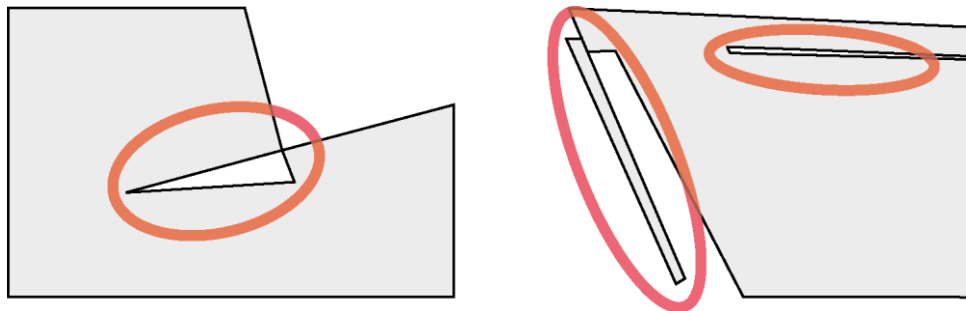
#### *United-Kingdom*

In United Kingdom the LAU 2 level does not fit with names given historically to cities. LAU 2 correspond to “electoral wards” (or “parts thereof”) and LAU1 to “district/unitary authorities”. The city names of the “urban areas” (morphological agglomerations built by the Ordnance Survey) do not necessary fit with LAU2 or LAU1, so that we have used these urban areas as the reference units in the automatic algorithms. Three different data bases have been used:

- *Urban areas* of England and Wales (2001)<sup>13</sup>
- *Settlements* of Scotland (2009)<sup>14</sup>
- *Settlements Development Limits* of Northern Ireland (2005)<sup>15</sup>

The vector format versions of these databases have been kindly sent by the National Statistics Office of United Kingdom. They could be used to give names by spatial requests, after correcting a hundred of topological errors that hampered the application of the automatic processes (Figure 10).

**Figure 10 : Some examples of topological errors**



#### *Ireland*

In Ireland, LAU2 corresponds to “electoral districts” and LAU1 do not systematically fit with the city names given to Census Towns by the Central Statistics Office of Ireland (for example when the LAU1 is a county). We have then used the Census Towns of Ireland (2006) data base, sent in vector format by the National Statistics Office of Ireland.

<sup>13</sup> <http://www.statistics.gov.uk/>

<sup>14</sup> <http://www.gro-scotland.gov.uk/>

<sup>15</sup> <http://www.cso.ie/>

## **2.3.2 Countries without population density grid**

For the 149 UMZ larger than 10 000 inhabitants that are located in Balkan countries (Albania, Bosnia-Herzegovina, Kosovo, Macedonia and Serbia), the population has not been attributed by EEA using the Population density grid but using other sources ([www.citypopulation.de](http://www.citypopulation.de)). Consequently, the automatic algorithms have not been applied to these 149 UMZ and we have used the same source for giving names.

In order to ensure a good comparability in thematic explorations (section 3) these countries have not been included in the analyses.

## **2.3.3 Validation process**

### **2.3.3.1 Sources**

Implementing an automatic process is essential in order to quickly adapt the naming method to new sources, to avoid errors and to establish the process traceability. Yet it is equally important to validate the resulting names by comparing them to other existing urban databases. Two sources were used to check the quality of the method:

- Geopolis database (Moriconi-Ebard, 1994)
- Eurostat compilation of national city names: database "Geographical names: Settlements"<sup>16</sup>

In each of these databases, the cities are only represented by points (centroids) which are associated to a name. The checking method relies on successive steps:

- First, a spatial overlay of the names attributed by the algorithm and of the names associated to Geopolis and Eurostat databases. This comparison is based on a spatial request that retrieves the centroids intersecting UMZ. Specific spatial patterns have to be taken into account (for instance when Eurostat or Geopolis centroids intersect UMZ "holes").
- Secondly, a semantic comparison. The associated names are gathered into common tables and UMZ naming is validated if the names are the same. In order to optimize the matching process, it is necessary to realize textual corrections: lowercases everywhere, same local abbreviations, same spellings and universal translations (differences like Warszawa/Varsovie, Praha/Prague, Aix en Provence/Aix-en-Provence, etc., have been corrected by choosing the name of the referent database -LAU or national settlement areas-). These corrections enabled to identify and correct 1081 mismatches in names.

---

<sup>16</sup> <http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/geodata/archives>



### 2.3.3.2 Typology of errors

85 names on 4437 UMZ did not match after those first checks and required case by case semantic modifications. A typology of errors has been proposed in order to make easier the future checks. These errors refer to 4 types of mismatches:

- The name of the reference unit (LAU or national settlement area) does not fit with any urban locality. This case happens generally when the name is related to some topographical features, rather than to the settlement itself (for instance the UMZ named Farum in Denmark was first called Fureso by the algorithm, whereas it is the name of a lake near the city). It happens also sometimes when it refers to a general location (for instance the large LAU named Westland, located in the south of Den Haag, Netherlands, is not suitable for naming the UMZ which corresponds more precisely to Monster locality in Google Earth).

- The name of the reference unit was historically given by an eponymous city that is currently less populated than another city included in this unit. There has been a sort of reversal between historic names and population trends, so that the most important UMZ does not receive the name of the most populated locality. This is for instance the case of the UMZ which is named Pamela (Portugal) according to the algorithm, whereas it should receive the name of the largest city (Pinhal Novo) of this LAU, identified by using Google Earth.

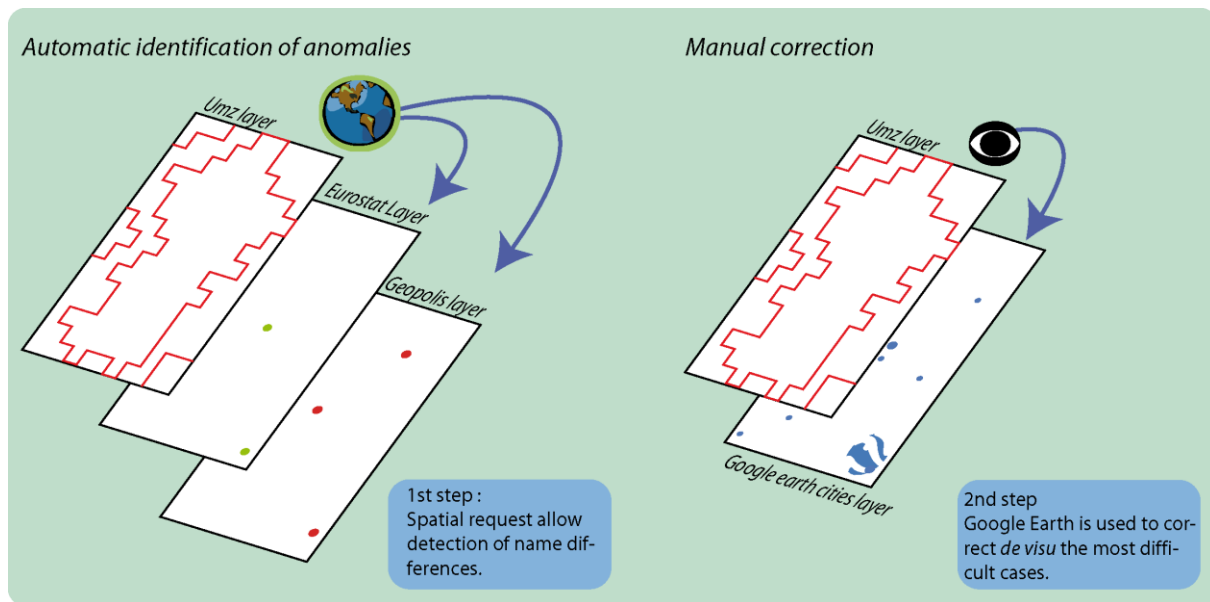
- The manual expertise of UMZ with identical names ("Lodz cases") has revealed another inconsistency: two UMZ included in the same reference unit can receive the same name even if they are very distant from each other. This is for instance the case of the two UMZ named Kristianstad-1 and Kristianstad-2 (Sweden): the UMZ of Kristianstad-2, which is distant from 10.5 km to Kristianstad-1, clearly overlays the locality of Åhus in Google Earth.

- In the peripheral parts of some industrial conurbations (Mannheim, Düsseldorf, Katowice...), a small UMZ located in the periphery of a reference unit takes the name of this unit whereas the most urbanized part of this unit belongs to the conurbation. As this urbanized part counts a relatively few population, the name of the reference unit is not taken into account in the name of the UMZ conurbation, so that this name is finally attributed only to the small UMZ. We have then chosen to give the Google Earth name and not the reference unit name to this small UMZ. An illustration can be given by the case of the locality of Ruchheim, in Germany, which is included in the LAU 2 Ludwigshafen, whose larger urbanised part belongs to Mannheim UMZ. As the automatic algorithm gives the name "Ludwigshafen" to the UMZ situated at Ruchheim place, we have corrected manually this name.

### 2.3.3.3 Solutions proposed

Ultimately these remaining mismatches are corrected by referring to the Google Earth database whose names layer is based on Multinet® from TeleAtlas®. The 85 UMZ have been converted to KLM format in order to be overlaid with other Google Earth layers. Names are then corrected *de visu* for the last mismatches (Figure 1).

**Figure 11 : Steps of the validation process**



## 2.4 Results : typology of naming situations

A simple count gives a first idea of the results obtained by automatic algorithms coupled with expertise on relevant administrative levels. We have considered UMZ larger than 10 000 inhabitants (4437 objects, including Balkans). The results have been summarized in Table 1:

**Table 1 : Naming UMZ through automatic methods**

	<b>SITUATION 1</b> "UMZ with one strong core"	<b>SITUATION 2</b> "UMZ with several cores"	<b>SITUATION 3</b> "UMZ with a weak core"
<b>Total number</b>	4164	193	80
<b>Percentage</b>	94%	4%	2%

Sources: LAU 2 (EuroBoundaryMap 2006, v2.0) from EuroGeographics, UMZ2000 from European Environment Agency, Population density Grid v.5 from Joint Research Center.

The typology presented in Table 1 has been mapped in Figure 12. If we focus first on "situation 2" (several cores), we recognise the industrial conurbations of the Midlands, the French and Belgium basin, the Ruhr basin, Silesia and Galicia regions. We also identify some sea-side conurbations, for example in Portugal, Spain, Italy or France. Another type of "UMZ with several cores" consists in large cities sprawling and connecting other large and close cities, like in Belgium (around Brussels) or in Romania.

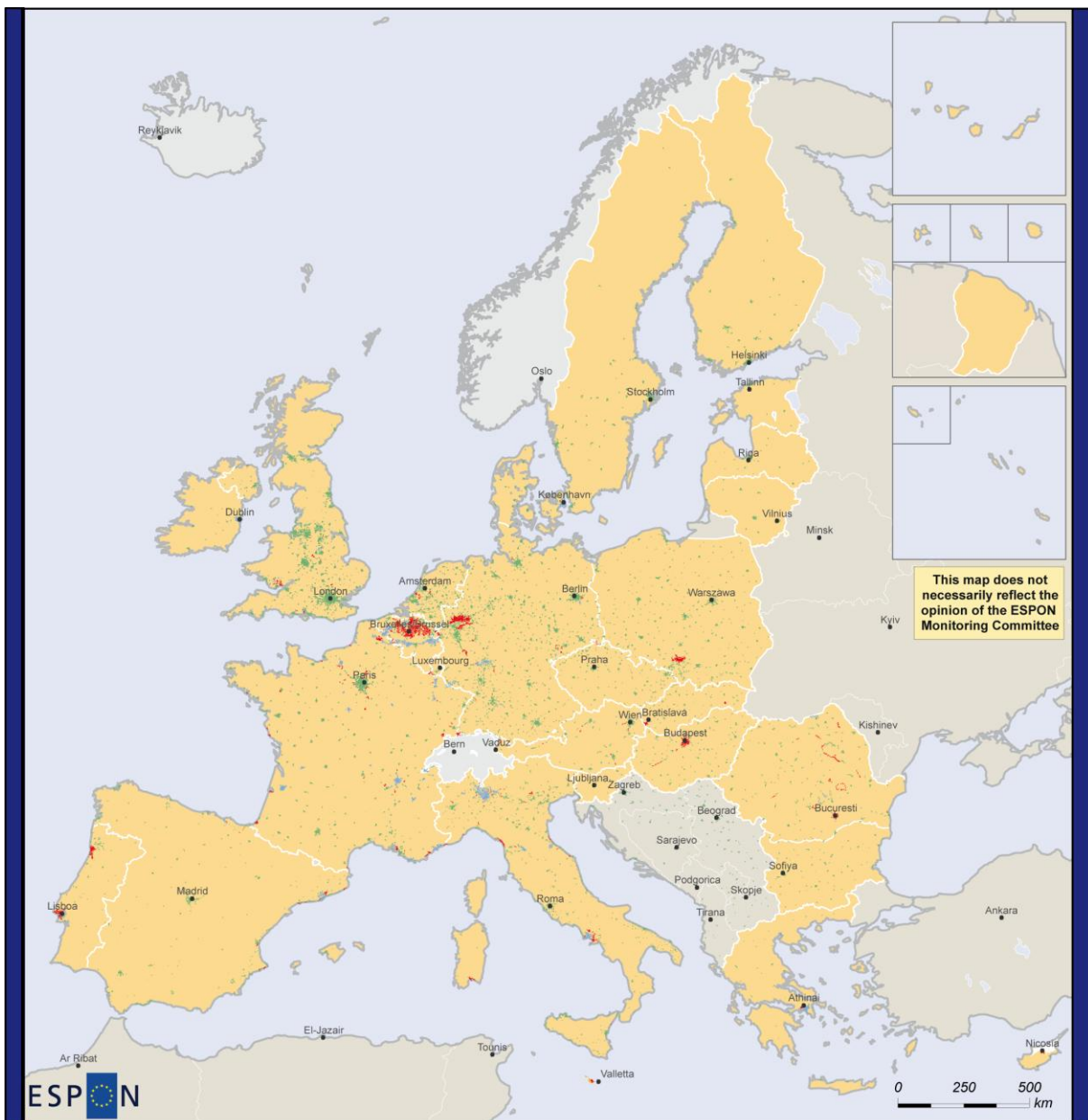
For the "situation 3" (one weak core, less strong than in situation 1), we can notice that locations are mostly the same than for UMZ with several cores (see in Italy, United Kingdom, Belgium, France...).

The "one strong core" cases, which represent the great majority (94% of the UMZ) are spread all around Europe but more represented in Northern Europe

(Sweden, Baltic countries, Denmark), characterised by relatively sparse urban settlements.





Let us notice that it is difficult to give more interpretations: situations 2 and 3 do not necessarily enlighten some "polycentric cities" but may result from the specific local or national average size of the reference units that have been used in the algorithm: we have more chances to obtain a "several cores" situation when this average size is little (like in France), and a real polycentric city could appear as "with one core" if the average size is large (like in Denmark).

**Figure 12 : UMZ typology according to naming results**




 EUROPEAN UNION  
 Part-financed by the European Regional Development Fund  
 INVESTING IN YOUR FUTURE

**UMZ type :**

-  Situation 1 : UMZ with one strong core
-  Situation 2 : UMZ with one weak core
-  Situation 3 : UMZ with several cores
-  no UMZ data

Regional level: NUTS 0  
 Source: ESPON DB, year 2010  
 Origin of data: The European Environment Agency (UMZ 2000 V.2), Joint Research Center (Density Grid V.5), LAU2 (2006, V.2) and national sources (see figure 9)  
 © EuroGeographics Association for administrative boundaries

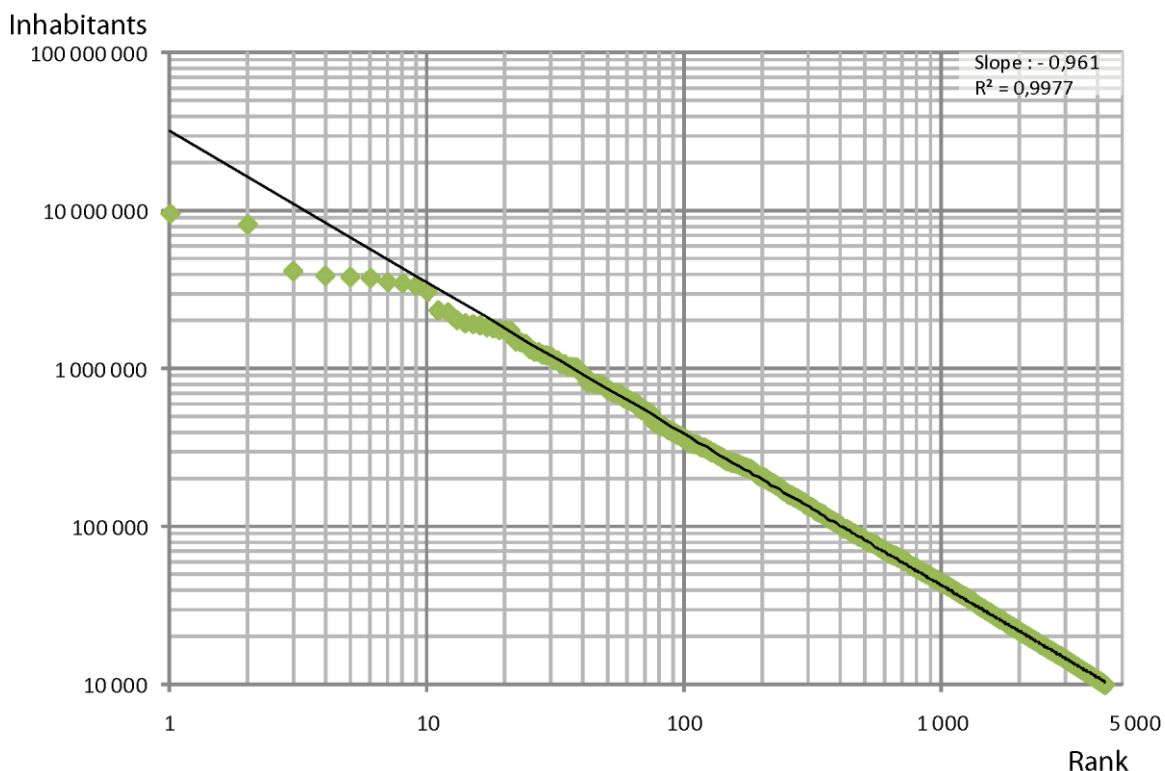
### 3 A thematic insight into European cities

The UMZ data base is now operational for a deep exploration of the common features and diversity of European urban settlements. Three types of analyses can be presented as a preview: in addition to classical indicators referring to the hierarchical structure of city systems or to the population density of cities, a new indicator has been created to identify and compare international UMZ.

#### 3.1 Urban hierarchy and city-size distribution

The classical rank-size distribution, plotted for the 4437 cities over 10 000 inhabitants (Figure 13), confirms the very high regularity of the hierarchical structure at the European level (the determination coefficient  $R^2$  equals to 0.99). The absolute value of the slope, used as an indicator of city size inequality level, is 0.96, very close to other values computed by European researchers with former databases (for instance, Geopolis data base, 1994). National studies and computation of primacy index should fruitfully complete this overview of urban hierarchy in Europe.

**Figure 13 : Pareto-Zipf distribution of city sizes (UMZ 2000 data base)**



Sources: LAU 2 (EuroBoundaryMap 2006, v2.0) from EuroGeographics, UMZ2000 from European Environment Agency, Population density Grid v.5 from Joint Research Center.

A closer look at the head of this hierarchy can be proposed through the “top ten” UMZ (Table 2), which are compared here to other urban rankings resulting from

European databases (Morphological Urban Areas from IGEAT<sup>17</sup> and Larger Urban Zones from Urban Audit<sup>18</sup>). Seven of these UMZ are also part of the largest set of MUA and LUZ. Main differences concern conurbations like Bruxelles-Antwerpen-Gent, Liverpool-Manchester and Essen-Dortmund-Duisburg, which are clearly overestimated by UMZ as compared to MUA which are built from similar morphological criteria. Further details about UMZ ranking are given in the table of the 50 first UMZ (see Annex, table 1).

**Table 2 : “Top ten” UMZ compared to MUA and LUZ (2000, population in thousand inhabitants)**

UMZ			MUA			LUZ		
Rank	Name	Pop.	Rank	Name	Pop.	Rank	Names	Pop.
<b>1</b>	Paris	9 476	<b>1</b>	Paris	9 591	<b>1</b>	London	11 917
<b>2</b>	London	8 208	<b>2</b>	London	8 265	<b>2</b>	Paris	11 089
<b>3</b>	Milano	4 156	<b>3</b>	Madrid	4 955	<b>3</b>	Madrid	5 805
<b>4</b>	Essen-Dortmund-Duisburg-Bochum	3 891	<b>4</b>	Berlin	3 776	<b>4</b>	Ruhrgebiet	5 302
<b>5</b>	Madrid	3 843	<b>5</b>	Barcelona	3 755	<b>5</b>	Berlin	4 971
<b>6</b>	Bruxelles-Antwerpen-Gent	3 790	<b>6</b>	Milano	3 698	<b>6</b>	Barcelona	4 234
<b>7</b>	Liverpool-Manchester	3 531	<b>7</b>	Athinai	3 331	<b>7</b>	Athina	4 013
<b>8</b>	Athinai	3 489	<b>8</b>	Roma	2 532	<b>8</b>	Roma	3 458
<b>9</b>	Berlin	3 435	<b>9</b>	Birmingham - Wolverhampton	2 363	<b>9</b>	Hamburg	3 135
<b>10</b>	Barcelona	3 106	<b>10</b>	Lisboa	2 315	<b>10</b>	Milano	3 077

Sources: LAU 2 (EuroBoundaryMap 2006, v2.0) from EuroGeographics, UMZ2000 from European Environment Agency, Population density Grid v.5 from Joint Research Center, ESPON 1-4-3, Urban Audit.

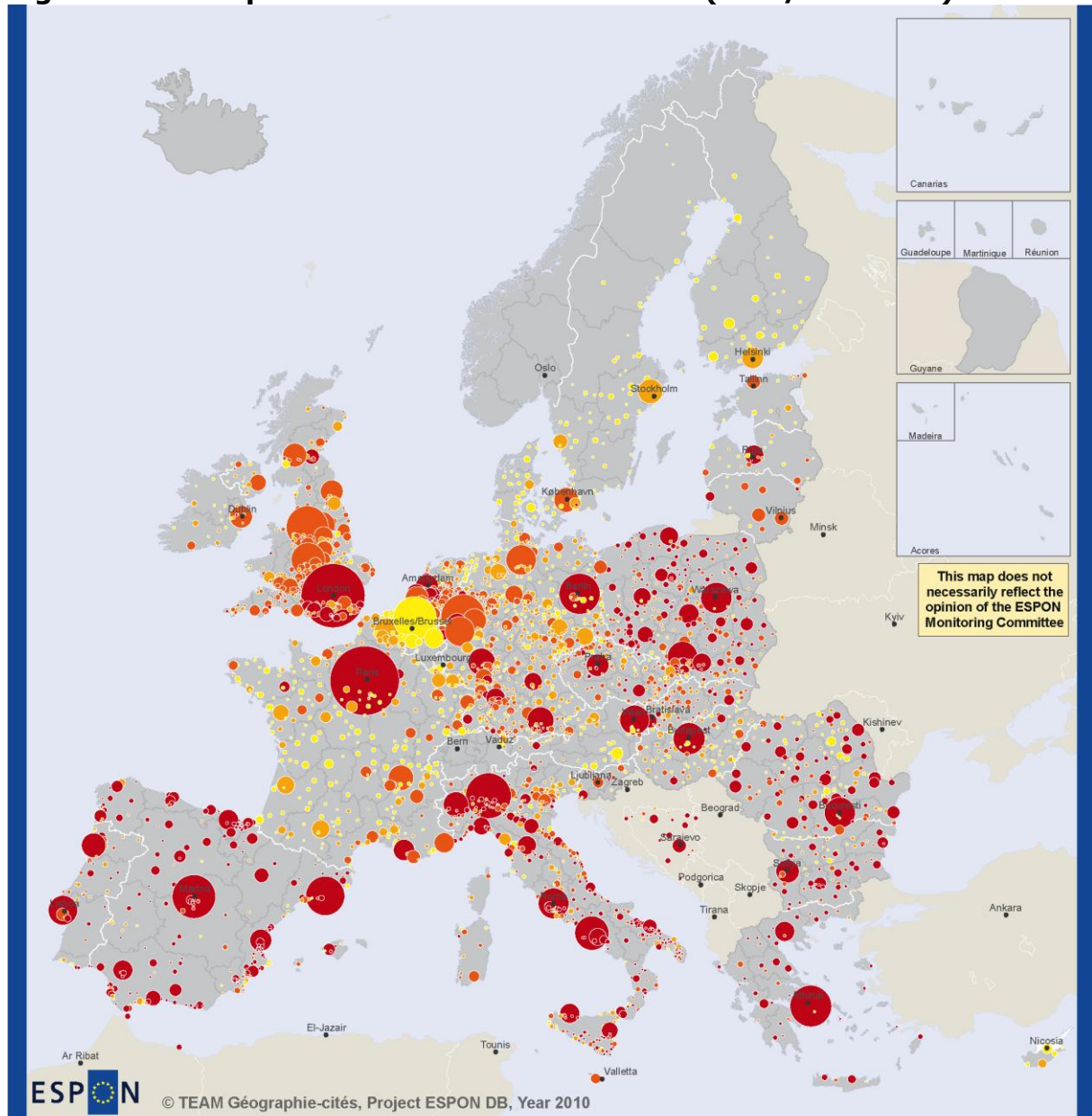
## 3.2 Density patterns

A multiscalar analysis of density levels in Europe gives striking results, with a major North-South gradient (Figure 14): for example, average urban density is lower than 2000 inh./km<sup>2</sup> in Sweden, Denmark, Finland, whereas it reaches 4000 inh./km<sup>2</sup> in Italy and more in Spain or Greece (Table 3).

<sup>17</sup> MUA have been defined in Vanderhoff et alii 1999 and in ESPON 1-4-3 « Study on urban functions ». IGEAT refers to Institut de Gestion de l'Environnement et d'Aménagement du Territoire, Université Libre de Bruxelles.

<sup>18</sup> <http://www.urbanaudit.org/>

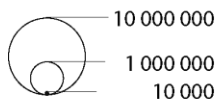
**Figure 14 : European cities sizes and densities (UMZ/CLC 2000)**



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

Regional level: NUTS 0  
Source: ESPON DB, year 2010  
Origin of data: The European Environment Agency (UMZ 2000 V.2), Joint Research Center (Density Grid V.5), LAU2 (2006, V.2) and national sources (see figure 9)  
© EuroGeographics Association for administrative boundaries

Number of inhabitants

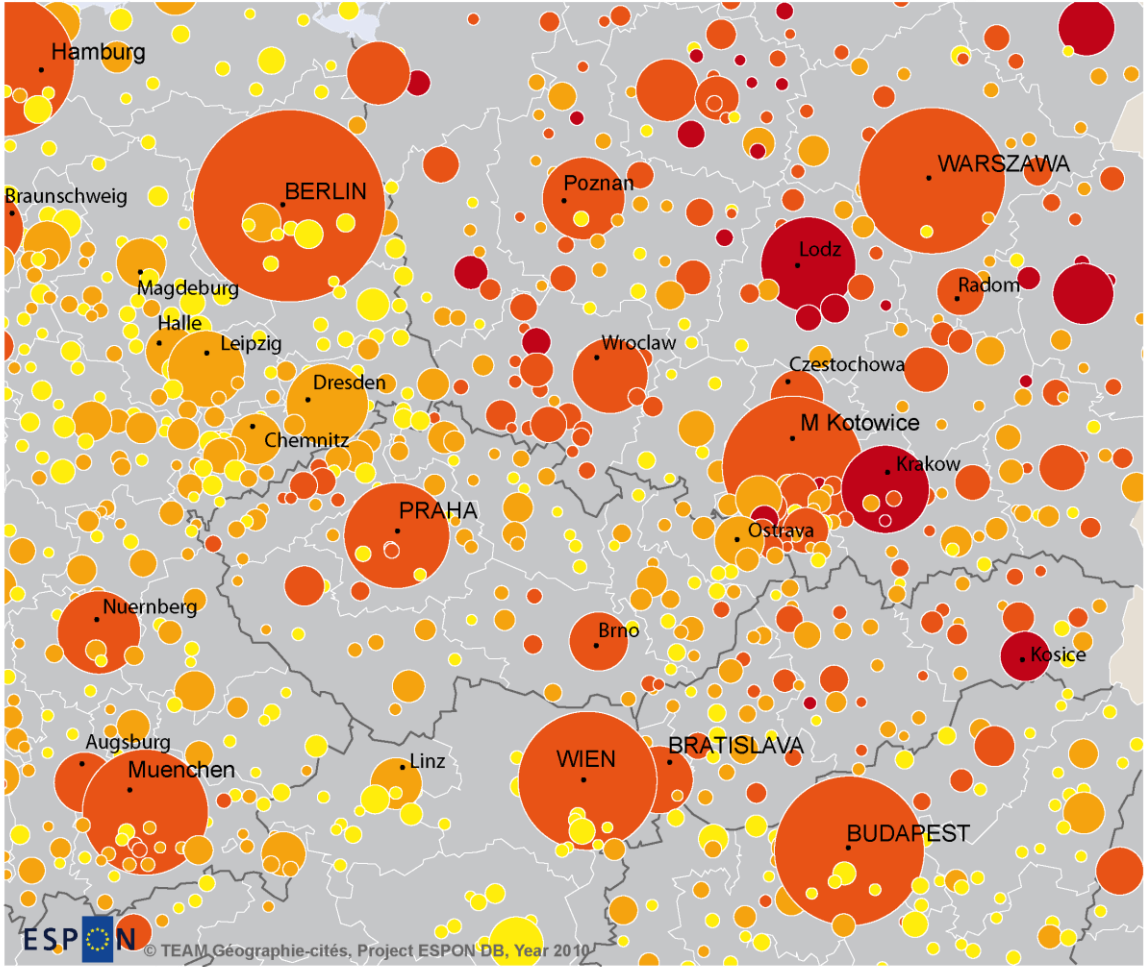


Density (inhab/km<sup>2</sup>)



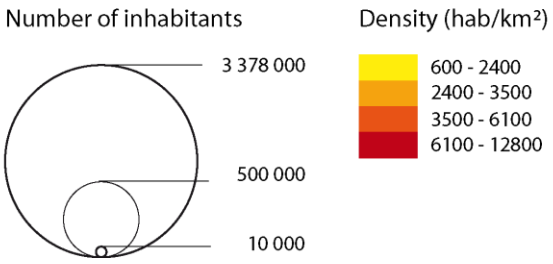
Some national specificities appear also very strongly, as revealed by the higher densities of Dutch cities, the strong discontinuities observed for instance at the Franco-Spanish frontier and at the German-Polish border (Figure 14), or as suggested by the high densities of some Eastern countries like Poland (see Annex, Table 2).

**Figure 15 : European City sizes and densities (UMZ/CLC 2000)**



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

Regional level: NUTS 0 and NUTS 2  
Source: ESPON DB, year 2010  
Origin of data: The European Environment Agency (UMZ 2000 V.2), Joint Research Center (Density Grid V.5), LAU2 (2006, V.2) and national sources (see figure 9)  
© EuroGeographics Association for administrative boundaries



Furthermore, a strong and regular relationship with city size levels can be enhanced (Table 3): densities exceed 5700 inh./km<sup>2</sup> in cities larger than 2 millions inhabitants, then decrease regularly until 3000 inh./km<sup>2</sup> for cities between 10 000 and 25 000 inhabitants. This higher level of densities in the largest cities can be interpreted as the result of a historical accumulation process and as the expression of a more pronounced centrality and competition for land.

Let us recall that density indicator is of high interest for urban planning issues, for example in environmental topics, especially when it can be coupled with other



transportations indicators. Even if current debates enlighten a lack of consensus between researchers, we can mention for instance the question of the minimal city or sub-district density level necessary for providing efficient public transportation networks, or the one of the possible link between average city density level and pollution gas emissions.

**Table 3 : Urban population density per class of population**

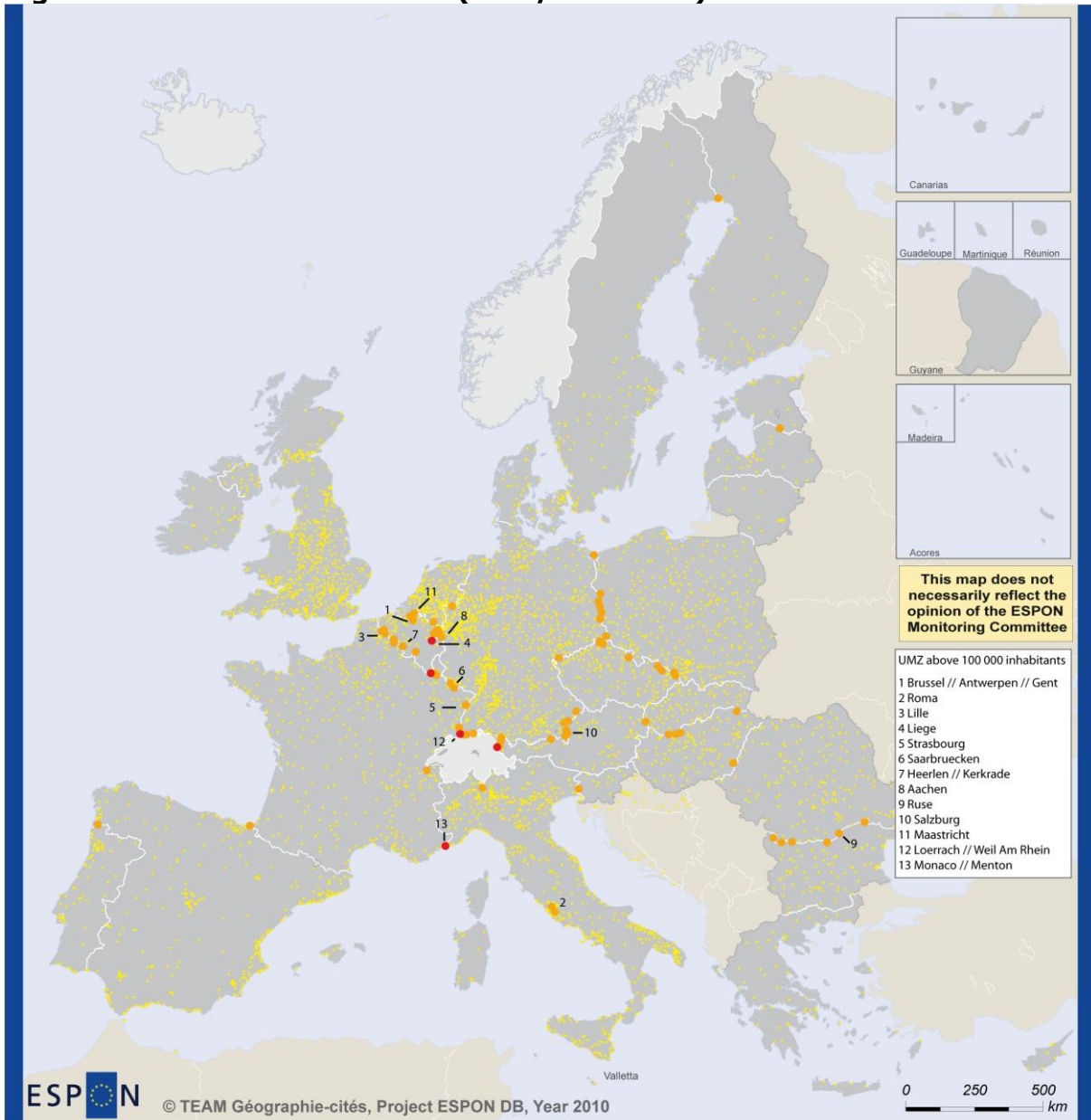
<b>Class of population</b>	<b>Number of UMZ</b>	<b>Density (inh./km<sup>2</sup>)</b>
<b>1 to 10 millions inh.</b>	39	4787
<b>0,5 to 1 millions inh.</b>	36	4 892
<b>250 to 500 thousands inh.</b>	136	4 235
<b>100 to 250 thousands inh.</b>	203	3 932
<b>50 to 100 thousands inh.</b>	512	3 469
<b>25 to 50 thousands inh.</b>	904	3 214
<b>10 to 25 thousands inh.</b>	2607	3 053

*Sources: LAU 2 (EuroBoundaryMap 2006, v2.0) from EuroGeographics, UMZ2000 from European Environment Agency, Population density Grid v.5 from Joint Research Center.*

### **3.3 International UMZ**

International UMZ can now be identified through a new indicator (international code) that describes the number of countries crossed by each UMZ. The distribution of international UMZ (Figure 16) is a first important result that offers an overview of cross-national cities, independent from institutional or administrative frames. Furthermore, an index of internalization (% of population living in one or more countries different than the main one) has been computed. It allows to qualify in a comparable way to what extent the city is embedded in a multi-national context and completes in a fruitful way the population indicator of these UMZ: for example, the most populated international UMZ is Brussels/Antwerpen/Gent, but it extends in a very small part in Netherlands (international index is only 1%). At the opposite, some UMZ located at the Poland/Germany, Slovakia/Hungary or Austria/Germany frontiers are not very populated but their international index is over 40% (Table 5). Two other tables are presented in Annex, with the most important international UMZ by countries, according to their population (see Annex, Table 3) and according to their international index (see Annex, Table 4).

**Figure 16 : International UMZ (UMZ/CLC 2000)**



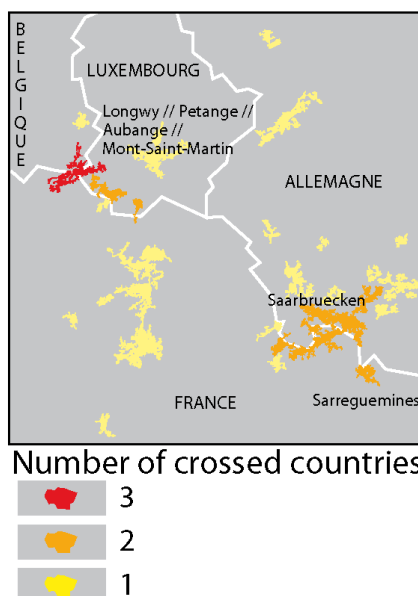
EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

Regional level: NUTS 0  
Source: ESPON DB, year 2010  
Origin of data: The European Environment Agency (UMZ 2000 V.2), Joint Research Center (Density Grid V.5), LAU2 (2006, V.2) and national sources (see figure 9)  
© EuroGeographics Association for administrative boundaries

**Number of crossed countries**

- 3
- 2
- 1
- no UMZ data

**Figure 17 : A zoom on some international UMZ**



Sources: LAU 2 (EuroBoundaryMap 2006, v2.0) from EuroGeographics, UMZ2000 from European Environment Agency, Population density Grid v.5 from Joint Research Center.

**Table 4 : Top Ten of main populated international UMZ**

UMZ Name	Population	Main Country	All countries*	International index**
<b>Brussels-Antwerpen-Gent</b>	3 769 885	BE	BE // NL	0,1
<b>Roma</b>	1 891 236	IT	IT // VA	0,03
<b>Lille</b>	1 335 026	FR	FR // BE	30,1
<b>Liege</b>	760 811	BE	BE // DE // NL	1,03
<b>Strasbourg</b>	435 410	FR	FR // DE	3,4
<b>Saarbruecken</b>	367 294	DE	DE // FR	28,5
<b>Heerlen-Kerkrade</b>	259 447	NL	NL // DE	12,6
<b>Aachen</b>	213 930	DE	DE // NL	3,9
<b>Ruse</b>	201 106	BG	BG // RO	35,1
<b>Salzburg</b>	181 407	AT	AT // DE	10,2

Sources: LAU 2 (EuroBoundaryMap 2006, v2.0) from EuroGeographics, UMZ2000 from European Environment Agency, Population density Grid v.5 from Joint Research Center.

\* Ranked by decreasing population.

\*\*% of population of UMZ that is not in the main country.

**Table 5: Top ten of most internationalised UMZ**

<b>UMZ Name</b>	<b>Population</b>	<b>Main country</b>	<b>All countries*</b>	<b>International index**</b>
<b>Comines/Wervik/ Comines-Warneton- Komen-Waasten</b>	33 796	FR	FR // BE	47,2
<b>Longwy- Petange – Aubange-Mont- Saint-Martin</b>	70 371	FR	FR // BE // LU	45
<b>Oberndorf bei Salzburg</b>	10 096	AT	AT // DE	44,1
<b>Braunau am Inn</b>	29 933	AT	AT // DE	42,2
<b>Monaco / Menton</b>	105 193	FR	FR // IT // MC	41,1
<b>Komarno</b>	44 404	SK	SK // HU	40,8
<b>Hamont</b>	16 854	NL	NL // BE	40,7
<b>Cieszyn</b>	53 152	PL	PL // CZ	40,6
<b>Guben</b>	34 413	DE	DE // PL	40,1
<b>Tui</b>	17 801	ES	ES // PT	39,9

Sources: LAU 2 (EuroBoundaryMap 2006, v2.0) from EuroGeographics, UMZ2000 from European Environment Agency, Population density Grid v.5 from Joint Research Center.

\* Ranked by decreasing population.

\*\*% of population of UMZ that is not in the main country.

## 4 Conclusion

UMZ present a great potential for future, as they result from the same building methodology and are defined with the same criteria in all the countries. They have been here completed and validated as a European database operational for urban studies: 4437 urban settlements over 10 000 inhabitants are now defined from CLC2000 with harmonized criteria (EEA, last version of Urban Morphological Zone shapes), population (JRC, last version of Population Density Grid), names and metadata. The establishment of an automated process for naming UMZ allows quick updating with new versions of sources or methods (EEA and JRC) or new dates (2006, 2010...). The validation of the method results from an expertise which selects the relevant data base for city names and relies on systematic matches with other sources (Eurostat, Geopolis). This protocol leads to a powerful data base for exploring the features of European cities in 2000 as regards to their settlement characteristics, distribution of city sizes, density patterns or international UMZ configurations.

Further work should improve the operational dimension of UMZ in future:

- Their integration in the ESPON Data Base could be hugely improved by building a zoning correspondence Table with LUZ or other functional database.
- Interoperability with other geo-referenced data bases (urban transport infrastructures, urban mobility, socio-economic LAU data...) opens a wide range of environmental and social studies. Enlargement of urban indicators, towards environmental (grid/raster) and socio-economic data should be realised by using the "OLAP Cube for Urban analysis" developed by UAB<sup>19</sup> and indicators collected at LAU2 level and aggregated in UMZ delineations
- Urban indicators were here computed for the year 2000. A very important challenge lies in the possibility of adding a temporal dimension to these indicators and making them vary in time. Development of a temporal urban data model would enable to follow UMZ urban indicators through time (1990, 2010 and future other dates).

---

<sup>19</sup> See Technical Report « Social/Environmental data », ESPON Database 2013, written by the Autonomous University of Barcelona.

## Annex:

**Table 1: UMZ ranking, "Top 50"**

Rank	Name	Country	Population 2001	Density (inh./km <sup>2</sup> )
1	Paris	FR	9 656 819	5 386
2	Greater London Urban Area	UK	8 221 307	4 861
3	Milano	IT	4 164 504	4 166
4	Essen / Duisburg / Dortmund / Bochum / Gelsenkirchen - 1	DE	3 892 380	3 674
5	Madrid - 1	ES	3 823 031	9 637
6	Brussel / Antwerpen / Gent	BE	3 769 885	1 841
7	Manchester	UK	3 546 819	3 538
8	Dimos Athinaion	GR	3 489 768	9 896
9	Berlin - 1	DE	3 367 457	4 648
10	Barcelona	ES	3 088 470	10 533
11	Napoli	IT	2 354 010	9 007
12	West Midlands Urban Area	UK	2 286 859	3 480
13	Budapest	HU	2 042 024	3 963
14	M. St. Warszawa - 1	PL	1 948 024	4 665
15	Bucaresti - 1	RO	1 925 741	9 066
16	Roma - 1	IT	1 891 236	6 300
17	Hamburg - 1	DE	1 838 019	3 549
18	M. Katowice / M. Sosnowiec / M. Gliwice / M. Zabrze / M. Bytom - 1	PL	1 810 260	3 845
19	Koeln - 1	DE	1 767 659	3 525
20	Wien	AT	1 756 034	4 304
21	Lisboa / Sintra	PT	1 749 316	5 906
22	Frankfurt Am Main - 1	DE	1 493 470	3 971
23	West Yorkshire Urban Area - 1	UK	1 473 892	3 541
24	Muenchen	DE	1 444 902	4 677
25	Lille	FR	1 335 026	2 831
26	Lyon	FR	1 287 802	3 092
27	Torino	IT	1 278 016	6 640
28	Stockholm	SE	1 233 147	2 822
29	Kobenhavn	DK	1 218 013	2 986
30	Porto / Vila Nova De Gaia / Matosinhos / Gondomar	PT	1 208 098	3 904
31	Wuppertal / Hagen /	DE	1 138 180	3 754
32	Glasgow - 1	UK	1 135 155	3 290
33	Sofia - 1	BG	1 079 088	6 271
34	Rotterdam - 1	NL	1 072 014	3 436
35	Tyneside - 1	UK	1 037 720	3 463
36	Dublin	IE	1 029 106	3 339

<b>Rank</b>	<b>Name</b>	<b>Country</b>	<b>Population 2001</b>	<b>Density (inh./km<sup>2</sup>)</b>
<b>37</b>	<b>Amsterdam</b>	NL	1 028 359	4 270
<b>38</b>	<b>Praha - 1</b>	CZ	1 020 584	4 303
<b>39</b>	<b>Valencia - 1</b>	ES	967 206	10 874
<b>40</b>	<b>Helsinki-helsingfors - 1</b>	FI	917 813	2 218
<b>41</b>	<b>Marseille</b>	FR	902 756	5 215
<b>42</b>	<b>M. Lodz - 1</b>	PL	822 545	5 232
<b>43</b>	<b>Bilbao</b>	ES	819 465	13 869
<b>44</b>	<b>Nice - 1</b>	FR	812 330	3 092
<b>45</b>	<b>Duesseldorf - 1</b>	DE	809 770	3 899
<b>46</b>	<b>Dimos Thessalonikis</b>	GR	804 095	11 600
<b>47</b>	<b>Sevilla</b>	ES	797 127	7 984
<b>48</b>	<b>Palermo</b>	IT	786 622	7 058
<b>49</b>	<b>Liege</b>	BE	760 811	1 401
<b>50</b>	<b>Gijon</b>	UK	736 438	3 276

Sources: LAU 2 (EuroBoundaryMap 2006, v2.0) from EuroGeographics, UMZ2000 from European Environment Agency, Population density Grid v.5 from Joint Research Center.

**Table 2 : Urban population density per country**

<b>Country</b>	<b>Number of UMZ</b>	<b>Density (inh./km<sup>2</sup>)</b>
<b>Albania</b>	19	-
<b>Austria</b>	60	1 986
<b>Bosnia and Herzegovina</b>	21	4 602
<b>Belgium</b>	50	1 449
<b>Bulgaria</b>	83	4 207
<b>Cyprus</b>	5	1 700
<b>Czech Republic</b>	116	2 889
<b>Germany</b>	846	2 799
<b>Denmark</b>	51	1 631
<b>Estonia</b>	16	2 396
<b>Spain</b>	348	5 752
<b>Finland</b>	44	1 135
<b>France</b>	391	2 037
<b>Greece</b>	51	4 493
<b>Croatia</b>	36	-
<b>Hungary</b>	107	2 202
<b>Ireland</b>	26	2 359
<b>North Ireland</b>	1	3 032
<b>Italy</b>	575	3 922
<b>Kosovo</b>	16	-
<b>Liechtenstein</b>	1	1 265
<b>Lithuania</b>	16	3 509
<b>Luxembourg</b>	3	2 532
<b>Latvia</b>	23	2 730
<b>Monaco</b>	1	5 818
<b>Macedonia</b>	20	-
<b>Malta</b>	1	3 297
<b>Netherlands</b>	201	3 398
<b>Poland</b>	327	3 658
<b>Portugal</b>	70	3 545
<b>Romania</b>	159	3 249
<b>Serbia</b>	58	-
<b>Sweden</b>	88	1 316
<b>Slovenia</b>	12	2 943
<b>Slovakia</b>	66	3 151
<b>San Marino</b>	1	4 314
<b>United Kingdom</b>	528	3 114

Sources: LAU 2 (EuroBoundaryMap 2006, v2.0) from EuroGeographics, UMZ2000 from European Environment Agency, Population density Grid v.5 from Joint Research Center.



**Table 3 : Main international UMZ by country (ranked by population)**

Country	Number of international UMZ	Main international UMZ		
		Population	International index*	Name
<b>Austria</b>	6	181 407	16,2	Salzburg
<b>Belgium</b>	4	3 769 885	0,1	Brussel / Antwerpen / Gent
<b>Bulgaria</b>	6	201 106	35,1	Ruse
<b>Czech Republic</b>	5	64 774	0,8	Karvina
<b>Germany</b>	19	367 294	28,5	Saarbruecken
<b>Estonia</b>	1	20 500	39,2	Valga
<b>Spain</b>	2	83 206	15,7	Irun
<b>Finland</b>	1	12 026	36,9	Tornio-Tornea
<b>France</b>	12	1 335 026	30,1	Lille
<b>Hungary</b>	4	26 606	36,9	Esztergom
<b>Italy</b>	4	1 891 236	0,03	Roma
<b>Lithuania</b>	1	10 029	5,8	Mauren / Eschen / Ruggell
<b>Luxembourg</b>	2	62 104	1,3	Esch-Alzette / Differdange
<b>Netherlands</b>	6	259 447	12,6	Heerlen / Kerkrade /
<b>Poland</b>	3	53 152	40,6	Cieszyn
<b>Slovakia</b>	2	44 404	40,8	Komarno

Sources: LAU 2 (EuroBoundaryMap 2006, v2.0) from EuroGeographics, UMZ2000 from European Environment Agency, Population density Grid v.5 from Joint Research Center.

\*% of population of UMZ that is not in the main country.

**Table 4 : Main international UMZ (ranked by International index)**

Country	Number of International UMZ	Main international UMZ		
		International index*	Name	Population
<b>Austria</b>	6	44,1	Oberndorf bei Salzburg	10 096
<b>Belgium</b>	4	8,5	Essen (BG)	14 407
<b>Bulgaria</b>	6	35,1	Ruse	201 106
<b>Czech Republic</b>	5	31,1	Nachod	29 071
<b>Germany</b>	19	40,1	Guben	34 413
<b>Estonia</b>	1	31	Valga	20 500
<b>Spain</b>	2	39,8	Tui	17 801
<b>Finland</b>	1	36,9	Tornio-Tornea	12 026
<b>France</b>	12	47,2	Comines / Wervik / Comines-Warneton - Komen-Waasten	33 796
<b>Hungary</b>	4	36,9	Esztergom	27 000
<b>Italy</b>	4	36,7	Gorizia	44 518
<b>Lithuania</b>	1	5,8	Mauren / Eschen / Ruggell	10 029
<b>Luxembourg</b>	2	8,3	Dudelange	18 284
<b>Netherlands</b>	6	40,7	Hamont	16 854
<b>Poland</b>	3	40,6	Cieszyn	53 152
<b>Slovakia</b>	2	40,8	Komarno	44 404

Sources: LAU 2 (EuroBoundaryMap 2006, v2.0) from EuroGeographics, UMZ2000 from European Environment Agency, Population density Grid v.5 from Joint Research Center.

\*% of population of UMZ that is not in the main country.

## Figures list:

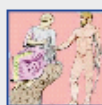
<b>Figure 1 : Naming methodology (Situation 1, 2 and 3)</b> .....	9
<b>Figure 2 : Leipzig (Germany), an UMZ with one strong core (Situation 1)</b> .....	9
<b>Figure 3 : Bayonne-Anglet-Biarritz (France), an UMZ with several cores (Situation 2)</b> .....	10
<b>Figure 4: Sandominic (Romania), an UMZ with one weak core (situation 3)</b> .....	11
<b>Figure 5 : Łódź (Poland), two UMZ with the same name</b> .....	12
<b>Figure 6 : UMZ 2000 with identical names</b> .....	13
<b>Figure 7 : The 12 UMZ sharing the same name "Roma"</b> .....	14
<b>Figure 8 : The different steps of the processing chain</b> .....	15
<b>Figure 9 : Relevant reference units for naming UMZ</b> .....	17
<b>Figure 10 : Some examples of topological errors</b> .....	19
<b>Figure 11 : Steps of the validation process</b> .....	22
<b>Figure 12 : UMZ typology according to naming results</b> .....	24
<b>Figure 13 : Pareto-Zipf distribution of city sizes (UMZ 2000 data base)</b>	25
<b>Figure 14 : European cities sizes and densities (UMZ/CLC 2000)</b> .....	27
<b>Figure 15 : European City sizes and densities (UMZ/CLC 2000)</b> .....	28
<b>Figure 16 : International UMZ (UMZ/CLC 2000)</b> .....	30
<b>Figure 17 : A zoom on some international UMZ</b> .....	31

## Tables:

<b>Table 1 : Naming UMZ through automatic methods and LAU2 level</b> .....	22
<b>Table 2 : "Top ten" UMZ compared to MUA and LUZ (2000, population in thousand inhabitants)</b> .....	26
<b>Table 3 : Urban population density per class of population</b> .....	29
<b>Table 4 : Top Ten of main populated international UMZ</b> .....	31
<b>Table 5:Top ten of most internationalised UMZ</b> .....	32

## Annex Tables:

<b>Table 1: UMZ ranking, "Top 50"</b> .....	34
<b>Table 2 : Urban population density per country</b> .....	36
<b>Table 3 : Main international UMZ by country (ranked by population)</b> ....	37
<b>Table 4 : Main international UMZ (ranked by International index)</b> .....	38



# LUZ specifications (Urban Audit 2004)

## MAIN RESULTS

- **Stakes:** Larger Urban Zones are based on various national definitions (functional areas, planning regions, local administrative units, etc.). Only a good knowledge of specifications allows identifying bias resulting from the national heterogeneity of these definitions
- **Methodology:** after collecting and expertizing documentation on national specifications, a common "syntax" has been used for categorizing specifications (rules of construction, building blocks, evolution since UA II etc.)
- **Results:** 1) Four general maps enlighten a synthetic typology of definitions, the evolution since UA II, the diversity of thresholds in commuting-based approaches and the specificity of capital city LUZ definitions 2) Thirty country-sheets summary specifications following a common syntax

ESPON 2013 DATABASE

MARCH 2011



# LIST OF AUTHORS

Anne Bretagnolle, University Paris 1, UMR Géographie-cités

François Delisle, UMR Géographie-cités

Hélène Mathian, C.N.R.S., UMR Géographie-cités

Liliane Lizzi, C.N.R.S., UMR Géographie-cités

Marianne Guérois, University Paris 7, UMR Géographie-cités

Guilhain Averlant, UMR Géographie-cités

## Contact

[anne.bretagnolle@parisgeo.cnrs.fr](mailto:anne.bretagnolle@parisgeo.cnrs.fr)

tel. + 33 1 01 40 46 40 00

# TABLE OF CONTENT

<b>1</b>	<b>STAKES AND MATTER .....</b>	<b>4</b>
1.1	A BOTTOM-UP APPROACH FOR BUILDING A EUROPEAN URBAN DATA BASE.....	4
1.2	DETECTING BIAS RESULTING FROM HETEROGENEITY IN LUZ DEFINITIONS .....	4
<b>2</b>	<b>LARGER URBAN ZONE PRESENTATION .....</b>	<b>6</b>
2.1	URBAN AUDIT ROUNDS.....	6
2.2	LUZ DEFINITION BY URBAN AUDIT .....	8
2.2.1	The three-level approach of the European Cities .....	8
2.2.2	Selection criteria of UA cities.....	9
2.2.3	Documentation (National Reports) .....	12
<b>3</b>	<b>EXPERTIZE METHODOLOGY .....</b>	<b>14</b>
3.1	DOCUMENTATION DATA PROCESS.....	14
3.1.1	Expertize of National Reports .....	14
3.1.2	Expertize of other documentation .....	15
3.2	CONSTRUCTION OF THE GENERAL SYNTAX.....	16
3.3	UNSOLVED PROBLEMS .....	17
<b>4</b>	<b>RESULTS.....</b>	<b>18</b>
4.1	TPOLOGY OF LUZ DELINEATIONS.....	18
4.1.1	LUZ as one elementary administrative unit .....	19
4.1.2	Aggregations of neighbouring units .....	19
4.1.3	Aggregations mainly based on commuting data .....	19
4.1.4	Aggregation with no specification.....	19
4.1.5	Planning regions or local consultations .....	20
4.1.6	No generic rules .....	20
4.1.7	No precise information .....	20
4.2	DIVERSITY OF COMMUTING THRESHOLDS.....	20
4.3	CAPITAL CITIES AS PARTICULAR CASES .....	21
4.4	TOWARDS MORE FUNCTIONAL APPROACHES .....	22
<b>5</b>	<b>CONCLUSION.....</b>	<b>24</b>
	<b>ANNEX : LUZ SPECIFICATIONS BY COUNTRY (UA III).....</b>	<b>25</b>
	<b>REFERENCES.....</b>	<b>72</b>

# **1 Stakes and matter**

## **1.1 A bottom-up approach for building a European urban data base**

Urban Audit<sup>1</sup> (UA) was conducted at the initiative of the Directorate-General for Regional Policy at the European Commission. It aims collecting comparable statistics and indicators for cities, at three different scales (Sub-Districts, City Core and Larger Urban Zones). A large number of indicators (more than one hundred) are sent by countries for each of these levels of definition, and the interest of such variables for urban studies is not to be proved. However, several questions concerning international comparability of indicators results are specifically raised at the larger scale definition, the LUZ one.

Indeed, the work engaged by Urban Audit for collecting and possibly harmonizing LUZ delineations represent a specific approach for trying to build a European set of cities. As opposed to Urban Morphological Zones<sup>2</sup>, it is not a top-down approach (starting from identical definition criteria and trying to enrich it by taking into account national diversity) but a bottom-up approach. Countries are required by Urban Audit to choose and send national definitions of LUZ, sometimes changing them when taking into account some recommendations.

## **1.2 Detecting bias resulting from heterogeneity in LUZ definitions**

Before using the large amount of available indicators of Urban Audit, some cautions have to be taken: LUZ definitions are so different from one country to another that an expertise of national specifications must be done in order to identify bias resulting from this heterogeneity.

The first step in our expertise was to collect a huge set of documentation on LUZ specifications (National Report sent at each round of Urban Audit, other information on Statistical Census Boards websites, annex documentations gathered by Urban Audit and kindly sent to us<sup>3</sup>). A second step was dedicated to the construction of a common syntax for describing these specifications, using a common vocabulary and approach in order to understand the national logics in these rules.

---

<sup>1</sup> <http://www.urbandaudit.org/>

<sup>2</sup> See Technical Report "Naming UMZ: making them more operational for urban studies", ESPON DB 2013.

<sup>3</sup> We want here to thank warmly Teodora Brandmuller, from Urban Audit, without her this work could not have been achieved. She helped us several times to understand some complicated parts of national documentations and she sent us, at different times, a lot of files that fortunately filled some lack in our collect.

This technical report is divided in three parts: in the first one we recall some specificities of Urban Audit and we present the Larger Urban Zones (criteria of selection, evolution, national sources etc.). In the second part we present the methodology used for analyzing LUZ specifications, in particular the general model of country-sheet that has been used for re-writing in common vocabulary and syntax LUZ specification for each of the 30 countries under study. In the third part, results (typologies and maps) are presented and discussed. The 30 country-sheet on LUZ specifications are presented in Annex.



## 2 Larger Urban Zone presentation

### 2.1 Urban Audit rounds

Urban Audit, under the coordination of Eurostat, aims gathering comparable data covering most aspects of urban life in European cities and towns. National Urban Audit Coordinators (NUAC), generally represented by National Statistics Offices, are the link between Eurostat and the cities involved. They collect and gather data in their country before transmitting it to Eurostat.

Three different rounds occurred until now (Table 1). A first phase (pilot phase) was launched in 1998, a second one between 2003 for Member States and 2004 for Candidate Countries (UA II 2001), then a third between 2006 and 2007 (UA III 2004). The next round (UA IV 2008) is ongoing and data dissemination is expected for 2011.

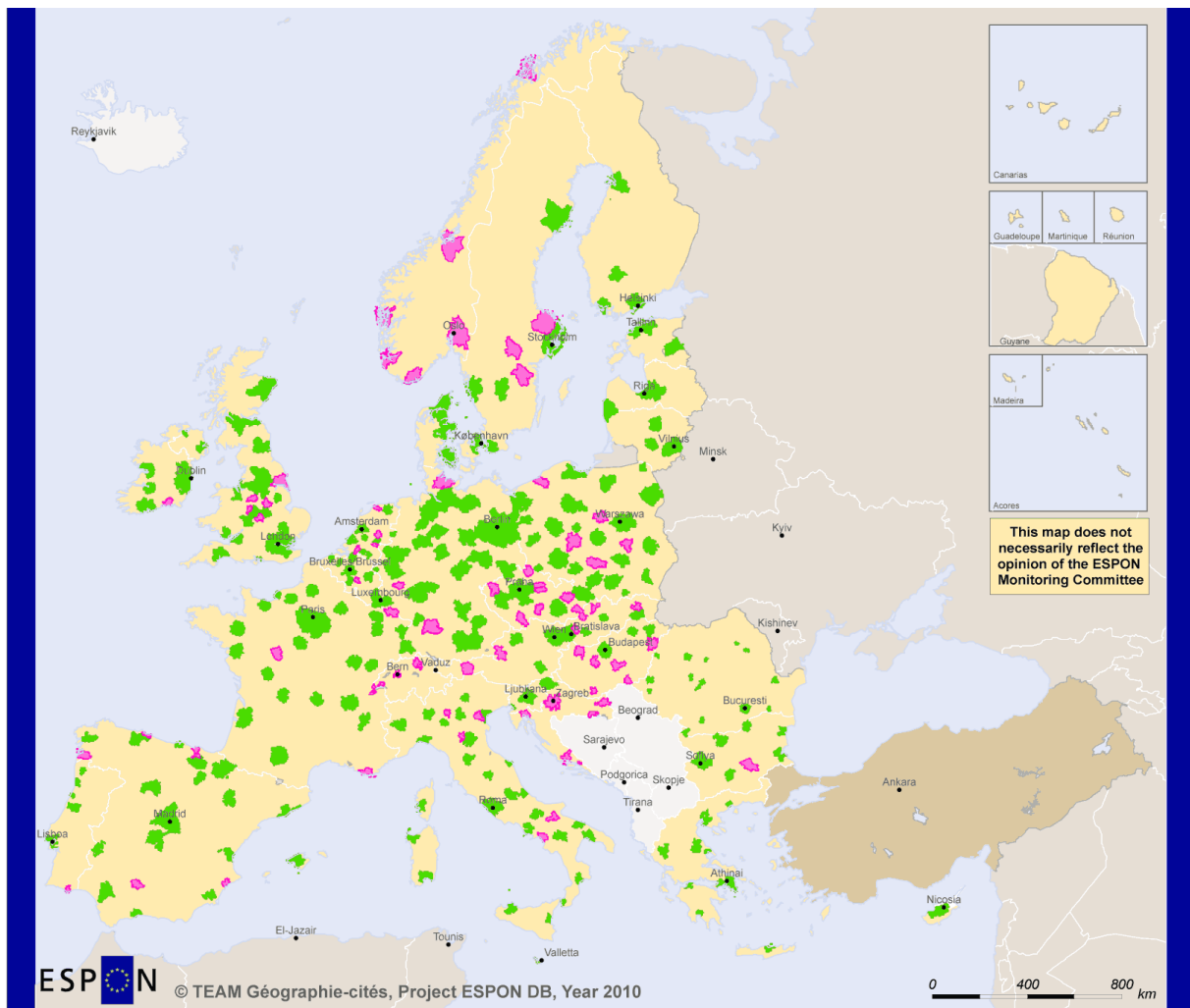
**Table 1: Number of cities, countries and indicators involved in the three Urban Audit rounds**

	<b>UAPP (1998)</b>	<b>UA II (2001)</b>	<b>UA III (2004)</b>
<b>Number of cities</b>	58	189 MS +69 CC = <b>258</b>	321+46 = <b>367</b>
<b>Number of countries</b>	15	15 +12 = <b>27</b>	27 MS + Norway + Switzerland + Croatia + Turkey = <b>31</b>
<b>Number of indicators</b>	500	336	338
<b>Reference year</b>	1981, 1991, 1996	2001	2004
<b>Launching year</b>	1998	2003-2004	2006-2007

UAPP: Urban Audit Pilot Phase; MS: Member State; CC: Candidate Countries

Some countries which are not included in the UE perimeter have participated to Urban Audit III and we have integrated them in this expertise when finding some data, i.e. Croatia, Norway and Switzerland. For Turkey, we did not found any information concerning LUZ so that we did not consider it here. We have worked on 30 countries in all (Figure 1)

**Figure 1: UA II and UA III LUZ**



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

- LUZs in UA II
- LUZs added in UA III

- No UA III
- Out of ESPON space
- No precise information

Regional level: NUTS 0  
Source: ESPON DB, year 2010  
Origin of data: Urban Audit, year 2010  
© EuroGeographics Association for administrative boundaries

Source: GISCO. Some mismatches between GISCO<sup>4</sup> and National Reports are described in the Country-sheets(Annex) and must be underlined on this map. Some LUZ are missing (the 26 LUZ from Turkey and 5 of the 9 LUZ from Switzerland), some others should be removed (Toulon LUZ in France).

<sup>4</sup> Geographical Information System of the European Commission ([http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)).

## 2.2 LUZ definition by Urban Audit

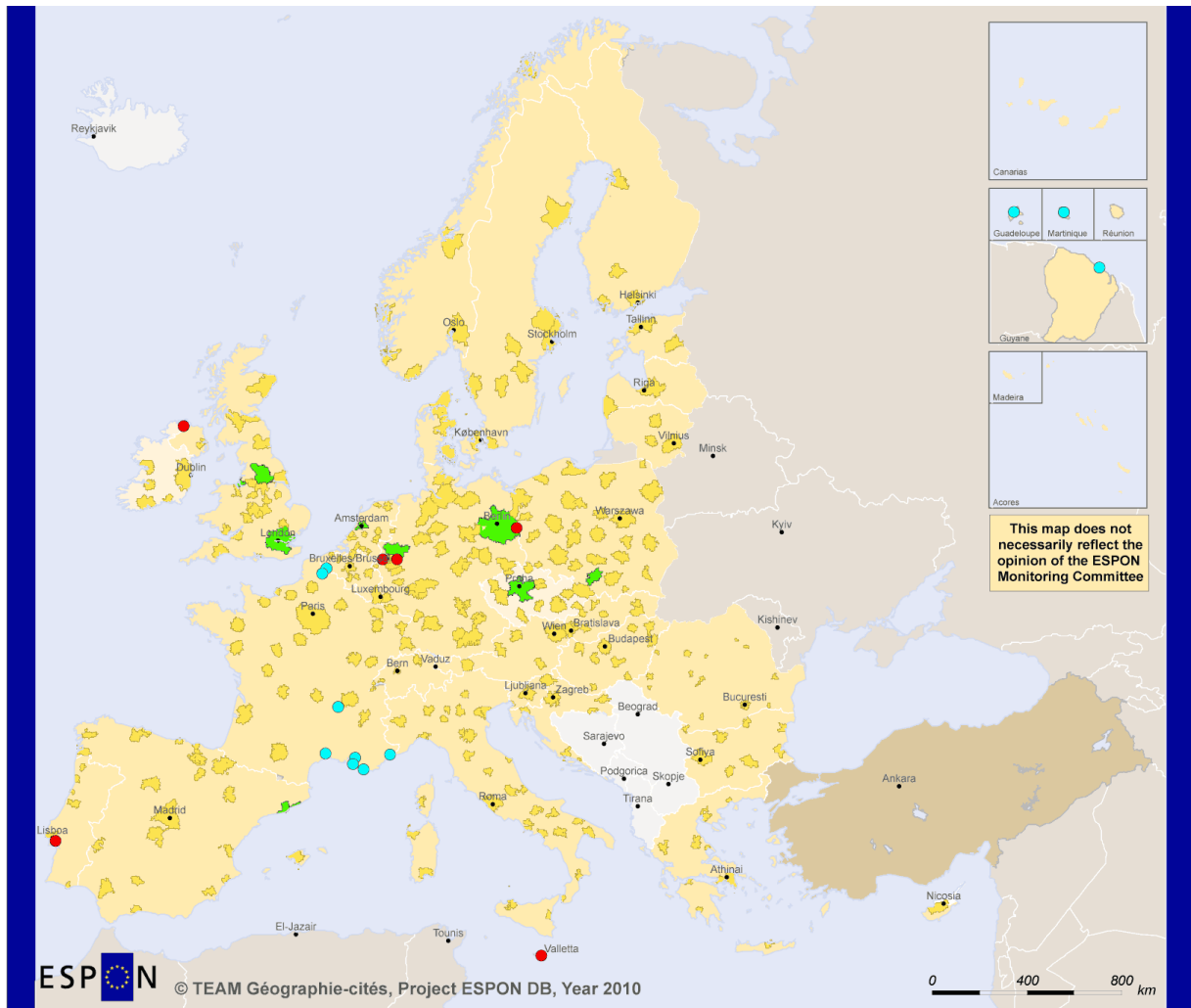
### 2.2.1 The three-level approach of the European Cities

Since Urban Audit II, urban entities that have been defined as “cities” may correspond to three representations:

- The **Core City** (CC) relates to an administrative approach of cities and towns, and generally fits with the eponymous central LAU2. In order to facilitate international comparisons, some *Kernels* have been created for larger European cities, and consist in an aggregation of LAU2.
- The **Sub-City District** (SCD) consists in a subdivision of the city according to population criteria.
- The **Larger Urban Zone** (LUZ) is conceived by Urban Audit to approach the functional urban region definition. It must “*allow a comparison between the city and its surroundings. The goal was to have an area from a significant share of the resident commute into the city, a concept known as the ‘functional urban region’.* To ensure a good data availability, the Urban Audit works with administrative boundaries that approximate the functional urban region” (<http://www.urbanaudit.org/help.aspx>).

In most of the cases, Urban Audit data collections concern European Cities simultaneously at the three different scale-levels. However, each Urban Audit “city” hasn’t systematically the three representations: some Urban Audit cities have no LUZ but one CC, other have the same perimeter for LUZ and CC; sometimes two CC share the same LUZ...The map presented below (Figure 2) gives a synthetic view of these particular cases in LUZ-CC articulations.

**Figure 2: A variety of representations of Urban Audit "cities" (UA III)**



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

Regional level: NUTS 0  
Source: ESPON DB, year 2010  
Origin of data: Urban Audit, GISCO, year 2010  
© EuroGeographics Association for administrative boundaries

**City representations in terms of city core and LUZ**

- LUZ with a City Core inside
- LUZ including several City Cores.
- City Core and LUZ have the same perimeter
- City Core has no LUZ definition
- No precise national information
- No UA III
- Out of ESPON space

**2.2.2 Selection criteria of UA cities**

Urban Audit cities and the three-scale perimeters for data collection are selected according to specific criteria, that are unusual in the field of urban studies in the sense that classical hierarchical criteria (minimal population threshold) are completed by geographical criteria (spatial dispersion within each country) and administrative criteria (inclusion of national and regional capital) (see Insert 1).

**Insert 1: Selection criteria of Urban Audit cities (<http://www.urbanaudit.org/help.aspx>)**

The results give a heterogeneous set of cities, as displayed on the following

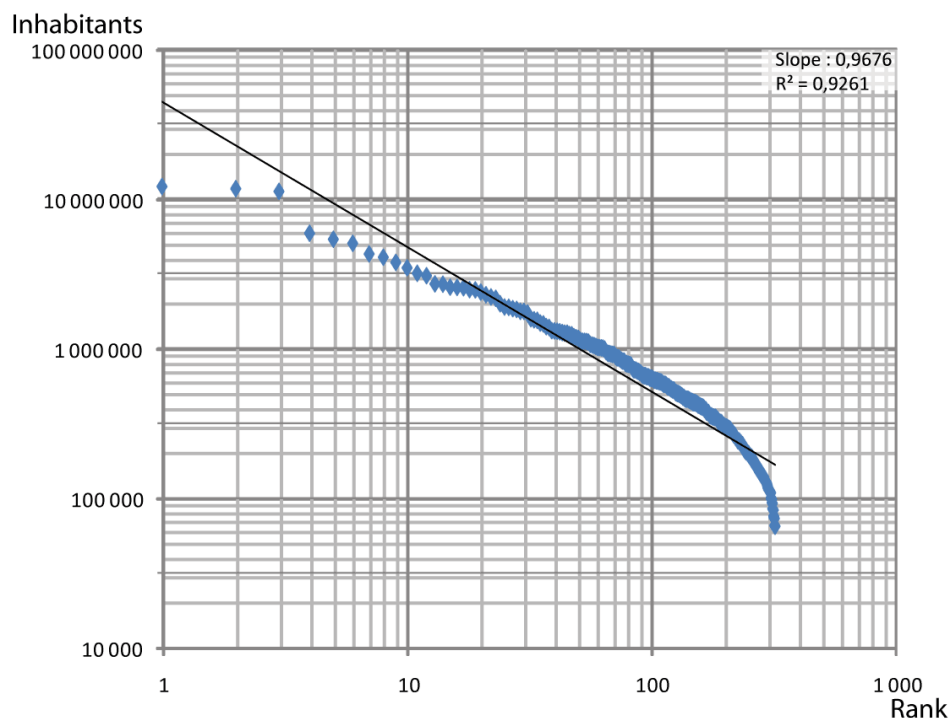
*"What is Urban? The Urban Audit aims at a balanced and representative sample of cities in Europe. To obtain such a selection, a few simple rules were followed:*

- 1. Approximately 20% of the national population should be covered by the Urban Audit.*
- 2. All capital cities were included.*
- 3. Where possible, regional capitals were included.*
- 4. Both large (more than 250 000 inhabitants) and medium-sized cities (minimum 50 000 and maximum 250 000 inhabitants) were included.*
- 5. The selected cities should be geographically dispersed within each Member State".*

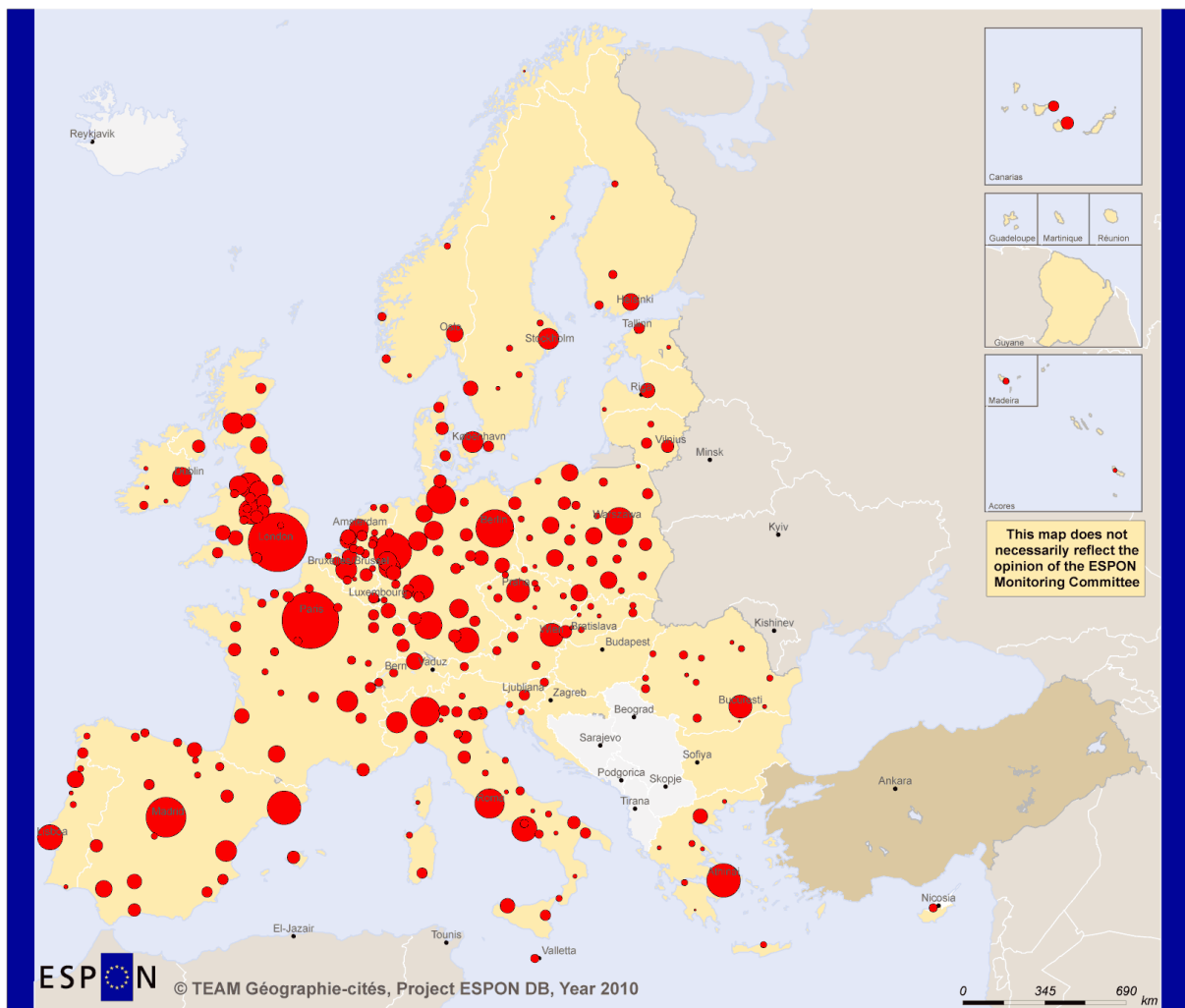
figure. On the map representing LUZ population (Figure 3), the information seems to be harmonious and close to the one obtained, for example, from the city-size patterns given by Urban Morphological Zone data base (see the Technical Report Naming UMZ, ESPON database 2013).

However, when plotting the distribution of LUZ sizes on a rank-size graph (figure 4), the queue-distribution appears very different from the rest of the city sizes. This is due to the very small number of LUZ under 120 000 inhabitants, under-represented compared to the size of other classes. Some of these small LUZ are indicated on Table 2 (see for example in Germany, Norway, Greece etc.).

**Figure 3: Pareto-Zipf distribution of LUZ size (UA III)**



**Figure 4: Population of LUZ in 2003-2006 (UA III)**



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

Regional level: NUTS 0  
Source: ESPON DB, year 2011  
Origin of data: Urban Audit, year 2010  
© EuroGeographics Association for administrative boundaries

**Inhabitants**

- 100 000
- 1 000 000
- 10 000 000

- No UA III
- Out of ESPON space
- No precise information

*NB: No data for Bulgaria and Hungary (see table 2)  
Some LUZ population are missing on Eurostat website<sup>4</sup>.*

In a more general way, one can observe the fact that the distribution is not well adjusted by the model (the coefficient of determination  $R^2$  is only 0.93) whereas it is not the case for more classical distributions of European city-sizes (see UMZ, in the Technical Report "Naming UMZ" of ESPON database 2013, or Geopolis data base, in François Moriconi-Ebrard, 1993, where the value is over 0.98).

<sup>5</sup> [http://epp.eurostat.ec.europa.eu/NavTree\\_prod/everybody/BulkDownloadListing?sort=1&file=data%2Furb\\_vluz.tsv.gz](http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing?sort=1&file=data%2Furb_vluz.tsv.gz)

**Table 2: LUZ number and min-max population by country (UA III)**

Countries	Nb of LUZ in UA III	Population min and max (2003-2006)	Countries	Nb of LUZ in UA III	Population min and max (2003-2006)
Austria	5	272 677 2 179 769	Latvia	2	131 788 1 003 949
Belgium	7	139 024 1 800 663	Lithuania	3	160 656 709 870
Bulgaria	8	No data	Luxembourg	1	451 600
Croatia	5	213 396 <sup>6</sup>	Malta	2	370 704
Cyprus	1	289 100	Netherland	14	158 883 1 443 258
Czech Republic	13	108 292 1 964 750	Norway	6	64 303 1 090 513
Denmark	4	475 082 1 822 569	Poland	27	82 539 2 660 406
Estonia	2	148 872 521410	Portugal	9	111 782 2 435 837
Finland	4	196 096 1 224 107	Romania	14	72 600 2 140 194
France	24	92 633 11 532 409	Slovakia	8	111 419 601 132
Germany	35	65 242 5 302 179	Slovenia	2	319 426 495 101
Greece	9	73 434 4 013 368	Spain	24	167 036 5 804 829
Hungary	9	No data	Sweden	8	139 588 1 860 872
Ireland	5	84 489 1 534 426	Switzerland	4	91 437 1 116 089
Italy	32	99 887 3 419 287	United Kingdom	26	253 500 11 917 000
			<b>Total</b>	313	

### 2.2.3 Documentation (National Reports)

The National Reports consist in a specific document sent by the NUAC (National UA Correspondent) at each UA round, which describes the general context and the specific conditions of the data collection. Even if National Reports are very different from one to another (for instance, the shortest is only 6 pages and the largest 124 pages), one can recognize more or less 4 main fields:

- *Overviews*: a general description of the data collection.
- *Spatial units description*: they are described for each of the three definition levels (City-Core, Sub-City Districts and LUZ). Usually the specifications of the LUZ definition are presented in this part of the Report.

---

<sup>6</sup> For Croatia, LUZ population is available only for one LUZ on Eurostat website ([http://epp.eurostat.ec.europa.eu/NavTree\\_prod/everybody/BulkDownloadListing?sort=1&file=data%2Furb\\_vluz.tsv.gz](http://epp.eurostat.ec.europa.eu/NavTree_prod/everybody/BulkDownloadListing?sort=1&file=data%2Furb_vluz.tsv.gz)).

- *Indicators descriptions*: this part may contain lists of data available, meta information, or quality aspects of variables.

- *Conclusion*: some recommendations for improvement or other particular aspects can be mentioned there.

The availability of National Reports by country depends not only on each Urban Audit round but also on particular cases (for example Czech Republic and Ireland didn't send National Report for UA III). A general Table sums up these national differences (Table 3). Every National Report can be uploaded for consultation at the Communication & Information Resource Centre Administrator (Circa) of Eurostat.

**Table 3: Availability of National Report per country and per Urban Audit round**

Countries	UA II	UA III	Countries	UA II	UA III
Austria	✓	✓	Ireland	✓	✗
Belgium	✓	✓	Italy	✓	✓
Bulgaria	✓	✓	Lithuania	✓	✓
Switzerland		✓	Luxembourg	✓	✓
Cyprus	✓	✓	Latvia	✓	✓
Czech Republic	✓	✗	Malta	✓	✗
Germany	✓	✓	Netherland	✓	✓
Denmark	✓	✓	Norway		✓
Estonia	✓	✓	Poland	✓	✓
Spain	✓	✓	Portugal	✓	✓
Finland	✓	✓	Romania	✓	✓
France	✓	✓	Sweden	✓	✓
Greece	✓	✓	Slovenia	✓	✓
Croatia		✓	Slovakia	✓	✓
Hungary	✓	✓	United Kingdom	✓	✓



## **3 Expertize methodology**

### **3.1 Documentation data process**

#### **3.1.1 Expertize of National Reports**

National Reports constitute the main source of information on LUZ specifications. They have been analyzed through different steps:

- Identification of coherence between the different phases of UA (which countries do participate, which cities, which LUZ?)
- Research of regularities in the national descriptions of the process, in order to point out categories that may be systematically extracted (city implications, LUZ evolutions...)
- Identification of the degree of completeness of the information regarding these categories.
- Research of correspondence with the geometric sources of LUZ, provided by GISCO (Geographical Information System of the European Commission<sup>7</sup>)

During these processes, different problems have been raised and have made the work highly complex. Some of them are purely formal (for example the National Reports which are not translated in English, a problem that we have naturally solved for France and Belgium but not for Norway). Some others are related to the content (for instance, allusive description of LUZ definition, or lack of explanation concerning LUZ definition) (Insert 2).

---

<sup>7</sup>[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)

## **Insert 2: Main problems encountered in National Report expertise (see country-sheets in Annex for more details)**

- National Reports or technical annexes are not written in English (France, Belgium, Norway) ;
- LUZ definitions refer to a national zoning whose construction rules are not specified (for example the *Stadsgewest* of the Netherlands, the *Stobyregioner* of Norway or the *Local Labour Market areas* of Sweden, resulting from a collaboration with Eurostat).
- LUZ building blocks are described in allusive terms (for example the *Amter* of Denmark).
- The National Report does not give any information about LUZ definition (for example in Luxembourg).
- There is no National Report (Ireland, Czech Republic).
- Building Blocks mentioned in National Reports are difficult to compare with the last version of NUTS or LAU available at European scale. Sometimes the version of LAU 2 is not the same, or LAU 1 are not available for the related country.
- Information given in National Report is sometimes in contradiction with information given in GISCO (for example when cities have their City Core inside another LUZ, see Wirral in United Kingdom).

### **3.1.2 Expertize of other documentation**

About one dozen of National Reports contained a fully specified description of LUZ definitions that did not need to be completed by other documentation (Austria, Bulgaria, Croatia, Cyprus, Estonia, Italy, Poland, Portugal, Romania, United Kingdom...).

For the rest of National Reports, we used other sources of information in order to try to fill the lack of information. Different situations occurred:

#### **Situation 1: LUZ is defined as a NUTS or LAU proxy<sup>8</sup>.**

We have then checked this information, using GISCO (Austria, Cyprus, Estonia, Latvia, Lithuania, Denmark, Slovakia) or search more information, especially when some reports or urban analyses were mentioned in the National Report for justifying the NUTS or LAU proxy (Austria, Latvia, Lithuania, Slovakia).

---

<sup>8</sup> See Section 4.1.1

**Situation 2: LUZ is defined using a national zoning not fully described in the National Report.**

We have then tried to find the documents giving the precise methodology on National Statistical Boards websites (for example, the *région urbaine* of Belgium, the *aire urbaine* of France, the *agglomeration* of Switzerland, the *planungsregionen* of Germany, the *stadsgewest* of Netherlands, the *storbyregioner* of Norway). For Netherlands and Norway, some documents were found but not in English.

**Situation 3: LUZ is defined using an allusive reference to a national zoning definition.**

Again, National Statistical Boards websites were visited (for example concerning “labour market areas” in Sweden or «suburban communes» in Romania). Another complementary method consisted to find information on national functional areas (Statistical Boards websites, publications) and to compare them to the LUZ perimeters (for example in Sweden).

**Situation 4: LUZ is not defined, or in a very imprecise way.**

We have then contacted Urban Audit Team (Teodora Brandmuller) and received in most of the cases some additional information (for example Greece, Spain, Finland).

## **3.2 Construction of the general syntax**

Different fields have been chosen for describing, in a common language and syntax, the rules used by each country to construct their LUZ. These fields and their content are described in Insert 3.

### Insert 3: Model of the Country-Sheet

NAME OF THE COUNTRY
<b>Summary</b>
◆ Number of cities with a LUZ in UA III and previous UA ■ Changes in LUZ definitions between UA II and UA III ● Description of LUZ delineation rules in UA III
<b>Building blocks</b>
◆ Identification of the building blocks of the LUZ ■ Links between building blocks when LUZ is an aggregation ● LUZ capital specificity (relatively to the other national LUZ)
<b>Particular cases</b>
Particular observations, not concerning LUZ capital
<b>Correspondence with GISCO</b>
Coherence between National Reports and GISCO shape files (number of LUZ, names etc.)
<b>References</b>
Documentation used for building the sheet (National Reports and other sources)

### 3.3 Unsolved problems

In some cases, it was not possible to achieve the expertise because of a remaining lack in information. In these cases, the country-sheet is not complete but missing information is specifically underlined. The most striking cases are:

- Netherlands and its "Stadsgewesten" (only a rough summary of the definition was found in English)
- Hungary and its LAU1 (statistical entities but without information on aggregation criteria)
- Sweden and its "Local Labour Market Areas" (defined in collaboration with Eurostat but we did not find more information)
- Finland and its LUZ (no information on LAU 2 aggregation criteria).

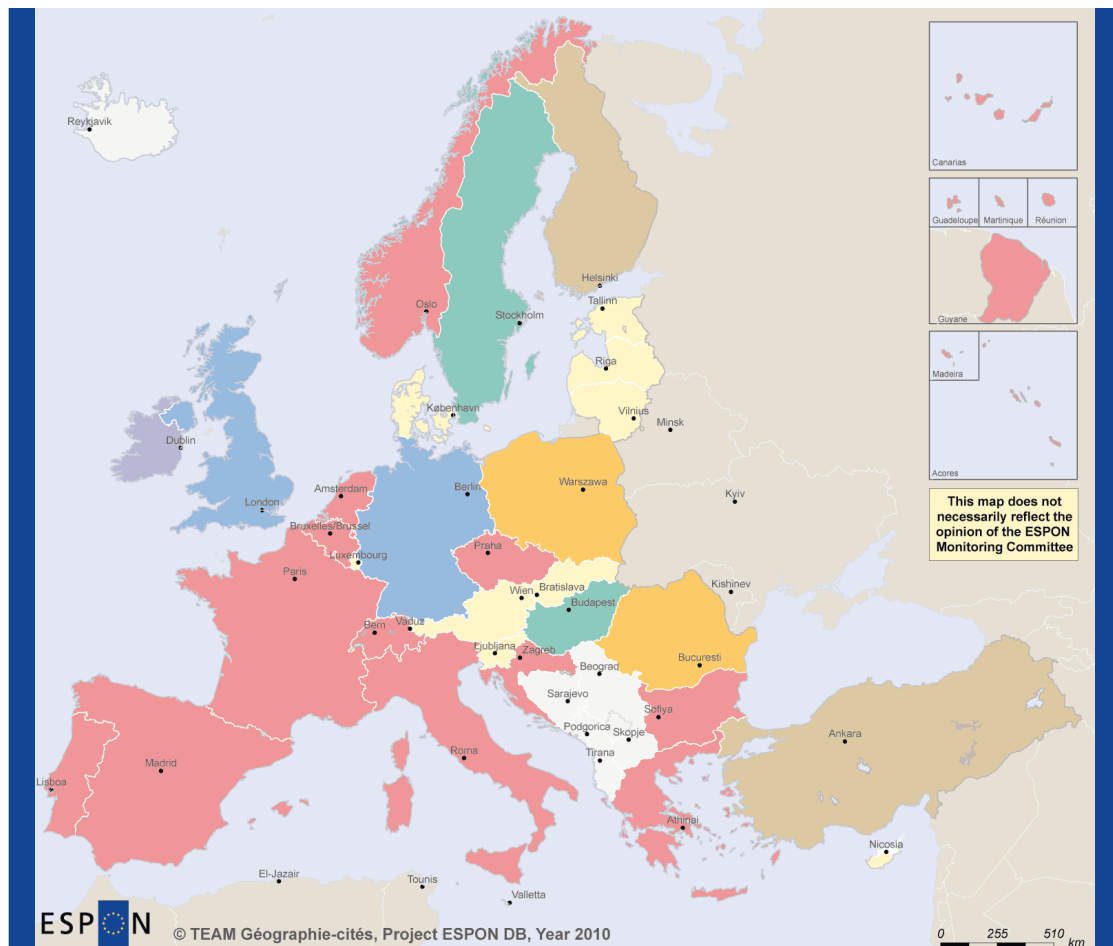
## 4 Results

After the construction of the 30 country-sheet specifications, different analyses have crossed the results in order to qualify in a synthetic way some national specificities in LUZ delineations.

### 4.1 Typology of LUZ delineations

A first synthesis of the 30 country-sheets is related to their type of definitions (Figure 5).

**Figure 5: Typology of LUZ delineation (UA III)**



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

National LUZ definitions  
based on:

One administrative unit

Aggregation of neighbouring units

Commuters mainly

Aggregation with no specification

Planning regions or local consultation

No generic rules

No precise information

No UA III

Out of ESPON space

Regional level: NUTS 0  
Source: ESPON DB, year 2010  
Origin of data: Urban Audit, year 2010  
© EuroGeographics Association for administrative boundaries

#### **4.1.1 LUZ as one elementary administrative unit**

A first approach characterizes countries that use a sole administrative unit as a LUZ, for example one LAU 1 (Cyprus or Estonia) or one NUTS 3 (Austria or Slovenia). Since these LUZ do not result from an aggregation of LAU, they are not considered as functional in our typology, even if most of the countries mention in their National Report some previous statistical analyses on commuters data that would justify the choice of such or such specific administrative level unit as a proxy (for example Estonia, or Slovakia).

#### **4.1.2 Aggregations of neighbouring units**

In two countries, Poland and Romania, LUZ consist in an aggregation of neighboring units, mainly based on distance or contiguity criteria. In Poland, these criteria are completed by hierarchical criteria (the extent of the ring depending on the City-Core size), and in Romania, it is completed by juridical criteria (the selected surrounding LAU 2 must be qualified as "urban" according to a former law, see the country-sheet in Annex).

#### **4.1.3 Aggregations mainly based on commuting data**

In about half of European countries, LUZ consist in functional aggregations mainly based on commuting data. However, we have to precise that construction methods are extremely different from one country to another: some take into account only commuters patterns towards a central pole (for example Greece or Croatia), whereas other also consider commuters patterns towards surrounding areas (for example Italy). Sometimes, school commuting are included (see Belgium), population growth (see Switzerland) or transport infrastructures (see Netherlands). In Portugal, the method is first based on LAU 2 commuter levels, then on LAU 1 commuter levels. And we will see in the next sub-section that, even if we examine only commuter thresholds, a great variety of situations appears. Nevertheless, this very rich set of functional approaches is worth studying, as it reflects results of national expertise on "what is a functional urban area" in such or such country.

#### **4.1.4 Aggregation with no specification**

In two countries, Sweden and Hungary, LUZ consist in an aggregation of administrative units but the criteria are not specified in the National Reports. In Hungary, for example, the LAU 1 is called a "statistical sub-region" and consists itself in an aggregation of elementary units based on functional criteria (see country-sheet in Annex).

#### **4.1.5 Planning regions or local consultations**

In Germany and United Kingdom, LUZ constructions are based on consultations at local or regional levels. LUZ correspond to "Planning regions" in Germany, whereas in United Kingdom "Office for National Statistics sought the recommendation of relevant Local Authorities and Government Office Regions when constructing the LUZ area for each of the 24 cities under analysis" (see country-sheets in Annex).

#### **4.1.6 No generic rules**

Ireland constitutes a particular case, in the sense that several approaches are used in this country (on a total of 5 LUZ, two of them are based on commuters and two others are based on planning regions) (see country-sheet in Annex).

#### **4.1.7 No precise information**

In Finland, the information given in the National Report or other sources collected by us is too incomplete about the way LUZ are constructed (see country-sheet in Annex).

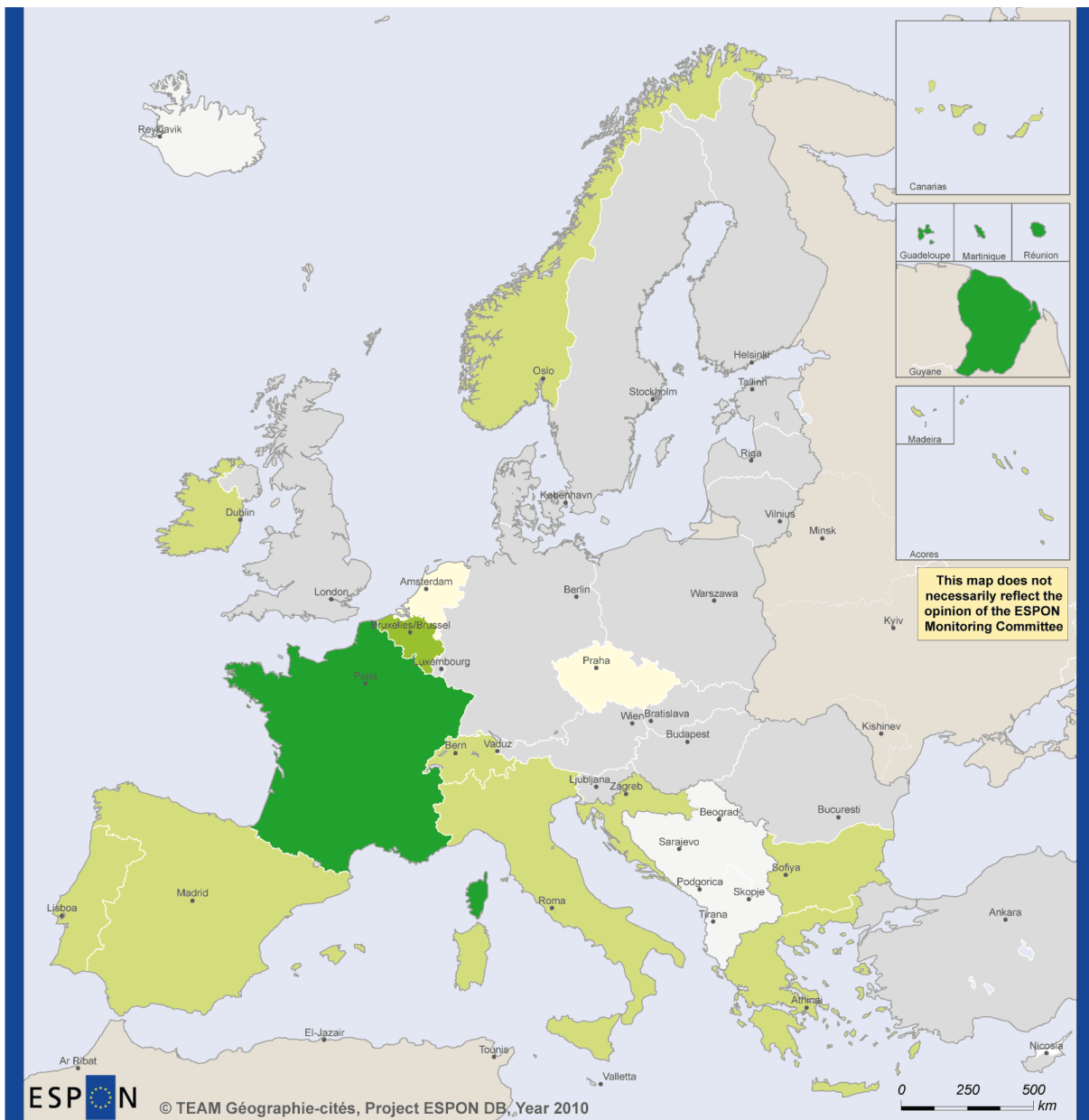
### **4.2 Diversity of commuting thresholds**

When the National Reports or other sources clearly mention the commuter thresholds used in functional definitions of LUZ, it is possible to display these data as values of a quantitative variable (Figure 6). The results seem to be very random and chaotic: countries with similar values are not located in the same part of Europe, and the map does not enlighten a general gradient or center-periphery structure or other macro-structure. Another possible explanation could be represented by the average size of administrative units (variation of thresholds corresponding to variation in LAU surfaces), but we did not find any statistical relations between these two variables. A more specific study should be addressed to the national researches that sustained the choice of these functional criteria in the different countries<sup>9</sup>.

---

<sup>9</sup> In France, for example, this context is given by Thomas Le Jeannic (1996), « Une nouvelle approche territoriale de la ville », INSEE – *Economie et Statistique* n°294-295.

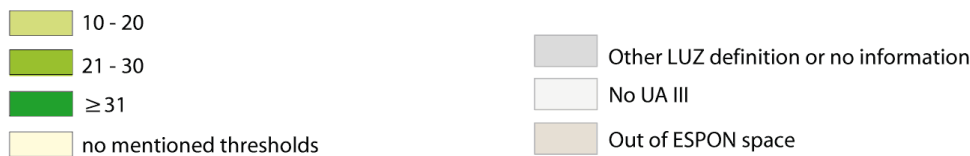
**Figure 6: A European variety of LUZ commuting thresholds**



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

Regional level: NUTS 0  
Source: ESPON DB, year 2010  
Origin of data: Urban Audit, year 2010  
© EuroGeographics Association for administrative boundaries

Level of commuters in LUZ functional approaches (%)



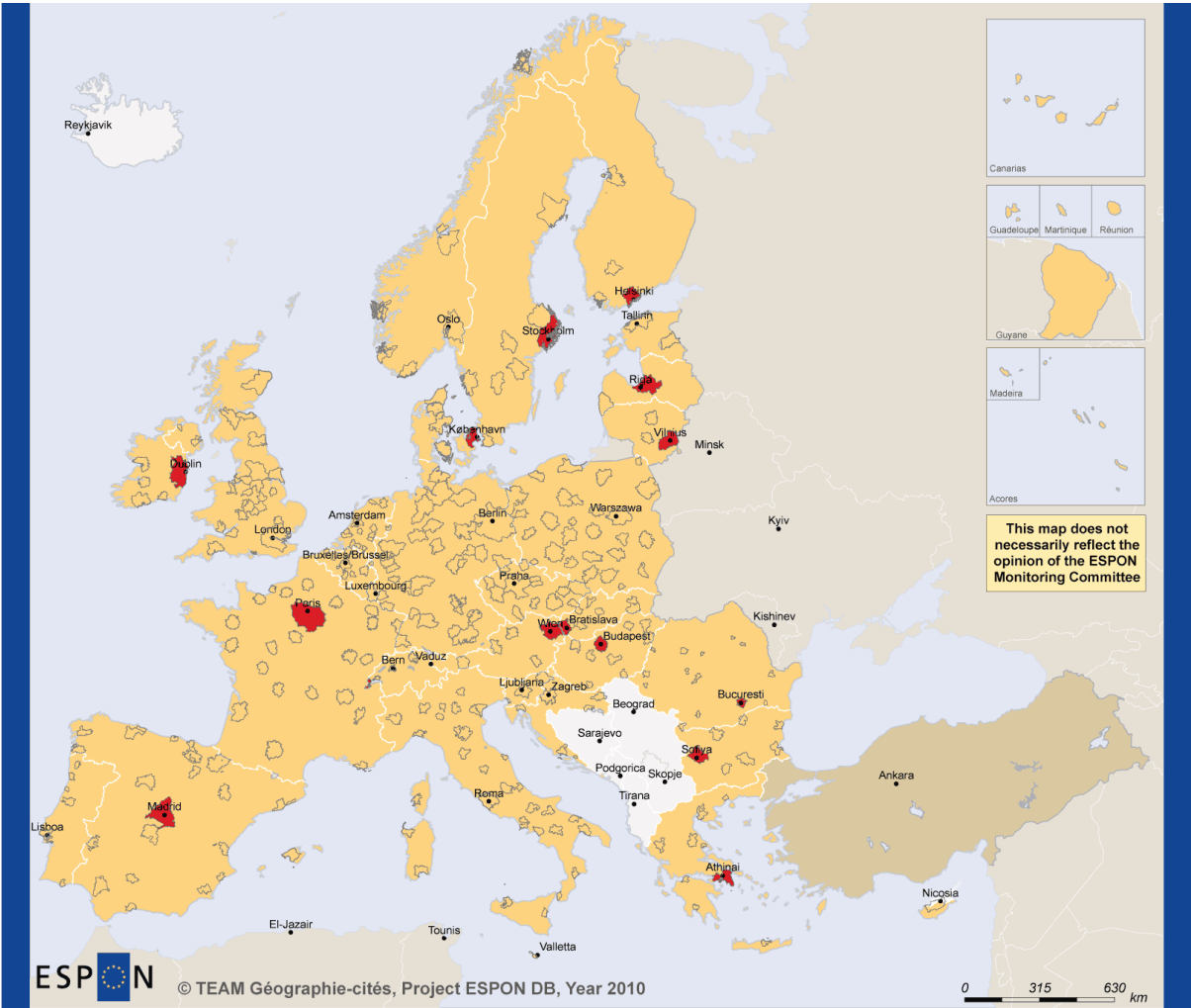
### 4.3 Capital cities as particular cases

In more than half of the European countries, the Capital cities constitute a particular case of the LUZ perimeters (Figure 7). The aim is generally to take into



account a particular influence field of this city (see for example Bucharest), or to fit better to a large administrative unit for questions of data availability (see for example the case of Paris). In most of the cases, NUTS perimeters are used : 4 Capital city LUZ fit with a NUTS 3, one fits with a NUTS 2, 5 are based on an aggregation of NUTS3, and 5 other are based on another type of aggregations (see country-sheets in Annex).

**Figure 7: Intra-national heterogeneity in UA III definitions**



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

Regional level: NUTS 0  
Source: ESPON DB, year 2010  
Origin of data: Urban Audit, year 2010  
© EuroGeographics Association for administrative boundaries

- 1 - Basic national delineation
- 2 - Particular zoning
- LUZ delineation
- Capital cities
- No UA III
- Out of ESPON space
- No precise information

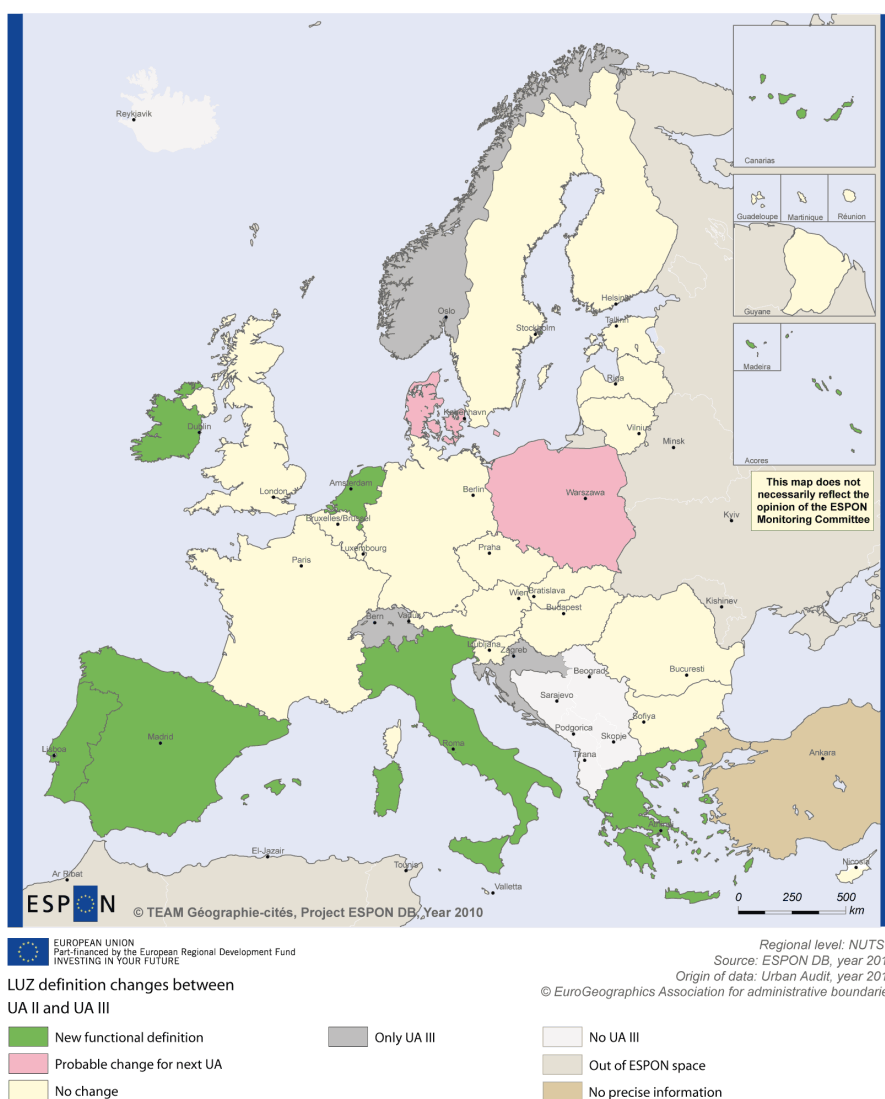
### 4.4 Towards more functional approaches

One of the fields that have been filled in the country-sheets concerns the evolution of LUZ definition between UA II and UA III. The results enlighten a

clear dynamic towards more functional approaches, largely encouraged by Urban Audit and Eurostat<sup>10</sup>.

Between UA II and UA III, six countries have changed to adopt functional definitions (Portugal, Spain, Greece, Italy, Netherlands, Ireland), and two countries have announced in their National Report or annex documentation that they would probably adopt new definition for the next Urban Audit (Poland and Denmark). It is also noteworthy that all the new participant members to UA III, which are not represented on this map because not concerned by evolution since UA II, i.e. Croatia, Norwegian and Switzerland, have all adopted functional definition (mainly based on commuters). We have no information for Turkey.

**Figure 8: New LUZ functional definitions**



<sup>10</sup> Torbiörn, Carlquist. « The Larger Urban Zones in the Urban Audit data collection. » *Globalisation impact on Regional and Urban Statistics*. Wroclaw, 2006.

## 5 Conclusion

As presented in this technical report, the work consisting in collecting and analyzing documentation on LUZ specifications was very complex but helped to re-write in a common way the different rules used by each country to define its LUZ. This work is just a step in the whole process, for three different reasons:

- It will have to be updated when the results of the next Urban Audit will be published
- Some explaining factors still have to be explored (for example, concerning the great variety of commuter thresholds in Europe)
- Some country profiles (country sheets) are not yet complete. Other documentation has to be found or to be translated in English.

Anyway, the different maps and the typology allowed to have a global overview and to enlighten different results. The large heterogeneity in the national approaches used to define LUZ engages researchers to be very cautious when interpreting some statistical results. But they also enlighten a very interesting evolution between UA 2001 and UA 2004, towards more functional definitions mainly based on commuters, even if the criteria (commuter thresholds, for instance) are very different from one country to another.

It confirms again that harmonization in definitions must not be only guided by the research of a unique rule and criteria for the whole Europe but must be based firstly on a good knowledge of the regional differences in settlement contexts and secondly on the political and historical ways each country defines cities. These differences are not an obstacle to harmonization when the metadata are fully specified.

## Annex : LUZ specifications by country (UA III)

### Model of the Country-Sheet

NAME OF THE COUNTRY
<b>Summary</b>
<ul style="list-style-type: none"> <li>◆ Number of cities with a LUZ in UA III and previous UA</li> <li>■ Changes in LUZ definitions between UA II and UA III</li> <li>● Description of LUZ delineation rules in UA III</li> </ul>
<b>Building blocks</b>
<ul style="list-style-type: none"> <li>◆ Identification of the building blocks of the LUZ</li> <li>■ Links between building blocks when LUZ is an aggregation</li> <li>● LUZ capital specificity (relatively to the other national LUZ)</li> </ul>
<b>Particular cases</b>
Particular observations, not concerning LUZ capital
<b>Correspondence with GISCO</b>
Coherence between National Reports and GISCO shape files (number of LUZ, names etc.)
<b>References</b>
Documentation used for building the sheet (National Reports and other sources)

## AUSTRIA

### Summary

◇ Austrian cities have been included in Urban Audit since the Pilot Phase (Wien, Graz) **(1; 2)**, then in UA II (Linz) **(2; 3)**, and in UA III (Salzburg, Innsbruck) **(4; 5)**. In total, five cities are concerned.

□ LUZ definitions have not changed between UA II and UA III (*"The spatial units of Vienna and Graz were the same as in the data collection 2001"* **(4)**).

○ Each LUZ corresponds to one NUTS 3 (except for Wien), not only in UA II (*"The proposal for the larger urban zones was to take NUTS 3- regions to get a functional urban region"* **(2)**) but also in UA III (*"For the new cities Salzburg (AT004C) and Innsbruck (AT005C) the Larger Urban Zone is the surrounding NUTS3-Region"* **(4)**). Apparently some commuting data have been used previously, but no details are provided in National Reports (*"For the definition of the larger urban zone of a city we used commuting data"* **(2)**).

### Building blocks

◇ NUTS 3, elementary administrative unit (i. e. *Gruppen von Politischen Bezirken*)

□ No aggregation

○ Wien: NUTS 3 aggregation (no details concerning the nature of the links)

### Particular cases

Linz: The LUZ *"is not the best solution, because in the NUTS3-region of Linz is another big city called Wels. But Linz and Wels are functionally related, so we decided to take also the NUTS3-region for Linz"* **(2)**.

### Correspondence with GISCO

Same number of LUZ<sup>11</sup>

[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)

### References

1. ERECO. (2000). *L'Audit Urbain, Vers un référentiel pour mesurer la qualité de la vie dans 58 villes européennes*. Luxembourg: Office des publications officielles des Communautés européennes.

2. Schrittwieser, Karin. (Undated). *Urban Audit II Final report for the European Commission*. Wien : Statistik Austria.

3. Eurostat. (2004). *Urban Audit, Methodological Handbook, 2004 edition*. Luxembourg: Office for Official Publications of the European Communities.

4. Schrittwieser, Karin. (Undated). *Urban Audit 2004 Austria Final Report for the European Commission*. Wien : Statistik Austria.

5. Eurostat. (2008). *European Regional and Urban Statistics Reference Guide*. Luxembourg : Office for Official Publications of the European Communities.

<sup>11</sup> Files downloaded and checked July 5, 2010

## BELGIUM

### Summary

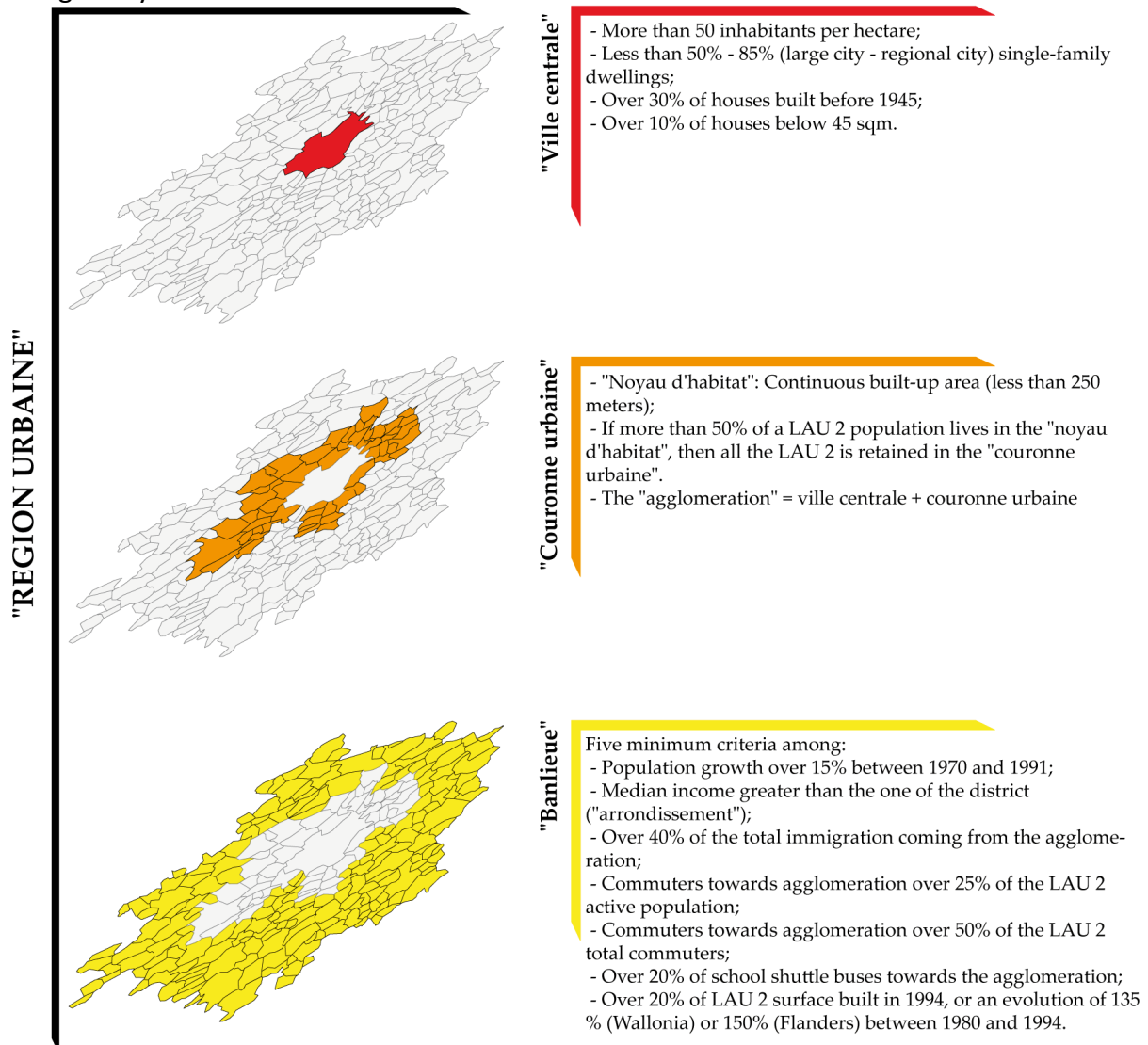
◇ Belgian cities have been included in Urban Audit since the Pilot Phase (Brussels, Anvers) **(1)**, then in UA II (Liège, Charleroi, Gand, Bruges) **(2; 3)**, and in UA III (Namur) **(4)**. In total, seven cities are concerned.

□ LUZ definitions have not changed between UA II and UA III **(5)**.

○ LUZ correspond to the Belgian functional region called « région urbaine » (« *Les « large urban zones » ont été établies à partir de : “Les régions urbaines belges en 1991”* » **(3)**). The criteria are functional ([elles] « *reposent sur des critères de fonctionnalité* » **(6)**). These criteria were formulated in 1996 **(7)**.

Construction data came from 1991 census **(8)**.

The figure below describes the different steps for building the “régions urbaines”, translated in English by us:



Source: Van der Haegen, H, Van Hecke, E et G., Juchtmans. *Les régions urbaines belges en 1991*. [éd.] Institut national de Statistique. *Etudes statistiques*. 1996, n°104

<b>Building blocks</b>
<p>◇ LAU 2, aggregation (i.e. Gemeenten / Communes): “ville centrale + couronne urbaine + banlieue”.</p> <p>□ Links: mainly based on commuters (threshold 25%).</p>
<b>Correspondence with GISCO</b>
<p>Same number of LUZ<sup>12</sup></p> <p><a href="http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1">http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1</a></p>
<b>References</b>
<ol style="list-style-type: none"> <li>1. ERECO. (2000). <i>L'Audit Urbain, Vers un référentiel pour mesurer la qualité de la vie dans 58 villes européennes</i>. Luxembourg: Office des publications officielles des Communautés européennes.</li> <li>2. Eurostat. (2004). <i>Urban Audit, Methodological Handbook, 2004 edition</i>. Luxembourg: Office for Official Publications of the European Communities.</li> <li>3. (Undated). <i>Deuxième rapport sur le projet Urban Audit II</i>.</li> <li>4. Eurostat. (2008). <i>European Regional and Urban Statistics Reference Guide</i>. Luxembourg : Office for Official Publications of the European Communities.</li> <li>5. (Undated). Rapport sur le projet Urban Audit II : Demande de données historiques 1991-1996 et données 2001.</li> <li>6. Doulou-Ouamba, Marlène. (2008). <i>Audit Urbain III</i>. s.l. : Service Publique Fédéral Economie, PME, Classes Moyennes et Energie.</li> <li>7. Van der Haegen, H, Van Hecke, E et G., Juchtman. (1996). <i>Les régions urbaines belges en 1991</i>. [éd.] Institut national de Statistique. <i>Etudes statistiques</i>. n°104.</li> <li>8. Hermia, J-P. (Undated). <i>Une nouvelle délimitation spatiale du phénomène périurbain bruxellois</i>.</li> </ol>

---

<sup>12</sup> Files downloaded and checked June24, 2010

## BULGARIA

### Summary

◇ Bulgarian cities have been included in UA II (Sofia, Plovdiv, Varna, Burgas, Pleven, Ruse, Vidin), then in UA III (Stara Zagora) **(1; 2; 3; 4)**. In total, eight cities are concerned.

□ LUZ definitions have not changed between UA II and UA III.

○ LUZ are defined with a functional approach. Two steps are followed:

1st step: *“Finding the settlements (LAU2 [i.e. Naseleni Mesta]) that have more than 20% commuters (out of the employed residents) in both directions to the City Core (incoming and outgoing migration are added)”*.

2nd step: *“Aggregating the data for the respective municipality (i. e. LAU 1) and checking again for the criterion” (4)*. Apparently, this choice is due to lack of data (*“Even though the NUTS5 data is the most suitable for defining the functional urban zones in the case of Bulgaria, due to data availability reasons a combination of NUTS4 units (Obshtini) was chosen” (5)*).

Data come from the 2001 Census (*“A particular question in the Census card addresses the location where the individual works/studies and lives” (5)*).

### Building blocks

◇ LAU 1, aggregation (i.e. *Obshtini*)

□ Links: based on commuting data (threshold 20%).

○ Sofia: Aggregation of one NUTS 3 (i.e. *Oblasti*) and several LAU 1 (*“The whole NUTS3 level unit (Oblast) was included in the newly formed LUZ, as well as several neighboring NUTS4 territorial units” (5)*).

### Correspondence with GISCO

Same number of LUZ<sup>13</sup>

### References

1. Eurostat. (2004). *Urban Audit, Methodological Handbook, 2004 edition*. Luxembourg: Office for Official Publications of the European Communities.
2. (Undated). *Urban Audit, Phase III - Historic data Interim Report 2006*.
3. Eurostat. (2009). *Annuaire régional d'Eurostat 2009*. Luxembourg : Office des publications officielles des Communautés européennes.
4. Tsvetarsky, Serguey. (Undated). *Phare 2005 - Project: Urban Audit IV*. Sofia.
5. Tsetarsky, Serguey et Kotzev, Alexander. (2004). *Eurostat Pilot Phase - Urban Audit II - Phase 1 - Final Report Bulgaria*. Sofia.

<sup>13</sup> Files provided by Urban Audit October 5, 2010



## CROATIA

### Summary

◇ Croatian cities have been included in UA III (Zagreb, Rijeka, Slavonski Brod, Osijek, Split) **(1)**. In total, five cities are concerned.

□ Croatia did not participate to UA II.

○ LUZ are defined with a functional approach, and the definition was specifically adopted for Urban Audit (*"Larger Urban Zones (LUZ) were created by CBS [Central Bureau of Statistics] only for UA needs on the basis of the nearest neighborhood"* **(2)**). Two steps are followed:

1st step: *"Larger urban zones around selected cities have been chosen in the following way: We have investigated for each city which municipalities and towns on the NUTS 5 level have commuting rate over 20% (according to the data from Census 2001)"*.

2nd step: *"we have created continuous larger urban zones (no holes or gaps)"* **(2)**.

The construction data come from the 2001 census **(2)**.

### Building blocks

◇ LAU 2, aggregation (i.e. *Gradovi i općine*)

□ Links: based on commuting data (threshold 20%)

### Particular cases

Split: *"Only one exception was Split LUZ in which municipality of Zagvozd was included even if it has no common border with the other components of the Larger urban zone"* **(2)**.

### Correspondence with GISCO

Same number of LUZ<sup>14</sup>

### References

1. Eurostat. (2008). *European Regional and Urban Statistics Reference Guide*. Luxembourg : Office for Official Publications of the European Communities.
2. Crostat. (2008). *Urban Audit Final Operational Report*. s.l. : Republic of Croatia Central Bureau of statistics.

---

<sup>14</sup> Files provided by Urban Audit, October 5, 2010

## CYPRUS

### Summary

- ◇ Only one city, in UA II and III **(1; 2)** (Lefkosia).
- LUZ definitions have not changed between UA II and UA III.
- Lefkosia LUZ corresponds to one LAU 1:  
*“The district of Lefkosia was taken as a proxy for the Larger Urban Zone (LUZ) based on data, from the 2001 Population Census, on commuting flows to the core city” (1): “in most municipalities and communities in the district of Lefkosia, other than those eight that are considered urban, reside a large percentage of commuters to the urban area of the district (for the purpose of employment): at least 40% (for the majority) when calculated individually” (2).*

### Building blocks

- ◇ LAU 1, elementary administrative unit (i.e. *Eparchie*)
- No aggregation

### Correspondence with GISCO

Same number of LUZ<sup>15</sup>

### References

1. (Undated). *Urban Audit 2001 Data Collection Project: Final Report (Reporting country: Cyprus)*.
2. (2008). *Urban Audit 2006/2007 Data Collection Exercise Final Report Member State: Cyprus*.

---

<sup>15</sup> Files provided by Urban Audit, October 5, 2010.

## CZECH REPUBLIC

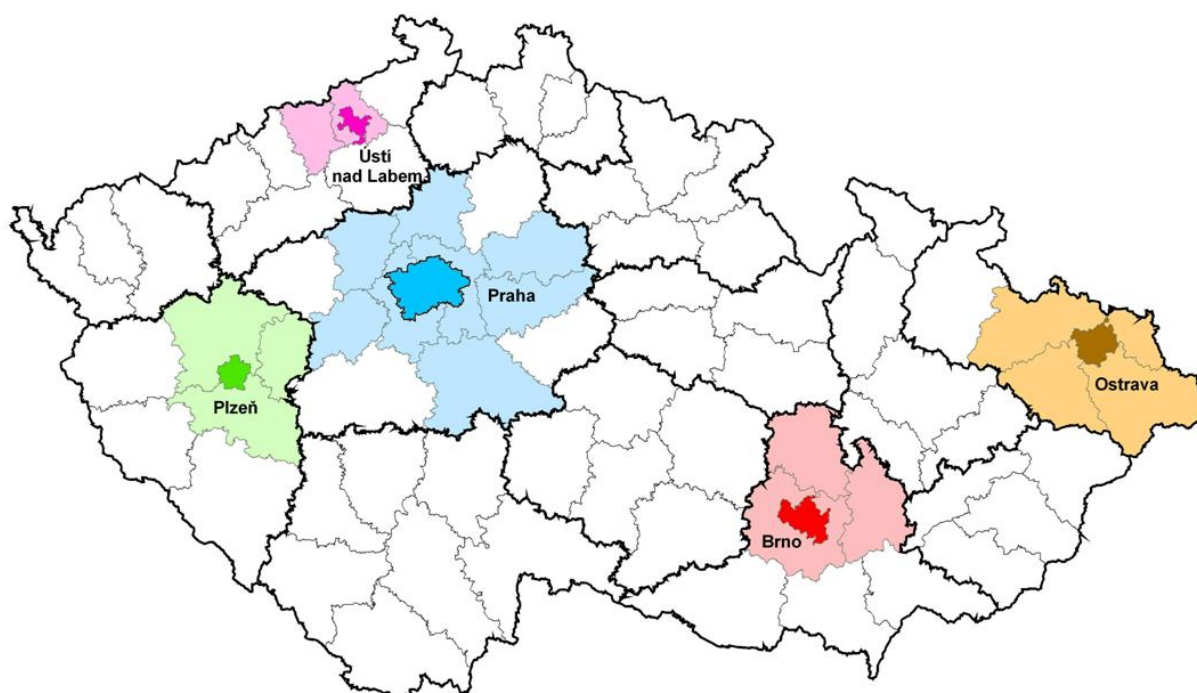
### Summary

◇ Czech cities have been included in UA II (Brno, Ostrava, Plzen, Praha, Usti nad Labem) **(1)**, then in UA III (Olomouc, Liberec, Ceske Budejovice, Hradec Kralove, Pardubice, Zlin, Karlovy Vary, Jihlava **(2)**). In total, thirteen cities are concerned.

□ Czech Republic did not transmit National Report for UA III.

○ LUZ are defined in UA II with a functional approach. They correspond to an aggregation of LAU 1 (*"Larger Urban Zones (LUZ) have been defined as groups of LAU 1 units, as the NUTS 3 regions were excessively large" (4)*). According to the UA II National Report, LUZ have been built using commuting data from Census 1991, but the methodology is not detailed (*"Delimitation of the adequate functional LUZs was based mainly on the commuting into work and schools from census 1991" (5)*). According to the Manual of Data Quality Aspects, the perimeters have been confirmed by Census 2001 (*"[LUZ] have been delimited as aggregates of LAU-1 units (okresy). (...) Inclusion of individual LAU 1 districts was (...) confirmed by Census 2001, once available" (3)*).

The map below gives the delimitation of the five LUZ of UA II. The grey lines correspond to LAU 1 units and the black lines to NUTS 3 **(3)**.



Source: Czech Statistical Office. (2005). *Urban Audit - Czech Republic - Manual of data quality aspects (Phase 1, 2 and historical data)*.

### Building blocks

◇ LAU 1, aggregation (i.e. *Okresy*)

□ Links: based on commuting data. Methodology has not been provided in National Report.

<b>Particular cases</b>
- Kladno: the City Core has no LUZ because it is part of Praha LUZ <sup>16</sup> .
<b>Correspondence with GISCO</b>
Same number of LUZ <sup>17</sup> <a href="http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1">http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1</a>
<b>References</b>
<ol style="list-style-type: none"> <li>1. Eurostat. (2004). <i>Urban Audit, Methodological Handbook, 2004 edition</i>. Luxembourg: Office for Official Publications of the European Communities.</li> <li>2. Eurostat. (2008). <i>European Regional and Urban Statistics Reference Guide</i>. Luxembourg : Office for Official Publications of the European Communities.</li> <li>3. Czech Statistical Office. (2005). <i>Urban Audit - Czech Republic - Manual of data quality aspects (phase 1, 2 and historical data)</i>.</li> <li>4. Czech Statistical Office. (2005). <i>Urban Audit - Final Report on Phase 2 - Czech Republic</i>. Praha.</li> <li>5. Czech Statistical Office. (2004). <i>Urban Audit II – Final Report on Phase 1</i>. Praha.</li> </ol>

---

<sup>16</sup> According to information transmitted by Urban Audit, January 2009.

<sup>17</sup> Files downloaded and checked July 21, 2010

## DENMARK

### Summary

◇ Danish cities have been included in Urban Audit since the Pilot Phase (Copenhagen) **(1)**, then in UA II (Århus, Odense, Aalborg) **(2; 3)**, and in UA III **(4; 5)**. In total, four cities are concerned.

□ LUZ definitions have not changed between UA II and UA III.

○ Each LUZ corresponds to one NUTS 3:

*"The level of the regions [i.e. Amter] in which the cities are placed, [called the larger urban zone], (...) is an administrative unit managed by a Council, and is elected every four years" (4).* These Amters have disappeared on December 31, 2006 but LUZ definitions remained the same for UA III: *"Denmark did not wish to change their LUZ yet. A big regional reform is foreseen for 2007 and in connection with this, new non-administrative NUTS 3 regions will have to replace the old "Amter" or counties" (6).*

Probable changes in LUZ definitions will occur for the next Urban Audit (UA IV).

### Building blocks

◇ NUTS 3, elementary administrative unit (i.e. *Amter*).

□ No aggregation

○ Copenhagen: aggregation of several LAU 1 (i.e. *Kommuner*) and several NUTS 3:

*"The capital region has been officially delineated to include the municipality and three counties (Amter):" (2)*, i.e, Copenhagen LUZ includes *"(...) the municipality of Copenhagen and Frederiksberg and the regions of Copenhagen, Frederiksberg and Roskilde" (4).*

### Correspondence with GISCO

Same number of LUZ<sup>18</sup>

[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)

### References

1. ERECO. (2000). *L'Audit Urbain, Vers un référentiel pour mesurer la qualité de la vie dans 58 villes européennes*. Luxembourg: Office des publications officielles des Communautés européennes.
2. Danmarks Statistik, Befolkning og Uddannelse. (2004). *Final report concerning agreement N° 2002CE160AT181 about Urban Audit*.
3. Eurostat. (2004). *Urban Audit, Methodological Handbook, 2004 edition*. Luxembourg: Office for Official Publications of the European Communities.
4. Statistics Denmark, Population. (2007). *Final Report Concerning Agreement N° About Urban Audit 2004 Collection*.
5. Eurostat. (2008). *European Regional and Urban Statistics Reference Guide*. Luxembourg : Office for Official Publications of the European Communities.
6. Carlquist, Torbiörn. (2006, August 30). The Larger Urban Zones in the Urban Audit data

<sup>18</sup>Files downloaded and checked July 5, 2010

collection. *Globalisation Impact on Regional and Urban Statistics*. Wrocław.

7. Christiansen, Henning. (Undated). *Urban Audit II - State of the Art - Country: Denmark*.

## ESTONIA

### Summary

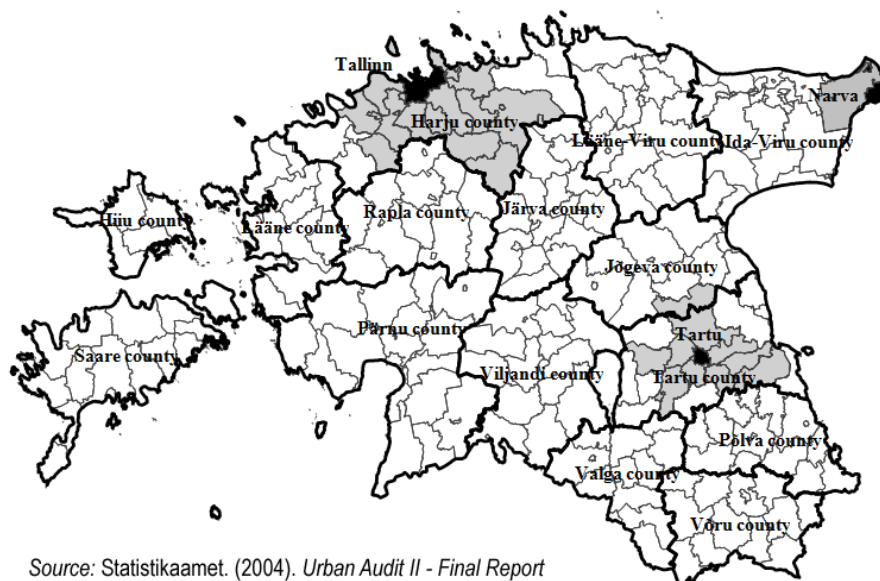
◇ Estonian cities have been included in Urban Audit since the second one (Tallinn et Tartu **(1)**), and in UA III **(2)**. In total, two cities are concerned.

□ LUZ definitions have not changed between UA II and UA III.

○ Each LUZ corresponds to one LAU 1:

*"In 2002 the working group headed by Professor Jussi S. Jauhiainen of Finland analyzed upon request of the Ministry of Internal Affairs the development potential of Urban Regions of Estonia " **(3)** . However, it seems that these Urban Regions were not used as LUZ and that LAU 1 were chosen instead. Indeed, a proposal was presented to Eurostat "to define Harju [the county that contains Tallinn] and Tartu counties as LUZs for the above-mentioned cities. The LUZ of Tartu does not exactly overlap with the Urban Region of Tartu specified by Professor Jussi S. Jauhiainen. However, the differences are not very big and considering the administrative concept, both the cities Tallinn and Tartu definitely influence the counties surrounding them" **(3)**.*

The map below gives Urban Regions (grey area) and LAU 1 (bold black line) **(4)**.



Source: Statistikaamet. (2004). *Urban Audit II - Final Report*

### Building blocks

◇ LAU1, elementary administrative unit (i.e. *Maakond*)

□ No aggregation

### Correspondence with GISCO

Same number of LUZ<sup>19</sup>

[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)

<sup>19</sup> Files downloaded and checked July 21, 2010.

## References

1. Eurostat. (2004). *Urban Audit, Methodological Handbook, 2004 edition*. Luxembourg: Office for Official Publications of the European Communities.
2. Eurostat. (2008). *European Regional and Urban Statistics Reference Guide*. Luxembourg : Office for Official Publications of the European Communities.
3. Statistikaamet. (2004). *Urban Audit II - Final Report*.
4. Statistikaamet. (Undated). *The Report on quality aspects and meta information*.



## FINLAND

### Summary

- ◇ Finnish cities have been included in Urban Audit since the Pilot Phase (Helsinki) **(1)**, then in UA II (Tampere, Turku, Oulu) **(2)**, and in UA III **(3)**. In total, four cities are concerned.
- LUZ definitions have not changed between UA II and UA III.
- LUZ correspond to an aggregation of LAU 2 and the list of building blocks is given in the UA III National Report **(4)**. However, the rules of aggregation are not described (*"The suitable area divisions for (...) LUZ (...) were defined together with Eurostat"* **(5)**).

### Building blocks

- ◇ LAU 2, aggregation (i.e. *Kunnat / Kommuner*)
- Links: No information
- Helsinki: The LUZ corresponds to *"the functional urban region called the Helsinki Region. This functional region is already used in the NORDSTAT database (Nordic Cities and Regions) and in the Finnish City Indicators database"*<sup>20</sup>

### Correspondence with GISCO

Same number of LUZ<sup>21</sup>  
[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)

### References

1. ERECO. (2000). *L'Audit Urbain, Vers un référentiel pour mesurer la qualité de la vie dans 58 villes européennes*. Luxembourg: Office des publications officielles des Communautés européennes.
2. Eurostat. (2004). *Urban Audit, Methodological Handbook, 2004 edition*. Luxembourg: Office for Official Publications of the European Communities.
3. Eurostat. (2008). *European Regional and Urban Statistics Reference Guide*. Luxembourg : Office for Official Publications of the European Communities.
4. Statistics Finland. (2007). *Eurostat - Urban Audit III - Final Country Report Finland*.
5. Statistics Finland. (2004). *Final Report: Urban Audit II Finland*.

---

<sup>20</sup> According to information provided by Urban Audit December 3, 2010.

<sup>21</sup> Files downloaded and checked June 24, 2010

## FRANCE

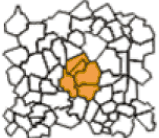
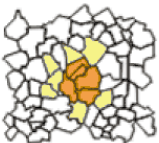
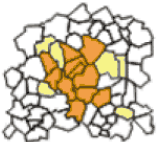

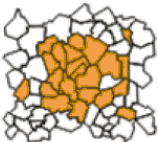

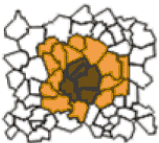
### Summary

◇ French cities have been included in Urban Audit since the Pilot Phase (Lyon, Toulouse, Strasbourg, Bordeaux, Nantes) **(1)**, then in UA II (Ajaccio, Amiens, Besancon, Caen, Clermont-Ferrand, Dijon, Grenoble, Le Havre, Limoges, Metz, Nancy, Orleans, Paris, Poitiers, Reims, Rennes, Rouen) **(2)**, and in UA III (Tours) **(3)**. In total, thirty-three cities are concerned.

□ LUZ definitions have not changed between UA II and UA III.

○ LUZ are defined with a functional approach. They correspond to the French definition of «*aire urbaine*» (“*Ce sont donc les aires urbaines qui ont été utilisées pour représenter les LUZ de l’audit*” **(4)**). Criteria were formulated in 1996 and the construction data come from the 1999 Census **(7)**.

The figure below describes the different steps for building the “aires urbaines”, translated in English by us:

1 - Determination of the urban pole (an "unité urbaine" concentrating more than 5000 jobs).	
2 - Rural LAU 2 and "unités urbaines" characterized by at least 40% of the active population working in the urban pole are added to the urban pole.	 <div style="display: flex; justify-content: flex-end; margin-top: 5px;"> <div style="width: 15px; height: 15px; background-color: orange; margin-right: 5px;"></div> Previous zone  <div style="width: 15px; height: 15px; background-color: yellow; margin-right: 5px; margin-left: 10px;"></div> Added LAU 2         </div>
3 - Rural LAU 2 and "unités urbaines" characterized by at least 40% of the active population working in this previous zone (resulting of the last step) are added.	 <div style="display: flex; justify-content: flex-end; margin-top: 5px;"> <div style="width: 15px; height: 15px; background-color: orange; margin-right: 5px;"></div> Previous zone  <div style="width: 15px; height: 15px; background-color: yellow; margin-right: 5px; margin-left: 10px;"></div> Added LAU 2         </div>
4 - Rural LAU 2 and "unités urbaines" characterized by at least 40% of the active population working in this previous zone (resulting of the last step) are added. The process is iterative.	 <div style="display: flex; justify-content: flex-end; margin-top: 5px;"> <div style="width: 15px; height: 15px; background-color: orange; margin-right: 5px;"></div> Previous zone  <div style="width: 15px; height: 15px; background-color: yellow; margin-right: 5px; margin-left: 10px;"></div> Added LAU 2         </div>
5 - No more rural LAU 2 and "unités urbaines" located outside the previous zone is characterized by at least 40% of the active population working in this previous zone. The iterative process ends.	 <div style="display: flex; justify-content: flex-end; margin-top: 5px;"> <div style="width: 15px; height: 15px; background-color: orange; margin-right: 5px;"></div> Previous zone  <div style="width: 15px; height: 15px; background-color: yellow; margin-right: 5px; margin-left: 10px;"></div> Added LAU 2         </div>
6 - The LAU 2 that are not contiguous to the area containing the urban pole are removed. The LAU 2 which are enclaved are added.	
7 - The "Aire urbaine" is constituted. Inside this area are distinguished LAU 2 of the urban pole and the others selected LAU 2 which compose the outer ring.	 <div style="display: flex; justify-content: flex-end; margin-top: 5px;"> <div style="width: 15px; height: 15px; background-color: #8B4513; margin-right: 5px;"></div> <b>Aire urbaine</b>  <span style="margin-left: 10px;">Urban pole</span>  <div style="width: 15px; height: 15px; background-color: #FF8C00; margin-right: 5px; margin-left: 10px;"></div> Outer ring         </div>

Source: [http://www.insee.fr/fr/regions/auvergne/default.asp?page=themes/donnees\\_detailles/aireurbaine/aire-construc.htm](http://www.insee.fr/fr/regions/auvergne/default.asp?page=themes/donnees_detailles/aireurbaine/aire-construc.htm)

<b>Building blocks</b>
<p>◇ LAU 2, aggregation (i.e. <i>Commune</i>)</p> <p>□ Links: based on commuting data (threshold 40%)</p> <p>○ Paris: NUTS 2 (i.e. <i>Régions</i>) (“<i>Une simplification sur Paris : la LUZ est la région Ile de France. L’aire urbaine n’est guère différente en étendue et les objectifs de l’audit pour la capitale étaient de toutes manières particuliers</i>” (4)).</p>
<b>Particular cases</b>
<p>Pointe-à-Pitre, Fort de France and Cayenne: these cities are not delineated as « <i>aire urbaine</i> », so that they don’t correspond to a Larger Urban Zone. This is the case for all DOM-TOM cities, as “aires urbaines” have been only defined for Metropolitan France (“<i>Pas de LUZ, là où le concept d’ « aire urbaine » n’avait pas été mis en œuvre. Au moment de la collecte, c’était le cas des départements d’outre-mer dont l’étendue territoriale des communes fait qu’il est difficile de reproduire le concept en place sur la métropole.</i>”(4));</p> <p>Saint-Etienne, Marseille, Nice: these cities don’t correspond to a LUZ as their City Core is too different from the urban pole of “aire urbaine” called “unité urbaine”. Indeed, City Core in France is defined as an EPCI<sup>22</sup> (“<i>Pas de LUZ, là où les EPCI différaient trop des agglomérations morphologiques. En effet le concept d’aire urbaine est conçu comme la zone d’influence d’un noyau qui est pris a priori comme l’unité urbaine, c’est à dire l’agglomération morphologique. (...) Saint-Etienne, dont l’EPCI est plus étendue que l’agglomération et absorbe même une seconde agglomération ; Marseille, et Nice dont les agglomérations sont chacune partagées sur trois EPCI, dont le rayonnement est donc nettement inférieur à l’aire urbaine</i> »”(4)).</p> <p>Lens: this city does not correspond to a LUZ, due to deviations between City Core and « unité urbaine » (“<i>Douai et Lens ne forment en fait qu’une seule agglomération, mais dont le contour est significativement différent du seul regroupement des deux EPCI</i>”(4))</p> <p>Aix-en-Provence, Lille, Montpellier and Toulon: these cities does not correspond to a LUZ (“<i>The cities of (...) Aix-en-Provence (...) do not have any LUZ</i>” (6)); For Lille and Montpellier, “LUZ are removed in 2007”, and for Toulon, “LUZ is removed in 2008”<sup>23</sup>.</p>
<b>Correspondence with GISCO</b>
Different number of LUZ: Toulon’s LUZ is not removed in GISCO <sup>24</sup>
<b>References</b>
<ol style="list-style-type: none"> <li>1. ERECO. (2000). <i>L’Audit Urbain, Vers un référentiel pour mesurer la qualité de la vie dans 58 villes européennes</i>. Luxembourg: Office des publications officielles des Communautés européennes.</li> <li>2. Eurostat. (2004). <i>Urban Audit, Methodological Handbook, 2004 edition</i>. Luxembourg: Office for Official Publications of the European Communities.</li> <li>3. Eurostat. (2008). <i>European Regional and Urban Statistics Reference Guide</i>. Luxembourg : Office for Official Publications of the European Communities.</li> </ol>

<sup>22</sup> Etablissement Public de Coopération Intercommunale: Political aggregation of LAU 2, especially for elaborating planning projects.

<sup>23</sup> According to information provided by Urban Audit, January 2009.

<sup>24</sup> Files provided by Urban Audit, October 5, 2010.

4. INSEE. (Undated). *Audit urbain 1999-2003, Bilan de collecte*.
5. INSEE. (2010). *Insee - Définitions et méthodes - Aire Urbaine*. Checked June 25, 2010, on Institut National de la Statistique et des Etudes Economiques : <http://www.insee.fr/fr/methodes/default.asp?page=definitions/aire-urbaine.htm>.
6. INSEE. (Undated). *Audit Urbain 2006, Bilan de collecte*.
7. *Mesurer un univers urbain en expansion*. Julien, Philippe. (2000). *Economie et Statistique* n°336, pp. 3-33.

## GERMANY

### Summary

◇ German cities have been included in Urban Audit since the Pilot Phase (Berlin, Hamburg, Munich, Cologne, Frankfurt am Main, Essen, Stuttgart, Leipzig, Dresden) **(1)**, then in UA II (Augsburg, Bielefeld, Bochum, Bonn, Bremen, Darmstadt, Dortmund, Düsseldorf, Erfurt, Frankfurt (Oder), Freiburg-im-Breisgau, Gottingen, Halle-an-der-Saale, Hannover, Karlsruhe, Magdeburg, Mainz, Moers, Monchengladbach, Mulheim-an-der Ruhr, Nurnberg, Regensburg, Schwerin, Trier, Weimar, Wiesbaden, Wuppertal) **(2)**, and in UA III (Kiel, Saarbrucken, Koblenz) **(3; 4)**. In total, forty cities are concerned.

□ LUZ definitions have not changed between UA II and UA III (*“Delineation of the territorial units of the cities of the previous rounds of data collection remained unchanged, i. e. of the 35 cities, the 28 Larger Urban Zones”* **(3)**).

○ Each LUZ corresponds to a « planning region », which is defined through a variety of criteria<sup>25</sup>**(6)**. According to Klaus Trutzel (National Urban Audit Coordinator), *“at this level, the statistical offices of the Länder can provide a particularly wide variety of planning policy data, not all of which are available for individual Kreise (NUTS 3)”* **(5)**. It is also noticed that *“the size of the LUZ is about adequate although the labour market area of the city does not exactly match the LUZ. It might be worth comparing the real commuting areas, delimited by LAU2 units, with the LUZ for better judgments of the adequacy of the LUZ and the information collected for them”***(3)**.

In 2002/2003, Eurostat checked the correlation between LUZ and functional area (defined with a threshold of 15% and 20%). The results show that LUZ fit quite well with functional areas **(7)**, except for Bielefeld and Schwerin<sup>26</sup>. However, according to Klaus Trutzel, these latter remain unchanged *“for comparability needs”* **(3)**.

### Building blocks

◇ NUTS 3, aggregation (i.e. *Kreis*)

□ Links: variety of criteria used to define “planning regions”, not specified in National Reports.

### Particular cases

- Essen, Dortmund, Bochum, Mulheim and Moers: One common LUZ, Ruhrgebiet, has been defined for these five cities in the Ruhr area (*“The common LUZ for 5 Urban Audit cities in the Ruhr area was kept”* **(3)**);

- Monchengladbach, Wuppertal and Frankfurt (Oder): these cities do not correspond to a LUZ (*“The 3 cities of Monchengladbach, Wuppertal and Frankfurt (Oder) have no LUZ around*

<sup>25</sup> The “planning regions” are defined at the federal level by the BBR (Bundesamt für Bauwesen und Raumordnung). They are transmitted to the Länder which adapt them taking into account the large-scale projects proposed by local authorities. The number of planning regions is officially 97. However, some of them are subdivided by the Länder into several sub-regions. That’s why the number of 115 planning regions can be found sometimes **(6)**.

<sup>26</sup> The comparison was made between the population of the LUZ and the population of the estimated functional areas.

them" (3));

- Potsdam: this city does not correspond to a LUZ because it is part of the LUZ of Berlin<sup>27</sup>.
- Stuttgart: this city was removed in UA II but added again in UA III (5)

### Correspondence with GISCO

Same number of LUZ<sup>28</sup>

[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)

### References

1. ERECO. (2000). *L'Audit Urbain, Vers un référentiel pour mesurer la qualité de la vie dans 58 villes européennes*. Luxembourg: Office des publications officielles des Communautés européennes.
2. Eurostat. (2004). *Urban Audit, Methodological Handbook, 2004 edition*. Luxembourg: Office for Official Publications of the European Communities.
3. Trutzel, Klaus. (2008). *Urban Audit III (2006): Final Country Report Germany*.
4. Eurostat. (2008). *European Regional and Urban Statistics Reference Guide*. Luxembourg : Office for Official Publications of the European Communities.
5. Trutzel, Klaus. (Undated). *URBAN AUDIT - State of the art*.
6. Queva, Christophe. (2007). « Les paradoxes de la Région en Allemagne, entre réseaux et territoires : la région, outil de déterritorialisation ? » *Annales de Géographie* (653).
7. Carlquist, Torbiörn. (2006, August 30). The Larger Urban Zones in the Urban Audit data collection. *Globalisation Impact on Regional and Urban Statistics*. Wroclaw.

---

<sup>27</sup> According to information transmitted by Urban Audit, January 2009.

<sup>28</sup> Files downloaded and checked July 21, 2010

## GREECE

### Summary

- ◇ Greek cities have been included in Urban Audit since the Pilot Phase (Athina, Thessaloniki et Patras) **(1)**, then in UA II and UA III (Volos, Iraklion, Kavala, Kalamata, Ioannina et Larisa) **(2; 3)**. In total, nine cities are concerned.
- LUZ definitions have changed between UA II and UA III. For UA II, NUTS 3 were used as proxy ("*NSSG [National Statistical Service of Greece] decided to use this NUTS3 level as a proxy for the large urban zone*" **(5)**).
- In UA III, a new definition is based on labour market areas ("*There was a new delimitation of the existing LUZ areas of the 9 cities according to the Labour Market Areas and the suggestion of Eurostat*" **(4)**). LUZ are defined with a functional approach, but the methodology is not specified in the National Report. According to documentation sent by Urban Audit<sup>29</sup>, commuting data have been used at LAU 1 level, with a threshold of 15% people commuting from a "suburb" to the central city (demos). Construction data come from 2001 Census.

### Building blocks

- ◇ LAU 1, aggregation (i.e. *Demoi/Koinotites*)
- Links: apparently based on commuting data (threshold 15%).
- Athina: the LUZ corresponds to one NUTS 3, elementary administrative unit (i.e. *Nomoi*). Some outlying islands have been excluded.

### Correspondence with GISCO

Same number of LUZ<sup>30</sup>  
[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)

### References

1. ERECO. (2000). *L'Audit Urbain, Vers un référentiel pour mesurer la qualité de la vie dans 58 villes européennes*. Luxembourg: Office des publications officielles des Communautés européennes.
2. Eurostat. (2008). *European Regional and Urban Statistics Reference Guide*. Luxembourg : Office for Official Publications of the European Communities.
3. Eurostat. (2004). *Urban Audit, Methodological Handbook, 2004 edition*. Luxembourg: Office for Official Publications of the European Communities.
4. General Secretariat of the National Statistical Service of Greece. (2008). *Urban Audit 2006/2007 Data Collection - Final Operation Report*.
5. (Undated). *Urban Statistics - Final Report*.

<sup>29</sup> According to information transmitted by Urban Audit, January 2009.

<sup>30</sup> Files downloaded and checked July 5, 2010

## HUNGARY

### Summary

◇ Hungarian cities have been included in UA II (Budapest, Miskolc, Pecs, Nyiregyhaza **(1)**) and in UA III (Debrecen, Szeged, Győr, Kecskemét, Székesfehérvár**(2)**). In total, nine cities are concerned.

□ LUZ definitions have not changed between UA II and UA III **(3)**.

○ LUZ correspond to LAU 1 aggregation, and the methodology is not fully specified:

*“LUZs consist of the NUTS 4 level units, i. e. statistical subregions of these cities. A statistical subregion is primarily a functional unit, established on the basis of actual working, residential, transport and secondary provisional (education, health care, and trade) connections between the central city and the urban zone around” **(4)**.*

### Building blocks

◇ LAU 1, aggregation (i.e. *Statisztikaikistérségek*)

□ Links: the methodology is not fully specified (see above)

○ Budapest: LAU 2, aggregation (i.e. *Települések*). The methodology is partly specified:

*“In the case of Budapest, the Larger Urban Zone is made up of the 79 settlements of the legally defined Budapest agglomeration, where the 78 settlements surrounding the capital are tightly connected with the centre. A part of the settlements in the urban zone show the morphological aspects too of an agglomeration. In 25 settlements, more than 50% of the resident population live near – the incidental distance is not more than 200 meters in the corresponding categories – urban (built-in) areas of the capital, although a morphological connection to a lesser extent can be observed in other settlements as well. However, there is a functional relation in the case of all settlements, that is why the former agglomeration, which had consisted of 44 settlements, was extended. The Budapest agglomeration with today’s boundaries and the list of the settlements included were most recently published in the Government Decree Nr. 89/1997. (V.28.). A specific feature of the past decade in Hungary was that suburbanisation started in the proximity of Budapest too. The migration loss of the capital was over 100 000 persons, and a significant proportion of those who moved out of Budapest went to live in settlements of the agglomeration” **(4)**.*

### Correspondence with GISCO

Same number of LUZ<sup>31</sup>

[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)

### References

1. Eurostat. (2004). *Urban Audit, Methodological Handbook, 2004 edition*. Luxembourg: Office for Official Publications of the European Communities.
2. Eurostat. (2008). *European Regional and Urban Statistics Reference Guide*. Luxembourg : Office for Official Publications of the European Communities.

<sup>31</sup> Files downloaded and checked July 19,2010



3. Hungarian Central Statistical Office. (Undated). *Urban Audit III - Final Report*.
4. Hungarian Central Statistical Office. (2004). *Urban Audit II - Final Report*.

## IRELAND

### Summary

- ◇ Irish cities have been included in Urban Audit since the Pilot Phase (Dublin, Cork **(1)**), then in UA II (Galway, Limerick **(2)**), and in UA III (Waterford **(3)**). In total, five cities are concerned.
- LUZ delineations have changed between UA II and UA III. Ireland did not transmit National Report for UA III, but some information was provided by Urban Audit.  
For UA II, *“the definition of the larger urban zones was based on local area development plans or in consultation with local authority planners. This applied to the LUZ zones used in the cities of Limerick and Cork”* **(4)**.
- In UA III, LUZ are defined through 3 different ways:  
Two LUZ are based on functional definitions: *“the LUZs for Limerick and Waterford cities are based on commuting data from the most recent census and a threshold of 20% commuters from the surrounding areas to the central city has been applied”* <sup>32</sup>. They correspond to an aggregation of LAU 2, probably built with the same commuters patterns than in UA II (threshold 20%).  
LUZ of Cork and Galway correspond to planning regions (*Cork Area Strategic Plan (CASP)*<sup>33</sup> and *Galway Transport and Planning Study*<sup>34</sup>).  
Dublin<sup>35</sup> corresponds to an aggregation of NUT 3.

### Building blocks

- ◇ LAU 2, aggregation (i.e. *Electoral districts*)
- Links: diversity of situations (see above). The methodology is not fully specified.
- Dublin: Aggregation of 2 NUTS 3 (i.e. *Regional Authority Regions*).

### Correspondence with GISCO

Same number of LUZ<sup>36</sup>  
[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)

### References

1. ERECO. (2000). *L'Audit Urbain, Vers un référentiel pour mesurer la qualité de la vie dans 58 villes européennes*. Luxembourg: Office des publications officielles des Communautés européennes.
2. Eurostat. (2004). *Urban Audit, Methodological Handbook, 2004 edition*. Luxembourg: Office for Official Publications of the European Communities.

<sup>32</sup> Information transmitted in October 14, 2010

<sup>33</sup> <http://www.corkcity.ie/casp/strategicplan/>

<sup>34</sup> <http://www.galwaycity.ie/AllServices/RoadsandTraffic/StudiesandSchemes/GalwayTransportandPlanningStudy>

<sup>35</sup> Information transmitted in October 14, 2010

<sup>36</sup> Files provided by Urban Audit October 5, 2010.

3. Eurostat. (2008). *European Regional and Urban Statistics Reference Guide*. Luxembourg : Office for Official Publications of the European Communities.
4. (2004). *Urban Audit - Irish National Report 2001*.

## ITALY

### Summary

◇ Italian cities have been included in Urban Audit since the Pilot Phase (Roma, Milan, Naples, Turin, Palermo, Genoa, Florence, Bari) **(1)** then in UA II (Bologna, Catania, Venice, Verona, Cremona, Trento, Trieste, Perugia, Ancona, L'Aquila, Pescara, Campobasso, Caserta, Taranto, Potenza, Catanzaro, Reggio di Calabria, Sassari, Cagliari) **(2)**, and in UA III (Padova, Brescia, Modena, Foggia, Salerno) **(3)**. In total, thirty-two cities are concerned.

□ LUZ definitions have changed between UA II and UA III.

During UA II, each LUZ corresponds to a NUTS 3: *“Provinces are administrative areas in some case without a credible geographical or statistical significance. There is no a credible alternative to consider Province as proxy of LUZ in Urban Audit II, at least at this stage”* **(4)**.

○ During UA III, LUZ definition is functional and corresponds to Local Labour System (LLS, i.e. *systemi locali del lavoro*). The methodology aggregates LAU 2 on the basis of employment (1000 jobs minimum in the LLS) and residence (*“occupied people working in A, occupied people resident in A, and occupied people resident and working in A”* **(5)**). Commuters are taken into account with a threshold of 10% (outflow) and 1% (inflow). The methodology uses a *“self-containment criteria”* with a threshold of 75% for building the LLS. This criteria is defined by *“occupied people resident and working in A/ occupied people resident in A”* **(5)**. Construction data come from *“the 1991 census intra-municipality daily commuting flows matrix”* **(4)**.

### Building blocks

◇ LAU 2, aggregation (i.e. *Comuni*)

□ Links: based on commuting data (threshold (10%). The methodology is provided in the National Report **(5)**.

### Correspondence with GISCO

Same number of LUZ<sup>37</sup>

[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)

### References

1. ERECO. (2000). *L'Audit Urbain, Vers un référentiel pour mesurer la qualité de la vie dans 58 villes européennes*. Luxembourg: Office des publications officielles des Communautés européennes.
2. Eurostat. (2004). *Urban Audit, Methodological Handbook, 2004 edition*. Luxembourg: Office for Official Publications of the European Communities.
3. Eurostat. (2008). *European Regional and Urban Statistics Reference Guide*. Luxembourg : Office for Official Publications of the European Communities.
4. ISTAT. (Undated). *URBAN AUDIT II THE IMPLEMENTATION IN ITALY Final Report submitted to the European Commission*.

<sup>37</sup> Files downloaded and checked July 5, 2010

5. Istat. (Undated). *Urban Audit III - Final Country Report ITALY - Grant n. 72501-2006-001-2006-492*.
6. Eurostat. (2008). *European Regional and Urban Statistics Reference Guide*. Luxembourg : Office for Official Publications of the European Communities.

## LATVIA

### Summary

- ◇ Latvian cities have been included in Urban Audit since UA II (Riga, Liepaja) **(1)**, and in UA III **(2)**. In total, two cities are concerned.
- LUZ definitions have not changed between UA II and UA III: “*The delineation of spatial units of Latvia’s territories participating in the project remained the same as in previous Urban Audit (UA) data collections*” **(3)**.
- Liepaja LUZ corresponds to one LAU 1: “*Also [LUZ] level was defined easily because functional urban zones for both cities were known. As they could be approximated with NUTS level 3 or 4 units, it was decided to create LUZ using NUTS 4 units*” **(4)**. Riga LUZ is an aggregation of LAU 1

### Building blocks

- ◇ Liepaja: LAU 1, elementary administrative unit (.i.e *Rajoni* and *republikas pilsētas*<sup>38</sup>)
- Riga: aggregation of LAU 1 (Riga and Ogre districts, and Jurmala city **(3)**).

### Correspondence with GISCO

Same number of LUZ<sup>39</sup>

[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)

### References

1. Eurostat. (2004). *Urban Audit, Methodological Handbook, 2004 edition*. Luxembourg: Office for Official Publications of the European Communities.
2. Eurostat. (2008). *European Regional and Urban Statistics Reference Guide*. Luxembourg : Office for Official Publications of the European Communities.
3. Latvijas statistika. (2008). *Urban Audit 2006 - Grant agreement No. 72501.2006.001-2006.478 - FINAL TECHNICAL IMPLEMENTATION REPORT LATVIA*.
4. Supe, Jolanta. (2004). *Data Collection Project - Urban Audit II - Latvia*. CSB of Latvia.

---

<sup>38</sup> A territorial reform occurred in Latvia in 2009. Before that (when UA II and UA III National Reports were written), LAU 1 were made of “districts” (i.e. *rajoni*) and “cities” (i.e. *republikaspilsētas*).

<sup>39</sup> Files downloaded and checked July 21, 2010

## LITHUANIA

### Summary

◇ Lithuanian cities have been included in UA II (Kaunas, Panevezys, Vilnius) **(1)** then in UA III **(2)**. In total, three cities are concerned.

□ LUZ definitions have not changed between UA II and UA III.

○ Each LUZ corresponds to one LAU 1. The National Report mentions some analyses before defining LUZ, but without more details: “*In co-operation with the cities, possible ways to define LUZ and SCD were examined. On the basis of analyses, the following territorial units [i.e. LAU 1] in the Urban Audit II were agreed to be used*” **(3)**.

The map below shows that one LUZ corresponds to two LAU 1 which are nested.



### Building blocks

◇ LAU 1, elementary administrative unit (i. e. *Savivaldybės*)

□ No aggregation

○ Vilnius: LAU 1 aggregation (City Core (Vilnius City) + Vilnius district + Elektrenai district + Trakai district)

### Correspondence with GISCO

Same number of LUZ<sup>40</sup>

[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)

### References

1. Eurostat. (2004). *Urban Audit, Methodological Handbook, 2004 edition*. Luxembourg: Office for Official Publications of the European Communities.

2. Eurostat. (2008). *European Regional and Urban Statistics Reference Guide*. Luxembourg : Office for Official Publications of the European Communities.

3. Lissauskaite, Jolita. (Undated). *FINAL REPORT ON URBAN AUDIT II*. Statistics Lithuania.

<sup>40</sup>Files downloaded and checked July 21, 2010

## LUXEMBOURG

### Summary

◇ Luxembourg city has been included in Urban Audit since the Pilot Phase (Luxembourg) **(1)**, then in UA II **(2)**, and in UA III **(3)**.

□ LUZ definition has changed between UA II and UA III.

For UA II, the LUZ corresponds to an aggregation of LAU2 (“we have defined the large urban zone (LUZ) regrouping 14 communes”**(2)**). However, according to Eurostat, the LUZ is underestimated **(4)**. For UA III, LUZ is enlarged to NUTS 0, as it can be deduced from GISCO.

○ In UA III, Luxembourg LUZ corresponds to Luxembourg country.

### Building blocks

◇ NUTS 0, elementary administrative unit (i.e. whole country)

□ No aggregation

### Correspondence with GISCO

Same number of LUZ<sup>41</sup>

[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)

### References

1. ERECO. (2000). *L'Audit Urbain, Vers un référentiel pour mesurer la qualité de la vie dans 58 villes européennes*. Luxembourg: Office des publications officielles des Communautés européennes.

2. CEPS/INSTEAD. (2003). *Urban Audit II Interim Report Luxembourg August 2003*.

3. Eurostat. (2008). *European Regional and Urban Statistics Reference Guide*. Luxembourg : Office for Official Publications of the European Communities.

4. Carlquist, Torbiörn. (2006, August 30). The Larger Urban Zones in the Urban Audit data collection. *Globalisation Impact on Regional and Urban Statistics*. Wroclaw.

---

<sup>41</sup>Files downloaded and checked July 19, 2010



## MALTA

### Summary

- ◇ Maltese cities have been included in UA II (Valetta, Gozo) **(1)**, then UA III **(2)**. In total, two cities are concerned. The National Report for UA III was not available.
- LUZ definitions have not changed between UA II and UA III (deduced from GISCO observations).
- Each LUZ corresponds to one NUTS 3 **(3)**.

### Building blocks

- ◇ NUTS 3, elementary administrative unit (i.e. *Gzejjer*)
- No aggregation

### Particular cases

Gozo: The City Core and the LUZ are both defined as the same NUTS 3

### Correspondence with GISCO

Same number of LUZ<sup>42</sup>

[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)

### References

1. Eurostat. (2004). *Urban Audit, Methodological Handbook, 2004 edition*. Luxembourg: Office for Official Publications of the European Communities.
2. Eurostat. (2008). *European Regional and Urban Statistics Reference Guide*. Luxembourg : Office for Official Publications of the European Communities.
3. National Statistics Office - Malta. (Undated). *Urban Audit Final Report*.

---

<sup>42</sup> Files downloaded and checked July 19, 2010

## NETHERLAND

### Summary

◇ Dutch cities have been included in Urban Audit since the Pilot Phase (Amsterdam, Rotterdam) **(1)**, then in UA II (Arnhem, Eindhoven, Enschede, Groningen, Heerlen, s'Gravenhage, Tilburg, Utrecht) **(2)**, and in UA III (Breda, Nijmegen, Apeldoorn, Leeuwarden) **(3)**. In total, fourteen cities are concerned.

□ LUZ definitions have changed between UA II and UA III.

For UA II, each LUZ corresponds to one NUTS 3: “Statistics Netherlands nevertheless proposed to use the Dutch NUTS 3 regions (called COROP Regions) as a proxy for the larger urban zones” **(4)**.

○ For UA III, LUZ correspond to the functional region called “stadsgewest”:

“More data came available [since UA II] for another proxy being 'stadsgewest', which represents better the relation of the city with its surroundings” **(5)**. The methodology used is described in a reference document **(6)** and on the Statistics Netherlands website **(7)**, but in Dutch language. It seems that two steps are followed:

The agglomeration is first delineated, starting from built-up area criteria (land use map from 1996), and then using some thresholds (population > 100 000 inh., employment > 50 000 persons and market area > 150 000 persons).<sup>43</sup>

The functional area surrounding the city is then built using following criteria<sup>44</sup>: commuters (data from Labour Force Survey realized between 1995 and 1997); residential migrations (Survey realized between 1996 and 1997); urban infrastructure information.

### Building blocks

◇ LAU 2, aggregation (i. e. *Gemeenten*)

□ Links: based on commuting data and residential migrations. The methodology is not specified in the National Report.

### Particular cases

Almere: “The exception is Almere, which territory is part of the stadsgewest Amsterdam, and by consequence has not a stadsgewest of its own” **(5)**.

### Correspondence with GISCO

Same number of LUZ<sup>45</sup>

[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)

### References

1. ERECO. (2000). *L'Audit Urbain, Vers un référentiel pour mesurer la qualité de la vie dans 58 villes européennes*. Luxembourg: Office des publications officielles des Communautés

<sup>43</sup><http://www.cbs.nl/nl-NL/menu/methoden/begrippen/default.htm?ConceptID=876>

<sup>44</sup><http://www.cbs.nl/nl-NL/menu/methoden/begrippen/default.htm?ConceptID=877>

<sup>45</sup>Files downloaded and checked July19, 2010

européennes.

2. Eurostat. (2004). *Urban Audit, Methodological Handbook, 2004 edition*. Luxembourg: Office for Official Publications of the European Communities.

3. Eurostat. (2008). *European Regional and Urban Statistics Reference Guide*. Luxembourg : Office for Official Publications of the European Communities.

4. Statistics Netherlands. (2003). *Urban Audit II - The implementation in the Netherlands*.

5. Statistics Netherlands. (2008). *Urban Audit 2006 - The implementation in the Netherlands*. 2008.

6. Vliegen, Mathieu. (2005). *Grootstedelijke agglomeraties en stadsgewesten afgebakend*. Voorburg/Heerlen : Centraal Bureau voor de Statistiek.

7. Centraal Bureau voor de Statistiek. (2010). *CBS - Home*. Checked July 19, 2010, on CBS: <http://www.cbs.nl/nl-NL/menu/home/default.htm>.

## NORWAY

### Summary

◇ Norwegian cities have been included in UA III (Oslo, Bergen, Trondheim, Stavanger, Kristiansand and Tromsø). A total of six cities are concerned.

□ Norway did not participated to UA II

○ LUZ correspond to the functional region called “storbyregioner” **(1)**. The documentation relating to LUZ has been transmitted by Statistics Norway to Urban Audit in Norwegian language **(2)**. It seems that LUZ definitions are largely based on criteria related to travel times and commuting data. According to information transmitted by Urban Audit, “Commuting was the leading criterion, travel times the adjustment instrument applied for the creation of the regions. A rule states that municipality’s center with over 10 percent commuting levels to a bigger center is affiliated with this center if it does not represent a separate commuting area”<sup>46</sup>.

### Building blocks

◇ LAU 2, aggregation (i.e. *Kommuner*)

□ Links: probably mainly based on commuting data (threshold 10%)

### Correspondence with GISCO

Same number of LUZ<sup>47</sup>

[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)

### References

1. ROG. (2005). *Data for storbyregioner/Urban Audit – avgrensing og inndeling av regionene*.
2. Det kongelige Kommunal og- regionaldepartement. (2003). *Storbymeldingen: Om utvikling av storbypolitikk*.

---

<sup>46</sup>Information provided by Urban Audit on December 3, 2010.

<sup>47</sup> Files downloaded and checked July 21, 2010.

## POLAND

### Summary

◇ Polish cities have been included in UA II (Warsaw, Lodz, Krakow, Wroclaw, Poznan, Gdansk, Szczecin, Bydgoszcz, Lublin, Katowice, Bialystok, Kielce, Torun, Olsztyn, Rzeszow, Opole, Gorzów Wielkopolski, Zielona Gora, Jelenia Gora, Nowy Sacz, Suwalki, Konin, Zory **(1)**), then in UA III (Czestochowa, Radom, Plock, Kalisz, Koszalin) **(2)**). In total, twenty-eight cities are concerned.

□ LUZ definitions have not changed between UA II and UA III.

○ LUZ definition is based on an aggregation of neighboring units (LAU 2 and LAU 1).

“Due to the lack of flow statistics (i.e. journeys to work) (...)” **(3)**, functional urban region could not be built.

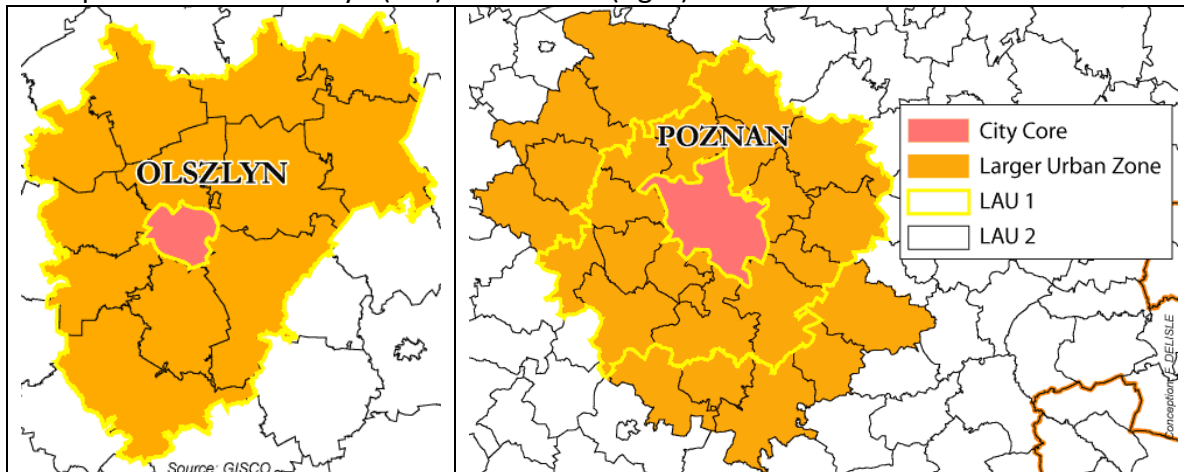
The methodology depends on City Core size **(3; 4; 5)**:

Population up to 100 000 inh.: LUZ corresponds to the surrounding LAU2 contiguous to the City Core (see for example, Suwalki at North-East);

Population between 100 and 250 000 inh.: LUZ corresponds to the surrounding LAU 1 that is contiguous to the City Core or to the LAU 1 containing the City Core (in most of the cases, the City Core is itself one LAU1, included inside another bigger LAU 1). (see for example Olszlyn, map below).

Population above 250 000 inh.: LUZ corresponds to the surrounding LAU 1 that is contiguous to the City Core, except if in some part of the ring, LAU 1 outer limit fits with local LAU 2 outer limit (see map below at the north-west and at the south-east of Poznan). Then, new LAU 2 are locally added (4 at the north-ouest and 5 at the south-east).

The map below shows Olszlyn (left) and Poznan (right) LUZ.



Probable changes in LUZ definitions will occur for next Urban Audit (UA IV).

### Building blocks

◇ LAU 2, aggregation (i.e. *Gminy*) and/or aggregation of LAU 1 (i.e. *Powiat*)

□ Links: based on contiguity and city core size.

## Correspondence with GISCO

Same number of LUZ<sup>48</sup>

[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)

## References

1. Eurostat. (2004). *Urban Audit, Methodological Handbook, 2004 edition*. Luxembourg: Office for Official Publications of the European Communities.
2. Eurostat. (2008). *European Regional and Urban Statistics Reference Guide*. Luxembourg : Office for Official Publications of the European Communities.
3. CENTRAL STATISTICAL OFFICE. (2008). *Urban Audit 2006 - Final Country Report (Poland) - Grant agreement no. 72501.2006.001-2006.484*.
4. Felczak, Dominika. (2003). *Urban Audit II Poland - Intermediate report summarising tasks 1 and 2 of Urban Audit program*. Central Statistical Office of Poland.
5. Młodak, Andrzej. (Undated). *Polish experiences and possibilities in realisation of the URBAN AUDIT programme*. Central Statistical Office.

---

<sup>48</sup> Files downloaded and checked June 24, 2010

## PORTUGAL

### Summary

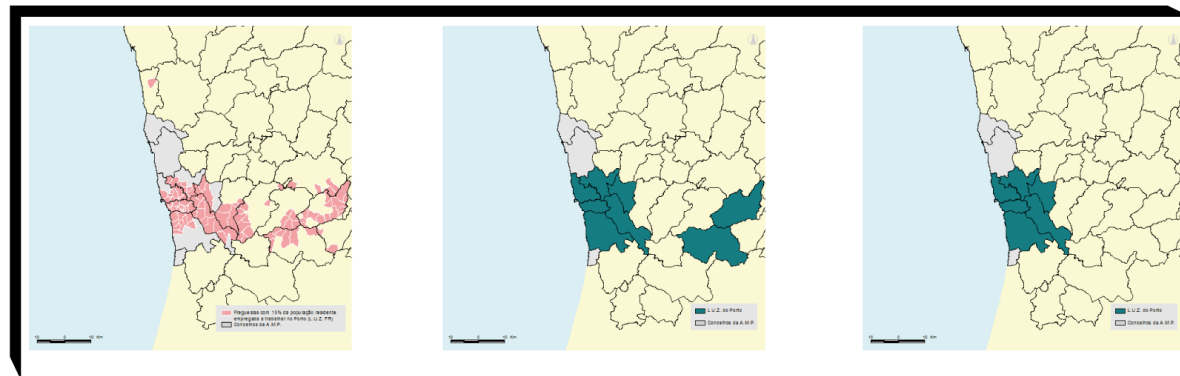
◇ Portuguese cities have been included in Urban Audit since the Pilot Phase (Lisbon, Oporto, Braga **(1)**), then in UA II (Aveiro, Coimbra, Setubal, Funchal, Ponta Delgada **(2)**), and in UA III (Faro **(3)**). In total, nine cities are concerned.

□ LUZ definitions have changed between UA II and UA III.

For UA II, each LUZ except Lisbon and Oporto, corresponds to one LAU 1: *“For all medium sized cities, it seems preferable to use NUTS 4 as a proxy for the functional urban region”* **(4)**. Lisbon and Oporto are defined as functional areas, with the methodology that has been used for all cities in UA III. For UA III, all LUZ are defined as functional areas.

○ LUZ are defined as functional areas, following three different steps, described on the map below **(5)**. The construction data come from Census 2001 **(5)**.

### LARGER URBAN ZONE



#### Etape 1

Identification of the “freguesias” (LAU level 2) that have a working commuting intensity to the urban audit’s cities (município) of more than 15% of the working resident.

#### Etape 2

Quantify the proportion of working people in all those “freguesias” belonging to a same “município” in the total of working people of the “município” they belong to. We have selected only the “municípios” where there was at least 50%

#### Etape 3

When needed, we have imposed the continuity spatial criterion to the core city

Source : Instituto Nacional de Estatística. Urban Audit III in Portugal, Final Report for the European Commission within the Framework of the Grant Agreement for an Action. 2008. Agreement Number 72501-2006-001-2006-485.

### Building blocks

◇ LAU 1, aggregation (i.e. *Municípios*)

□ Link: based on commuting data (threshold 15% at LAU 2 level, see map above)

### Particular cases

Setúbal: *“According to results Setúbal is the only city that still remains with “Luz identical to city” ”* **(5)**.

## Correspondence with GISCO

Same number of LUZ<sup>49</sup>

[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)

## References

1. ERECO. (2000). *L'Audit Urbain, Vers un référentiel pour mesurer la qualité de la vie dans 58 villes européennes*. Luxembourg: Office des publications officielles des Communautés européennes.
2. Eurostat. (2004). *Urban Audit, Methodological Handbook, 2004 edition*. Luxembourg: Office for Official Publications of the European Communities.
3. Eurostat. (2008). *European Regional and Urban Statistics Reference Guide*. Luxembourg : Office for Official Publications of the European Communities.
4. (Undated). *Urban Audit II in Portugal, Final Report for the European Commission within the Framework of the Grant Agreement for an Action*.
5. Instituto Nacional de Estatística. (2008). *Urban Audit III in Portugal, Final Report for the European Commission within the Framework of the Grant Agreement for an Action*.

---

<sup>49</sup> Files downloaded and checked June 24,2010



## ROMANIA

### Summary

◇ Romanian cities have been included in UA II (Alba Iulia, Arad, Bacau, Braila, Bucharest, Calarasi, Cluj-Napoca, Craiova, Giurgiu, Oradea, Piatra Neamt, Sibiu, TarguMures, Timisoara **(1)**) then in UA III **(2)**. In total, fourteen cities are concerned.

□ No information concerning evolution between UA II and UA III in National Reports

○ LUZ definition is based on an aggregation of neighboring units (LAU 2).

For UA II, LUZ are built with *“bordering communes situated around 15 km in the near vicinity of the cities” (3)*, except Bucharest (see below). These communes must be relevant with a law from 1968, *“abrogated in 1989, in which was considered suburban communes (term that is not used anymore) for each city of county” (3)*. *“In conclusion, LUZ were formed adding to the city the bordering suburban communes that are urbanized, industrialized and other well economic develop communes” (3)*.

For UA III, *“Romania proposed to be attached to the cities the bordering urbanized communes situated near vicinity of the municipalities selected as Urban Audit cities” (4)*.

### Building blocks

◇ LAU 2, aggregation (i.e. *Comuni/Orase/Municipiu*)

□ Links: distance criteria (15 km) and juridic criteria (law from 1968, see above)

○ Bucharest: aggregation of LAU 2, using a distance criteria (20 km) and a selection of LAU 2: *“we selected the communes and cities around 20 km from Bucharest, that are nearly integrated into city and well economically developed” (3)*.

### Correspondence with GISCO

Same number of LUZ<sup>50</sup>

[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)

### References

1. Eurostat. (2008). *European Regional and Urban Statistics Reference Guide*. Luxembourg : Office for Official Publications of the European Communities.
2. Eurostat. (2004). *Urban Audit, Methodological Handbook, 2004 edition*. Luxembourg: Office for Official Publications of the European Communities.
3. (Undated). *Urban Audit II for Candidat Countries Romania - Intermediate Report*.
4. Romania National Institute of Statistics. (2007). *URBAN AUDIT HISTORICAL DATA*.
5. Romania National Institute of Statistics. (2008). *Urban Audit - Final Operational Report Romania*.

<sup>50</sup>Files downloaded and checked July 21, 2010

## SLOVAKIA

### Summary

- ◇ Slovakian cities have been included in UA II (Bratislava, Košice, B. Bystrica, Nitra **(1)**), and in UA III (Prešov, Žilina, Trenčín, Trnava**(2)**). In total, eight cities are concerned.
- LUZ definitions have not changed between UA II and UA III.
- Each LUZ corresponds to one LAU 1 (“*District at the level of LAU1 was selected as LUZ in all cities*” **(3)**), except the LUZ of Bratislava (see below) **(4)**. According to the National Report, “*all proposals for LUZs are based on the employment zones, which were specified within the research work done for the SO SR [Statistical Office of the Slovak Republic] last year and we have agreed them also with the cities contact persons*” **(4)**.

### Building blocks

- ◇ LAU 1, elementary administrative unit (i.e. *Okresy*)
- No aggregation
- Bratislava: NUTS 3 (i.e. *Kraje*) **(4)**.

### Correspondence with GISCO

Same number of LUZ<sup>51</sup>

[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)

### References

1. Eurostat. (2004). *Urban Audit, Methodological Handbook, 2004 edition*. Luxembourg: Office for Official Publications of the European Communities.
2. Eurostat. (2008). *European Regional and Urban Statistics Reference Guide*. Luxembourg : Office for Official Publications of the European Communities.
3. The Statistical Office of the Slovak Republic. (2008). *Final Technical Report of the Grant 72501.2006.001-2006.487 Urban Audit data collection 2006/2007*. Bratislava.
4. Olexa, Michel. (2005). *Urban Audit II - PHARE 2001 Project - Intermediate Report for the Slovak Republic*. Bratislava.

---

<sup>51</sup>Files downloaded and checked July 19, 2010

## SLOVENIA

### Summary

- ◇ Slovenian cities have been included in UA II (Ljubljana, Maribor) **(1)**, and in UA III **(2)**. In total, two cities are concerned.
- LUZ definitions have not changed between UA II and UA III.
- Each LUZ corresponds to one NUTS 3 (“NUTS3 level [equals LUZ]” **(3)**). In Slovenia, NUTS 3 correspond to statistical regions, used as functional and planning areas<sup>52</sup> (“Data were collected for city Ljubljana (large size city) and for city Maribor (medium size city) namely for the following spatial units: Podravska statistical region (Ljubljana LUZ) and Osrednjeslovenska statistical region (Maribor LUZ)”**(4)**).

### Building blocks

- ◇ NUTS 3, elementary administrative unit (i.e. *Statistične regije*).
- No aggregation

### Correspondence with GISCO

Same number of LUZ<sup>53</sup>

[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)

### References

1. Eurostat. (2004). *Urban Audit, Methodological Handbook, 2004 edition*. Luxembourg: Office for Official Publications of the European Communities.
2. Eurostat. (2008). *European Regional and Urban Statistics Reference Guide*. Luxembourg : Office for Official Publications of the European Communities.
3. Statistical Office of the Republic of Slovenia. (Undated). *Urban Audit II - Quality Report on Data Collection - Country: Slovenia*.
4. Statistical Office of the Republic of Slovenia (SORS). (Undated). *Final Report - URBAN AUDIT 2006 - (Grant agreement NO 72501-2006-001-486)*.

---

<sup>52</sup>[http://www.stat.si/vodic\\_oglej.asp?ID=360&PodrocjeID=2](http://www.stat.si/vodic_oglej.asp?ID=360&PodrocjeID=2)

<sup>53</sup>Files downloaded and checked July 19,2010

## SPAIN

### Summary

◇ Spanish cities have been included in Urban Audit since the Pilot Phase (Madrid, Barcelona, Valence, Seville, Saragossa, Malaga) **(1)** then in UA II (Badajoz, Las Palmas, Logrono, Murcia, Oviedo, Palma di Mallorca, Pamplona/Iruña, Santander, Santiago de Compostela, Toledo, Valladolid, Vitoria/Gasteiz) **(2)**, and in UA III (Bilbao, Cordoba, Alicante/Alacant, Vigo, Gijon, Santa Cruz de Tenerife) **(3)**. In total, twenty-four cities are concerned.

□ LUZ definitions have changed between UA II and UA III.

For UA II, each LUZ correspond to one NUTS 3 **(2)**:

*“The administrative Spanish units were adapted to the three spatial levels defined in the Project. The Spanish statistical information has been obtained for these different spatial units: national, region (autonomous communities), provinces, municipalities, districts and sections”* **(4)**. However, in the years after UA II, some analyses showed that Spanish LUZ were overbounded (*“Countries where the LUZ were over-bound include (...) Spain”* **(5)**).

○ In UA III, LUZ definition is functional and based on commuting data:

*“A new design for the LUZ has been done for the 25 core cities, according to the community census table program for 2000/2001 which contains data on commuting between municipios. In some cases we tried to approach the non official Spanish NUTS level 4 division”* **(3)**. This definition was adopted specifically for Urban Audit needs.

Construction data come from census 2001.

### Building blocks

◇ LAU 2, aggregation (i.e. *municipios*)

□ Links: based on commuting data (threshold 15%)<sup>54</sup>

○ Madrid: NUTS 3, elementary administrative unit (i.e. *Provincias*)<sup>55</sup>

### Particular cases

L’Hopitalet de Llobregat: this city does not correspond to a LUZ because it is part of Barcelona LUZ.

### Correspondence with GISCO

Same number of LUZ<sup>56</sup>

[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)

<sup>54</sup>This threshold is deduced from our computations and observations, based on an Excel file sent by Urban Audit on December 7, 2010.

<sup>55</sup> According to information transmitted by Urban Audit January 2009. However, two LAU 2 (i.e. *Municipios*) are missing on GISCO.

<sup>56</sup> Files downloaded and checked July 5, 2010.

## References

1. ERECO. (2000). *L'Audit Urbain, Vers un référentiel pour mesurer la qualité de la vie dans 58 villes européennes*. Luxembourg: Office des publications officielles des Communautés européennes.
2. Eurostat. (2004). *Urban Audit, Methodological Handbook, 2004 edition*. Luxembourg: Office for Official Publications of the European Communities.
3. (2004). *Final Report, Grant Agreement of Urban Audit II in Spain Agreement number-2002 CE 16 0 AT 186*.
4. (Undated). *Final Country Report Spain*.
5. Carlquist, Torbiörn. (2006, August 30). The Larger Urban Zones in the Urban Audit data collection. *Globalisation Impact on Regional and Urban Statistics*. Wroclaw.

## SWEDEN

### Summary

◇ Swedish cities have been included in Urban Audit since the Pilot Phase (Stockholm, Göteborg) **(1)**, then in UA II (Malmö, Jönköping, Umeå) **(2)**, and in UA III (Uppsala, Linköping, Örebro) **(3)**. In total, eight cities are concerned.

□ LUZ definitions have changed between UA II and UA III.

For UA II, LUZ definition of Jönköping and Umea correspond to “Local Labour Market Areas” (“for the last two cities (Jönköping and Umeå) ‘Local Labour Market areas’ were used to form the LUZ” **(2)**).

○ In UA III, LUZ definition corresponds to the Local Labour Market Areas defined by Eurostat (“For the other seven cities ‘Local Labour Market areas’ defined by Eurostat were used to form the LUZ” **(3; 2)**). By checking GISCO data and maps, one can deduce that these LUZ delineations do not correspond to the two kinds of functional areas used by the Swedish statistics, the A-Region (*arbetsmarknadsregioner*) and FA-Region (*funktionella analysregioner*) **(4;5)**.

### Building blocks

◇ LAU 2, aggregation (i.e. *Kommuner*)

□ Links: No information

○ Stockholm: NUTS 3, elementary administrative unit (i.e. *län*) **(3)**

### Correspondence with GISCO

Same number of LUZ<sup>57</sup>

[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)

### References

1. ERECO. (2000). *L'Audit Urbain, Vers un référentiel pour mesurer la qualité de la vie dans 58 villes européennes*. Luxembourg: Office des publications officielles des Communautés européennes.

2. Statistiska centralbyrån. (Undated). *Urban Audit II - Final Report*. Stockholm : s.n., Undated.

3. Final Country Report [Sweden].

4. COMMIN. (2010). *COMMIN | The Baltic Sea Conceptshare*. Checked October 29, 2010, on COMMIN | The Baltic Sea Conceptshare: <http://commin.org/en/bsr-glossaries/national-glossaries/sweden/arbetsmarknadsregion.html>.

5. Statistics Sweden. (2010). *Lokala arbetsmarknader - – egenskaper, utveckling och funktion*. Örebro : Statistiska centralbyrån.

<sup>57</sup> Files downloaded and checked June 24, 2010.

## SWITZERLAND

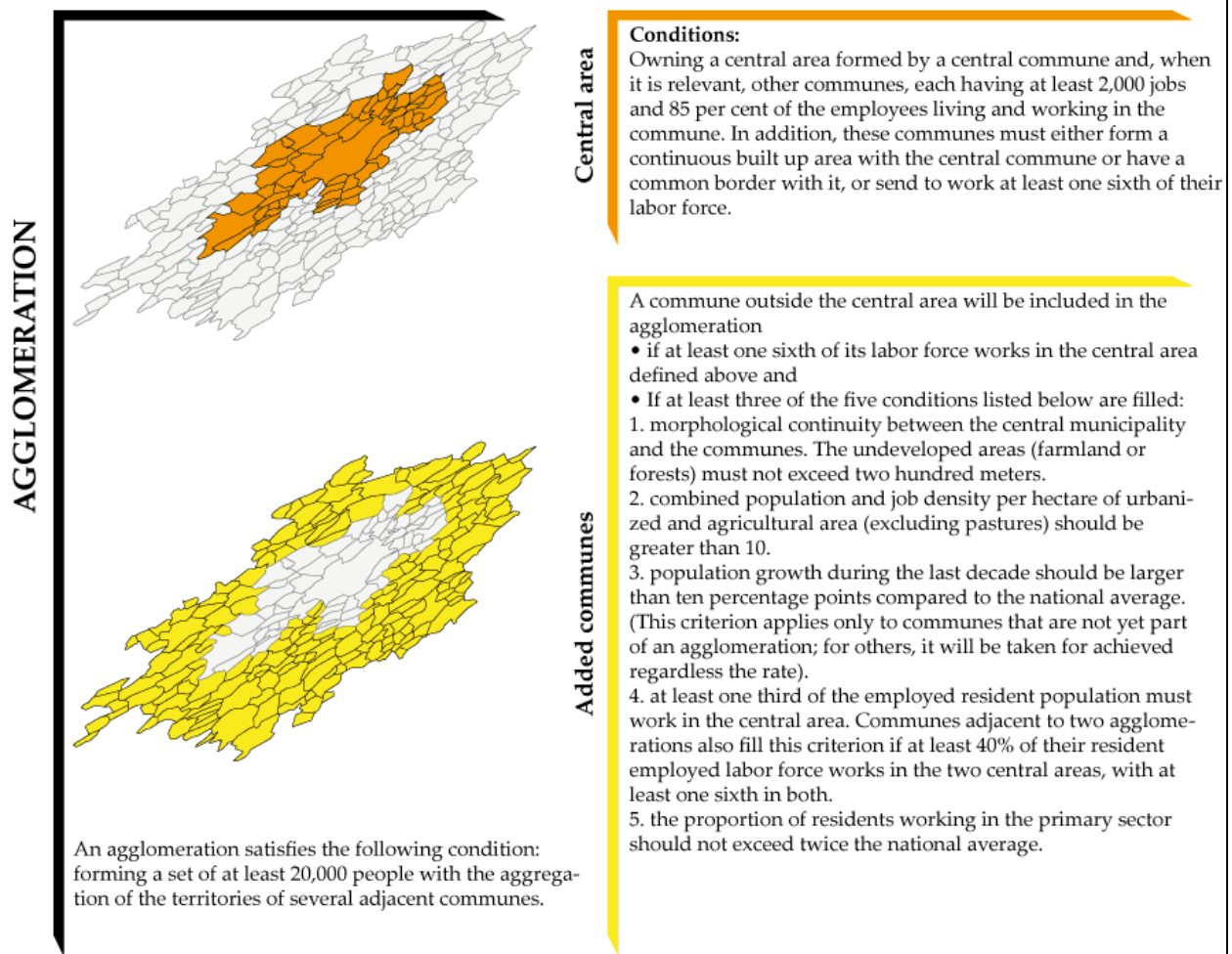
### Summary

◇ Swiss cities have been included in UA III as a pilot phase (Zürich, Genève, Bern, Lausanne), and five more cities have been added (Basel, Winterthur, St Gallen, Luzern et Lugano) **(1)**. In total, nine cities are concerned.

□ Switzerland did not participate to UA II

○ LUZ definitions correspond to the Swiss functional region called « agglomération », which is described in 2000 by the Office Fédéral de la Statistique **(2)**. Criteria were formulated in 1990, actualized in 2000. Construction data come from Census 2000<sup>58</sup>.

The figure below describes the different steps for building the “agglomérations”, translated in English by us:



Source : Schuler, Martin, Dessemontet, Pierre et Joye, Dominique. **Recensement fédéral de la population 2000 - Les niveaux géographiques de la Suisse.** Neuchâtel : Office fédéral de la statistique, 2005.

<sup>58</sup> <http://www.bfs.admin.ch/bfs/portal/fr/index/regionen/11/pro/01.html>

<b>Building blocks</b>
<ul style="list-style-type: none"> <li>◇ LAU 2, aggregation (i.e. <i>Gemeinden/Communes/Comuni</i>)</li> <li>□ Links: based on commuting data (threshold 16,66%)</li> <li>○ Geneva: Aggregation of one NUTS 3 (the <i>Canton</i> of Genève) and one LAU 1 (the <i>District</i> of Nyon, in the canton of Vaud).</li> </ul>
<b>Correspondence with GISCO</b>
<p>Different number of LUZ: 9 LUZs but only 4 in Gisco<sup>59</sup>. The five cities included in the second phase are missing.</p> <p><a href="http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1">http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1</a></p>
<b>References</b>
<ol style="list-style-type: none"> <li>1. Eurostat. (2008). <i>European Regional and Urban Statistics Reference Guide</i>. Luxembourg : Office for Official Publications of the European Communities.</li> <li>2. Jeanneret, Barbara. (2009). <i>Audit urbain 2006/07. Rapport final pour l'Union européenne (Suisse)</i>. Office fédéral de la statistique OFS.</li> <li>3. Schuler, Martin, Dessemontet, Pierre et Joye, Dominique. (2005). <i>Recensement fédéral de la population 2000 - Les niveaux géographiques de la Suisse</i>. Neuchâtel : Office fédéral de la statistique.</li> </ol>

---

<sup>59</sup> Files provided by Urban Audit September 5, 2010



## UNITED KINGDOM

### Summary

◇ English cities have been included in Urban Audit since the Pilot Phase (Birmingham, Leeds, Glasgow, Bradford, Liverpool, Edinburgh, Manchester, Cardiff) **(1)**, then in UA II (Aberdeen, Belfast, Bristol, Cambridge, Derry, Exeter, Leicester, Lincoln, London, Newcastle upon Tyne, Portsmouth, Sheffield, Worcester, Wrexham) **(2; 3)**, and in UA III (Coventry, Kingston-Upon-Hull, Stoke-on-Trent, Wolverhampton, Nottingham) **(4; 5)**. In total, twenty-seven cities are concerned.

□ LUZ definitions have not changed between UA II and UA III.

○ LUZ are defined according to a local consultation process:

*“The Larger Urban Zone or LUZ (...) are built up from LAU1 (Local Authority District) spatial units” **(5)**. “ONS [Office for National Statistics] sought the recommendation of relevant Local Authorities and Government Office Regions when constructing the LUZ area for each of the 24 cities under analysis. This produced a range of forms for the LUZs, that reflect the diversity of urban experience in the UK (...). The majority of cities assembled LUZs consisting of other Local Administrative Units. These were not necessarily arranged around the city in a ‘doughnut’ formation” **(3)**.*

### Building blocks

◇ LAU 1, aggregation (i.e. Lower tier authorities (districts) or individual unitary authorities, Individual unitary authorities or LECs (or parts thereof), Districts).

□ Links: Local consultation

### Particular cases

Derry: *“For Northern Ireland, Derry City was not given a LUZ. This was due to its rural surroundings as to have included neighbouring LAUs would not have been representative of the area” **(3)** ;*

Stevenage and Gravesham: *“Stevenage and Gravesham are included within the London LUZ as well as being Urban Audit cities on their own account” **(5)**;*

Lincoln: *“Lincoln City Council (...) opted not to have a LUZ comprised of neighbouring LAUs. Instead, a number of electoral wards were specified as representing the true reach of the city of Lincoln in its rural locale” **(3)**;*

Leeds and Bradford: *“Leeds and Bradford share a single LUZ” **(5)**; “It was virtually impossible to distinguish the urban reach of Leeds, discounting the effect of Bradford, and vice versa. There was also the pull which Bradford had on Leeds, etc.” **(3)**.*

Aberdeen: *“The LUZ for Aberdeen contained only one other LAU, Aberdeenshire. The total area of this LUZ, however, came to nearly 6500 km<sup>2</sup>. This is second only to the LUZ for London. It was decided that the reach of Aberdeen’s LUZ into its surrounding area was in part due to the relatively sparse population density in the vicinity” **(3)**.*

Wirral: This city does not correspond a LUZ<sup>60</sup>.

<sup>60</sup>According to information transmitted by Urban Audit January 2009 and GISCO September 5, 2010.

## Correspondence with GISCO

Same number of LUZ<sup>61</sup>

[http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative\\_units\\_statistical\\_units\\_1](http://epp.eurostat.ec.europa.eu/portal/page/portal/gisco/popups/references/administrative_units_statistical_units_1)

## References

1. ERECO. (2000). *L'Audit Urbain, Vers un référentiel pour mesurer la qualité de la vie dans 58 villes européennes*. Luxembourg: Office des publications officielles des Communautés européennes.
2. Eurostat. (2004). *Urban Audit, Methodological Handbook, 2004 edition*. Luxembourg: Office for Official Publications of the European Communities.
3. Office for National Statistics - Regional & Local Division. (Undated). *URBAN AUDIT II Implementation for the United Kingdom Final Report for the European Commission (Eurostat)*.
4. Eurostat. (2008). *European Regional and Urban Statistics Reference Guide*.
5. (2008). *Urban Audit 3 Final Country Report: United Kingdom*.

---

<sup>61</sup> Files downloaded and checked July 5, 2010

## References

ERECO. (2000). *L'Audit Urbain, Vers un référentiel pour mesurer la qualité de la vie dans 58 villes européennes*. Luxembourg: Office des publications officielles des Communautés européennes.

STATISTIKA CENTRALBYRAN. (Undated). *Urban Audit II - Final Report*. Stockholm : s.n., Undated.

Final Country Report [Sweden].

COMMIN. (2010). *COMMIN | The Baltic Sea Conceptshare*. Checked October 29, 2010, on COMMIN | The Baltic Sea Conceptshare: <http://commin.org/en/bsr-glossaries/national-glossaries/sweden/arbetsmarknadsregion.html>.

STATISTICS SWEDEN. (2010). *Lokala arbetsmarknader - - egenskaper, utveckling och funktion*. Orebro : Statistiska centralbyrå.

EUROSTAT. (2004). *Urban Audit, Methodological Handbook, 2004 edition*. Luxembourg: Office for Official Publications of the European Communities.

EUROSTAT. (2008). *European Regional and Urban Statistics Reference Guide*. Luxembourg : Office for Official Publications of the European Communities.

HUNGARIAN CENTRAL STATISTICAL OFFICE. (Undated). *Urban Audit III - Final Report*.

HUNGARIAN CENTRAL STATISTICAL OFFICE. (2004). *Urban Audit II - Final Report*.

STATISTICS NETHERLANDS. (2003). *Urban Audit II - The implementation in the Netherlands*.

STATISTICS NETHERLANDS. (2008). *Urban Audit 2006 - The implementation in the Netherlands*. 2008.

VLIEGEN, M. (2005). *Grootstedelijke agglomeraties en stadsgewesten afgebakend*. Voorburg/Heerlen : Centraal Bureau voor de Statistiek.

CENTRAAL BUREAU VOOR DE STATISTIEK. (2010). *CBS - Home*. Checked July 19, 2010, on CBS: <http://www.cbs.nl/nl-NL/menu/home/default.htm>.

STATISTICS FINLAND. (2007). *Eurostat - Urban Audit III - Final Country Report Finland*.

STATISTICS FINLAND. (2004). *Final Report: Urban Audit II Finland*.

**Figures list:**

**Figure 1: UA II and UA III LUZ..... 7**

**Figure 2: A variety of representations of Urban Audit “cities” (UA III) ... 9**

**Figure 3: Pareto-Zipf distribution of LUZ size (UA III) ..... 10**

**Figure 4: Population of LUZ in 2003-2006 (UA III) ..... 11**

**Figure 5: Typology of LUZ delineation (UA III) ..... 18**

**Figure 6: A European variety of LUZ commuting thresholds ..... 21**

**Figure 7: Intra-national heterogeneity in UA III definitions..... 22**

**Figure 8: New LUZ functional definitions..... 23**

**Tables:**

**Table 1: Number of cities, countries and indicators involved in the three Urban Audit rounds ..... 6**

**Table 2: LUZ number and min-max population by country (UA III)..... 12**

**Table 3: Availability of National Report per country and per Urban Audit round ..... 13**

**Insert:**

**Insert 1: Selection criteria of Urban Audit cities (<http://www.urbanaudit.org/help.aspx>)..... 10**

**Insert 2: Main problems encountered in National Report expertise (see country-sheets in Annex for more details) ..... 15**

**Insert 3: Model of the Country-Sheet ..... 17**

**Annex:**

**Austria ..... 26**

**Belgium ..... 27**

**Bulgaria..... 29**

**Croatia ..... 30**

<b>Cyprus</b> .....	<b>31</b>
<b>Czech Republic</b> .....	<b>32</b>
<b>Denmark</b> .....	<b>34</b>
<b>Estonia</b> .....	<b>36</b>
<b>Finland</b> .....	<b>38</b>
<b>France</b> .....	<b>39</b>
<b>Germany</b> .....	<b>42</b>
<b>Greece</b> .....	<b>44</b>
<b>Hungary</b> .....	<b>45</b>
<b>Ireland</b> .....	<b>47</b>
<b>Italy</b> .....	<b>49</b>
<b>Latvia</b> .....	<b>51</b>
<b>Lithuania</b> .....	<b>52</b>
<b>Luxembourg</b> .....	<b>53</b>
<b>Malta</b> .....	<b>54</b>
<b>Netherland</b> .....	<b>55</b>
<b>Norway</b> .....	<b>57</b>
<b>Poland</b> .....	<b>58</b>
<b>Portugal</b> .....	<b>60</b>
<b>Romania</b> .....	<b>62</b>
<b>Slovakia</b> .....	<b>63</b>
<b>Slovenia</b> .....	<b>64</b>
<b>Spain</b> .....	<b>65</b>
<b>Sweden</b> .....	<b>67</b>
<b>Switzerland</b> .....	<b>68</b>
<b>United Kingdom</b> .....	<b>70</b>



## THE FUNCTIONAL URBAN AREAS DATABASE

### CONTENT

- As a joint venture between 3 challenges of the Espon DB project (Urban data, Local data and Time series) and starting from the results of the previous Espon program we provide an update of the database of the Functional Urban Areas (FUAs) and Morphological Urban Areas (MUAs), as well as their inter-relations. Not only is it enhanced, it is also fundamentally enriched by the quality of the data provided, as the Functional Urban Areas (FUAs) are now delineated for most of the European countries of the Espon space at the LAU2 level.
- The FUAs are defined as labor basins of the MUAs which are themselves defined as densely populated areas, all this independently from any national, administrative or political definitions, but based instead on pure statistics.
- The main quality and advantage of these FUAs are their simple and universal definition throughout Europe, making them comparable in all the countries where they were delineated.
- Moreover we have also produced a list of indicators for these FUAs.

**ESPON 2013 DATABASE**

**MARCH 2011**



EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

18 pages

# LIST OF AUTHORS

Didier Peeters, Free University of Brussels, IGEAT

## **Contact**

dpeeter1@ulb.ac.be

tel. +32 2 650 50 77

# TABLE OF CONTENTS

<b>1</b>	<b>General Methodology</b> .....	<b>3</b>
1.1	Structure of the FUAs .....	4
1.2	Linking the different databases .....	6
1.3	The transposition of the MUAs into the new system .....	7
1.4	The correspondence between the SIRE codes and the eurogeographics codes .....	7
<b>2</b>	<b>The data processing</b> .....	<b>8</b>
2.1	The selection of the LAU2s in relation to the MUAs .....	8
2.2	The spatial operations .....	8
<b>3</b>	<b>Known issues</b> .....	<b>10</b>
<b>4</b>	<b>The indicators</b> .....	<b>11</b>
4.1	Methodology .....	11
4.2	Displaying FUAs database on maps, two examples .....	14
4.3	Missing data issues .....	16
4.4	Synthesis of the problems encountered to the attention of the Espon MC national representatives .....	17



# 1 General Methodology

Theoretically each MUA has a FUA, each FUA has a MUA near its center. But reality can be less straightforward and we find a good number of slightly more or much more complex cases.

First a MUA can be a secondary MUA in another's FUA. This happens when the extent of a FUA takes in a small MUA or when the population of the MUA has a commuting rate higher than the chosen threshold to another MUA. Most cases are obvious but a few are not, since the commuting flows are measured at the level of the LAU2 and a MUA can have several of these: one can therefore see that one part of a MUA is commuting toward another one while the rest of it is not. Those cases were solved by first checking if the secondary MUA was landlocked into the rest of the «main» FUA which would lead to consider the secondary MUA as actually part of the main FUA, second considering the literature and the commonly accepted facts. We don't see any major rule that could lead us to elaborate a methodology to classify these rare cases.

Second a FUA can have two «twin» MUAs (or even more) when the commuting flows from one to another are crossing each other (they are usually touching each other) and the LAU2s around them send commuters on a relatively high and equivalent level to the both of them. An example of this is Douai and Lens, or Béthune and Bruay-la-Buissière near Lille in France, or Locarno-Bellinzona in Switzerland.

We must here mention an unfortunate limitation of the database: except for the French residents there are no transnational commuters data provided, which is a non-sense for a European perspective. We have trans-border FUAs (see project Espon Metroborder) around the French frontier but only for the French residents working in another country, not for the opposite although we know that there are commuters from Germany, Belgium or Italy working in French FUAs (Strasbourg, Nice, ...). In the Espon Metroborder project this weakness was bypassed by assuming that the FUAs on both sides of the border are forming one entity, considering the results of Espon 1.4.3 based themselves on the literature.

Besides all the data limitations mentioned above we have encountered here and there difficulties in some areas like Scotland or former East-Germany, due to too big incoherencies between the different database, and this because of radical reshaping of the LAU2s during the last ten years.

*The poly-FUAs* : Espon 1.4.3 provided also a higher urbanization level called the "Polycentric Metropolitan areas" (poly-FUAs), which were made of groups of FUAs of the same neighborhood. These were based on a distance criteria :

*" In some cases, we have to consider the situation where different metropolises, with the centre of their cores distant from less than 60 km, are contiguous, or are only separated one from the other by other cities, with their own labour pool, or yet are bordered by other large, medium or small cities, distant from less than 30 km, also with their own individualised manpower basin. In these cases, we have identified conurbations of POLYCENTRIC METROPOLITAN AREAS (poly-FUAs). We have also considered as forming a POLYCENTRIC METROPOLITAN AREA two large cities distant one from the other less than 30 km and reaching together the level of 500,000 inhabitants. For the rest, we don't have considered as being a polycentric metropolitan*

*area two or more large, medium or small cities with contiguous manpower basins, even if they reach together the threshold of 500,000 inhabitants.*

*So to form a poly-FUA structure we must have either :*

- metropolises (> 500 000 inh.) with their centres less than 60 km apart, and labour basins touching each other*
- 2 large cities (> 250 000 inh.) with their centres less than 30 km apart, and labour basins touching each other*
- 1 metropolis and 1 large or medium city (> 100 000 inh.) with their centres less than 30 km apart, and labour basins touching each other*
- metropolises with their centres less than 60 km apart, labour basins separated only by the labour basin of a smaller FUA touching the both of them »*

These poly-FUAs were not taken into account into this project since the delimitation of the FUA has no impact on them but they remain useful and necessary

## **1.1 Structure of the FUAs**

The FUAs may be elements of the higher structures named poly-FUA (see above General Methodology) but this "super-structure" is also related to their internal structure which reflects their functional nature.

The following diagrams summarize for instance four different situations in a high-density area, implying quite different realities as regards functions, economy, management of mobility and territorial planning, but which could be confused if the analysis did not sufficiently explicit the definitions used. Even if these four patterns are purely theoretical, they are respectively globally based on the situation of an old coal basin for the first one (type1), the Ile-de-France Region for the second (type 2, with new cities functionally not much independent from Paris), the Belgian central metropolitan area (type 3) and the big London metropolitan area (type 4), where secondary centers of the external fringe of the FUA have more decisional autonomy and are moreover doubled by a belt of important or specialized cities (cf. Cambridge, Oxford) inside the FUA.

These should be understood as milestones of possible urbanization situations or evolution paths, rather than an exhaustive typology. There can be numerous intermediary states between any of these types and it would be vain to classify all the FUAs.

The following diagrams summarize for instance four different situations in a high-density area, implying quite different realities as regards functions, economy, management of mobility and territorial planning, but which could be confused if the analysis did not sufficiently explicit the definitions used. Even if these four patterns are purely theoretical, they are respectively globally based on the situation of an old coal basin for the first one (type1), the Ile-de-France Region for the second (type 2, with new cities functionally not much independent from Paris), the Belgian central metropolitan area (type 3) and the big London metropolitan area (type 4), where secondary centers of the external fringe of the FUA have more decisional autonomy and are moreover doubled by a belt of important or specialized cities (cf. Cambridge, Oxford) inside the FUA.

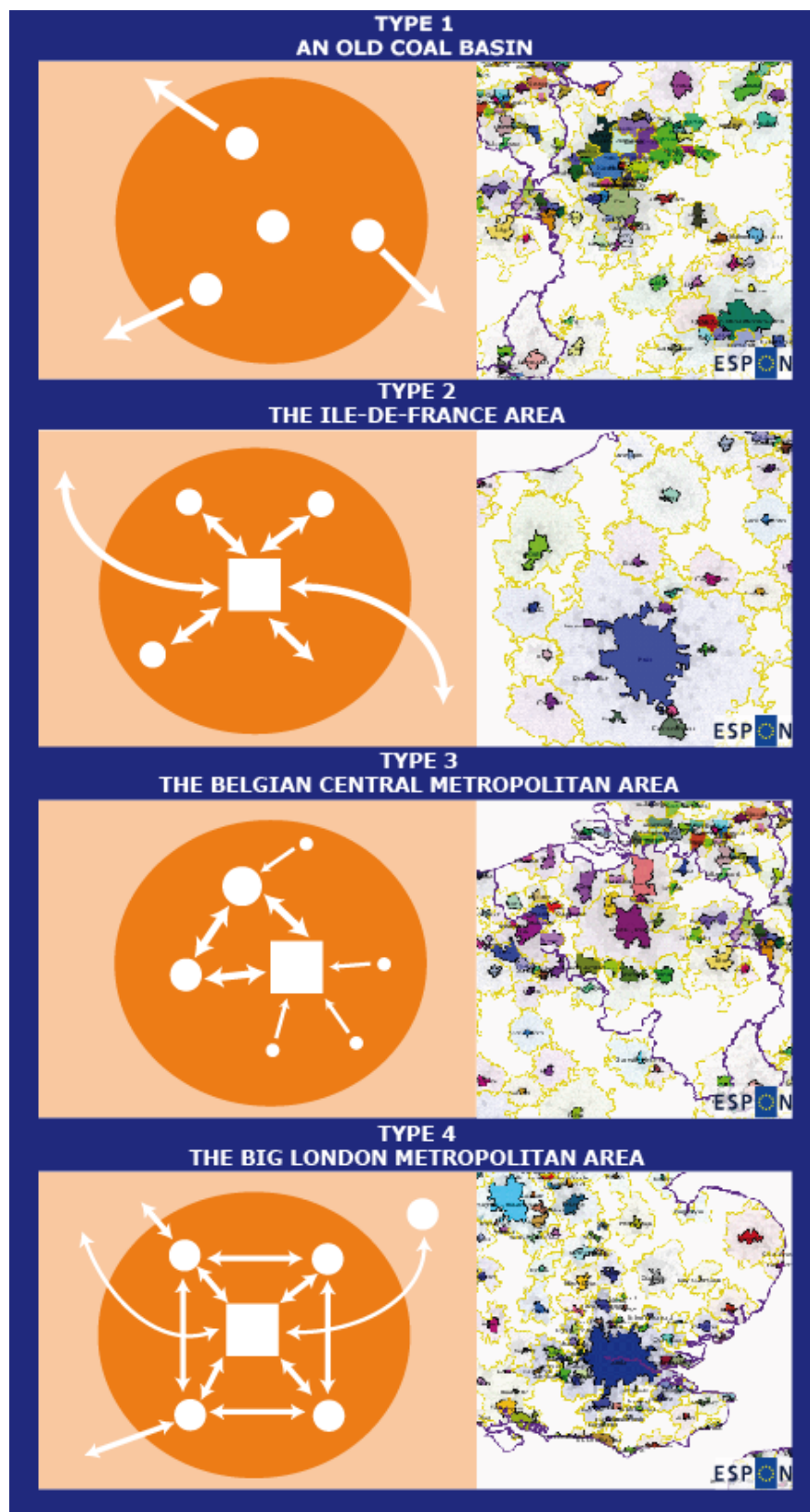


Figure 2 - Four different FUA types illustrated by the Ostrava region (1), the Ile-de-France Region (2), the Belgian central metropolitan region (3) and the London region (4).

## Available databases

Different databases had to be joined together to achieve this work:

- Eurogeographics (GISCO) 2008 geometries of the LAU2s with identification code "SHN"
- MUAs' compositions in NUTS-5s with identification code from "CMRGCD97"
- SIRE database from Eurostat, tables emp\_place\_tot\_mat\_01 and c\_emp\_place\_tot\_mat\_01 for the commuting numbers between LAU2s, emp\_place\_tot\_01 and c\_emp\_place\_tot\_01 for the non-commuters numbers (working in area), LAU2s with identification code "CODCOM"

The database providing the commuting numbers between LAU2s is named SIRE and is provided by Eurostat. Most of its content comes initially from the 2001 censuses.

Data are provided only at the LAU1 level for Bulgaria and Slovenia, the data seem to be insufficient for Hungary. 2001 is quite old now but we know by experience that the FUAs are relatively stable, they are built on a quite low commuting rate and made simpler and coherent by filling the holes in them (LAU2s that don't reach the commuting threshold but landlocked in the FUA) and by eliminating the LAU2s remotely detached from the rest of the FUA. So chances are that any changes in the commuting rates would merely affect the general shape of the FUA, moreover if any they would probably affect small LAU2s at the fringes that don't have too much influence. The objective is not to compute the populations of the FUAs with accuracy close to one inhabitant but to have a good magnitude and to establish functional relations. But still changes can happen and it would be a good thing to have the opportunity to make this work on more recent data.

<i>Entity</i>	<i>Year</i>	<i>Source</i>
LAU2s spatial definition	2008	Eurogeographics
MUAs population	2001	Espon 1.4.3
FUAs population	2001 and 2006*	LAU2s population, Eurostat
Commuting numbers	2001*	SIRE database, Eurostat

*Table 1 - Sources of the different data used in the elaboration of the FUAs and the MUAs*

## 1.2 Linking the different databases

The first step of the work was to link all these tables with their different identification codes. We got the help from Eurostat which provided us with a table of correspondence between CODCOM and 3 other codes (COMM\_ID, LAU2\_code and nsi\_code), as well as the NUTS-3 code for 2004 and 2006, which was easier to match

with the other available codes although not straightforward. A resulting table was produced, certainly not perfect, especially because of the difficulty to deal with complex cases where the matching was a n-to-n type. This is of course a time series issue, since we are dealing with data from 1997, 1999 and 2006 and it is not always possible to determine what became what. At first we made a spatial join between the old CMRGCD97 geometries and the 2008 geometries but the matching was not very good, due to some distortion in the projection. By checking the names and the code similarities we kept what appeared to be good and worked the imperfectly matched codes differently. For this we carefully used the names of the entities in the different sources and the code similarities, starting with the steadier methods and using progressively less steady solutions. We always used the code similarities and checked with the names. Progressively we eliminated most of the difficulties.

The objective of this first step was to transpose the LAU2s components of the old MUAs from the 1997 system into the new 2008 one, and also to match the CODCOM from the SIRE database with the EUROGEOGRAPHICS 2008 boundaries. So we stopped when we got a result satisfying for this objective but it should be improved by splitting this single big table into 3 or 4 more simple tables, but we didn't have enough time to produce something more rigorous.

### **1.3 The transposition of the MUAs into the new system**

Espon 1.4.3 had produced the delineation of about 2000 MUAs (the 300 smaller not published) with a total of about 11000 NUTS-5 in them. All these NUTS-5 were transposed into the EUROGEOGRAPHICS 2008 LAU2s delimitations as explained above. Only a fistful of new MUAs was added during the FUA production but no further research was made on the identification of the MUAs. This might be done by inserting the UMZ database (produced in this same Espon Database project) into this urban areas delimitation tool.

### **1.4 The correspondence between the SIRE codes and the eurogeographics codes**

CODCOM (SIRE) and SHN (EUROGEOGRAPHICS) share no similarities but the correspondence file provided by EUROSTAT was used and we have completed some relations due to differences between 2006 (SHN in the EUROSTAT file) and 2008 (EUROGEOGRAPHICS), by the same method based on the code similarities and the names correspondences.

## 2 The data processing

All the following steps are realized in SQL (MySQL or PostgreSQL) with postgis 1.5 for the spatial steps.

### 2.1 The selection of the LAU2s in relation to the MUAs

The initial work was done with the SIRE matrix. The table `emp_place_tot_mat_01` (and `c_emp_place_tot_mat_01`) is a simple matrix of about 2 millions lines containing the residence LAU2 (CODCOM), the working place LAU2 (DESCODCOM) and the number of commuters. The table `emp_place_tot_mat_01` (and `c_emp_place_tot_mat_01`) was completed with the active population working in their residence LAU2, in order to take into account all the working places. The MUAs were added in another column (this is not an orthodox way to proceed but it made things easier) in relation to the work place LAU2.

The process then consisted in computing the total number of commuters going in a MUA, for each CODCOM and each MUA. In the same time a commuting rate is computed by comparing this number to the «economically active population» from the tables `emp_place_tot_01` (EU15) and `c_emp_place_tot_01` (former candidate countries). Only the rates of 10 % or higher were kept.

Then specific cases are processed because of the incoherencies between the different databases that were not solved by the correspondence table mentioned above, especially the countries where the commuting data are provided at the LAU1 level or cases of big cities considered as a single entity in one database and a set of multiple LAU2s in the other (Budapest, Bratislava, Paris), or transborder data provided by other sources (Luxembourg, Belgian border), or Hungary where there is no active population data provided.

We then obtain a list of LAU2s with their respective MUA work places and their commuting rates, all this forming kind of «proto-FUAs» since each MUA is considered as the seed of a potential FUA and since the LAU2s can be part of several «proto-FUAs». This is a raw material to build the real FUAs, with their complex elements.

### 2.2 The spatial operations

The geographical objects for a map representation are produced.

The MUAs are simply aggregated from the LAU2s in the Eurogeographic 2008 layer. This step is easy because the MUAs were previously delineated as explained above.

The FUAs are much more complicated. The objective is to obtain coherent areas, without holes in them and no isolated parts. Here are the different steps.

- Selection of the main destination of the commuters for each LAU2, among all the MUAs toward which the commuting rate is higher than 10 % . Actually the different destinations are ranked and the first one is kept. In case of equality (same number of commuters toward 2 MUAs) the bigger MUA is selected as main destination.
- The MUAs are grouped into FUAs, according to the main destination of the commuters of each MUA and/or according to the literature as explained in the main report and above.
- The exterior rings of the FUAs are created and the bigger part is kept. Several parts can be kept where the separation is due to geographical particularities (islands, both sides of a fjord,...).
- The LAU2s spatially enclosed in the exterior rings are selected and attributed to their FUA, so that the holes are filled.

### 3 Known issues

Besides the unavoidable incoherencies between the different databases already mentioned above or in the main report there are specific problems we could not bypass :

- The original definition of the FUA is based on a proportion of the 'occupied' active population but all we have is the 'economically' active population, which includes the unemployed population. So this distorts the commuting rate by overestimating it by a maximum factor of 1.1, considering a 10 % unemployment.
- The active population in Germany is wrong in at least half of the municipalities! This problem is somehow minimized because, first, when a LAU2 has a commuting flow toward more than one MUA (for example in Rhein-Ruhr region) what matters is the highest number of commuters whatever the active population is, considering that this happens in areas with a strong peri-urbanization and without free space between the FUAs. Second, the process of building the FUAs eliminates the LAU2s that are not among the others for a same FUA and homogenizes the area by making a ring around the LAU2s of the FUA. Third, the LAU2s where the commuters' number is higher than the active population are ignored. Fourth, the overestimation of the FUAs (there can't be underestimation) happens by including probably small municipalities at the fringes of the areas. But still, errors remain.
- Same remark for Slovakia, but apparently with a smaller proportion of erroneous data.
- In former East-Germany, especially in the Saxe-Anhalt, there were many municipalities merging after the reunification. This leads to many incoherencies between SIRE and EUROGEOGRAPHICS, i.e. we have commuting numbers for LAU2s that don't exist anymore and we didn't spend time on trying to redistribute the commuters in the new LAU2s. This might perhaps be done but not in this project.
- In the area between Glasgow and Edinburg we have the same problem than in Saxe-Anhalt (see above).
- In Portugal the commuting numbers in the core city areas are provided at the LAU1 level but this is not a problem because in every case we have a ring of municipalities (LAU2s) around this LAU1 that are included in the FUA. So any central municipality that would be out of the MUA but into that LAU1 area would be included in the FUA anyway.
- Technically speaking, instead of making a ring surrounding the FUAs with the Postgis `ST_ExteriorRing` function we could explore the possibility to use the `ST_convexHull` function that minimizes the quirks in the shape of some FUAs, but it could lead also to exaggerate their size.

Whenever new data would become available, like especially the active population in Germany, we would rebuild the FUAs immediately.



## 4 The indicators

In collaboration with the project Espon METROBORDER we have produced a list of indicators for all the FUAs. Since the FUAs are defined at the LAU2 level any information available at that same level can be used to characterize the FUAs, otherwise it is sometimes possible to use the NUTS3 values, as explained below.

<i>Indicator</i>	<i>Year*</i>	<i>Source</i>	<i>Unit</i>
Population	2001	LAU2, EUROSTAT	inhab.
	2006	LAU2, EUROSTAT	inhab.
Population variation	2001 - 2006	LAU2, EUROSTAT	%
FUAs Areas	2008	EUROGEOGRAPHICS	km <sup>2</sup>
MUAs Areas	2008	EUROGEOGRAPHICS	km <sup>2</sup>
Compactness	synthetic indicator		%
GDP	2006	NUTS-3, EUROSTAT	Euro
GDP/inhab.***	synthetic indicator		Euro/inhab.
Economical structure	2006	NUTS-3, EUROSTAT (6 big NACE sectors)	%
Unemployment**	2006	NUTS-3, EUROSTAT	%

*Table 2 - list of the indicators available in the Espon database*

\* : For some indicators some data are coming from different years. See the missing data issues below.

\*\* : Unemployment values should be used «with caution» !

\*\*\* : see the map below

### 4.1 Methodology

For the population and area indicators we have simply computed the FUA values from the LAU2 values. The compactness is the % of the population of the FUA actually living in the main MUA.

The economical indicators are computed by using the NUTS-3 values on which we apply a population ratio between the NUTS-3 and the intersection of the FUA and the NUTS-3. This is possible because we now have the LAU2 composition of the FUAs which make the link between those and the FUAs possible. The methodology is

therefore useful to assess the urban areas throughout Europe (except in the countries where no commuting data are available – see above).

By way of illustration let's look at the Basel case. Basel is a complex case, it is a transborder metropolitan urban region with 5 MUAs and involving 7 NUTS-3.

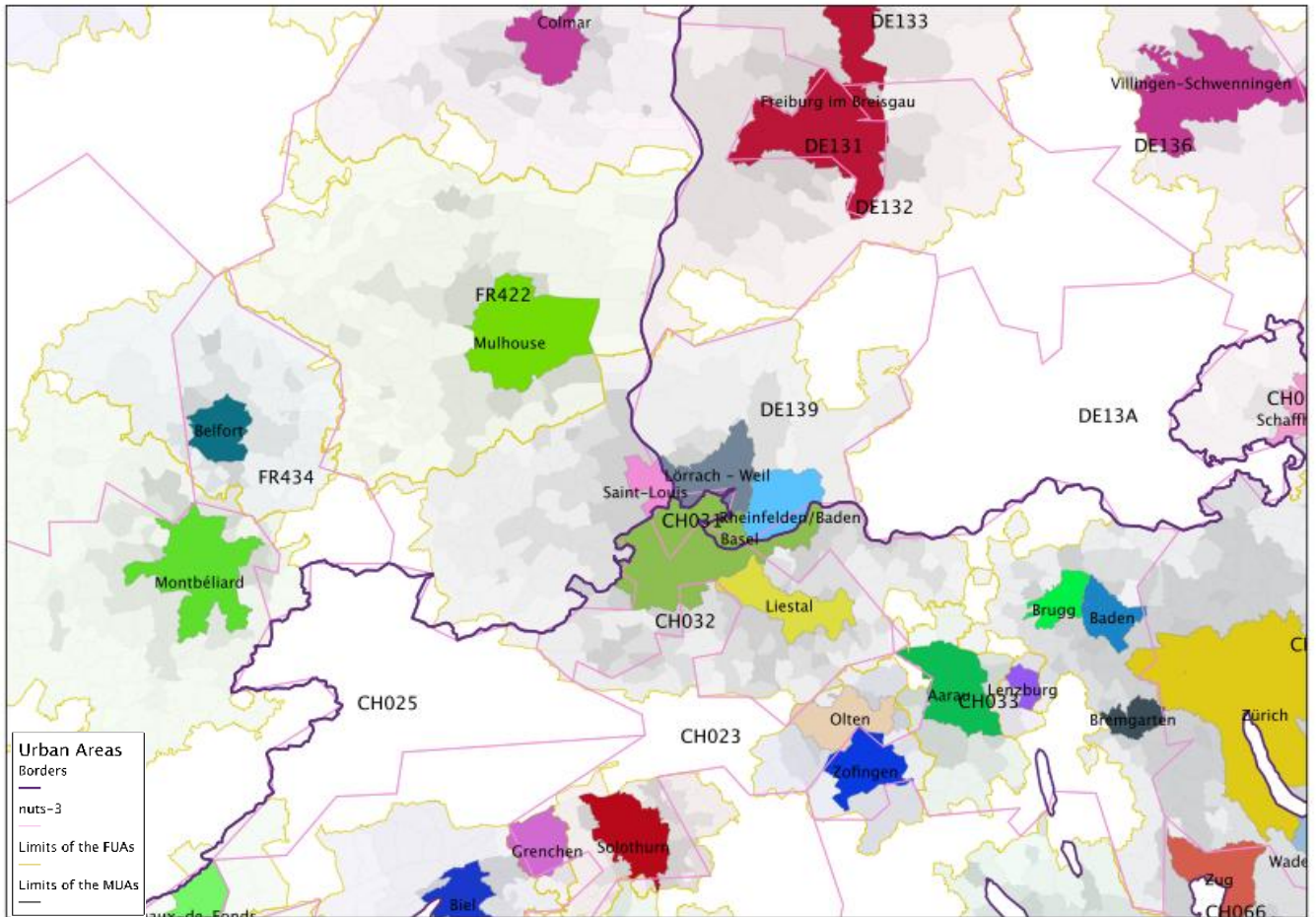


Figure 1 - The FUA of Basel in its region

The MUAs are shown in plain colors, the FUAs are in light colors delimited by a yellow line, the NUTS-3 are represented by the pink lines, and finally the gray shade shows the intensity of the commuting (the more grayish the higher the number of commuters is). The national borders are violet.

nuts3_06	pop_nuts3	pop_fua	pop_fua in nuts3	coef
CH023	248	827,3	31,7	13
CH025	69	827,3	0,1	0
CH031	185	827,3	188,1	100
CH032	266	827,3	258,7	100
CH033	572	827,3	52,6	10
DE139	221	827,3	215,6	100
FR422	738	827,3	91,4	13

*Table 3 - correspondence between the nuts-3 and the FUA of Basel.*

We see here (table 3) that the FUA of Basel extends over 7 nuts-3 in 3 countries, and we see in the table that this corresponds to different population values according to the nuts-3. We have computed for each of them a coefficient ('coef') in % giving what part of the nuts-3 indicator (for instance the GDP) we take from every nuts-3, the total giving the indicator value (for instance the GDP) of the FUA. In the particular case of the GDP or added values we assume that the productivity is equal everywhere inside each NUTS-3, which is not a bold assumption. For every other indicator we use the same methodology as long as we can make the same kind of uniformity hypothesis.

## 4.2 Displaying FUAs database on maps, two examples

The FUAs database allows displaying innovative results for the all ESPON Area. For instance, the maps of GDP per capita 2006 (figure 3) and evolution of population 2001-2006 (figure 4) show strong contrasts between FUAs, and give an additional picture of the situation as compared to classical maps produced at NUTS3 level. However, it is important to have a look to keep in mind that some of the values are estimated (cf section 4.3 below) and such results must be interpreted carefully.

Gross domestic product per capita in the Functional Urban Areas (2006)

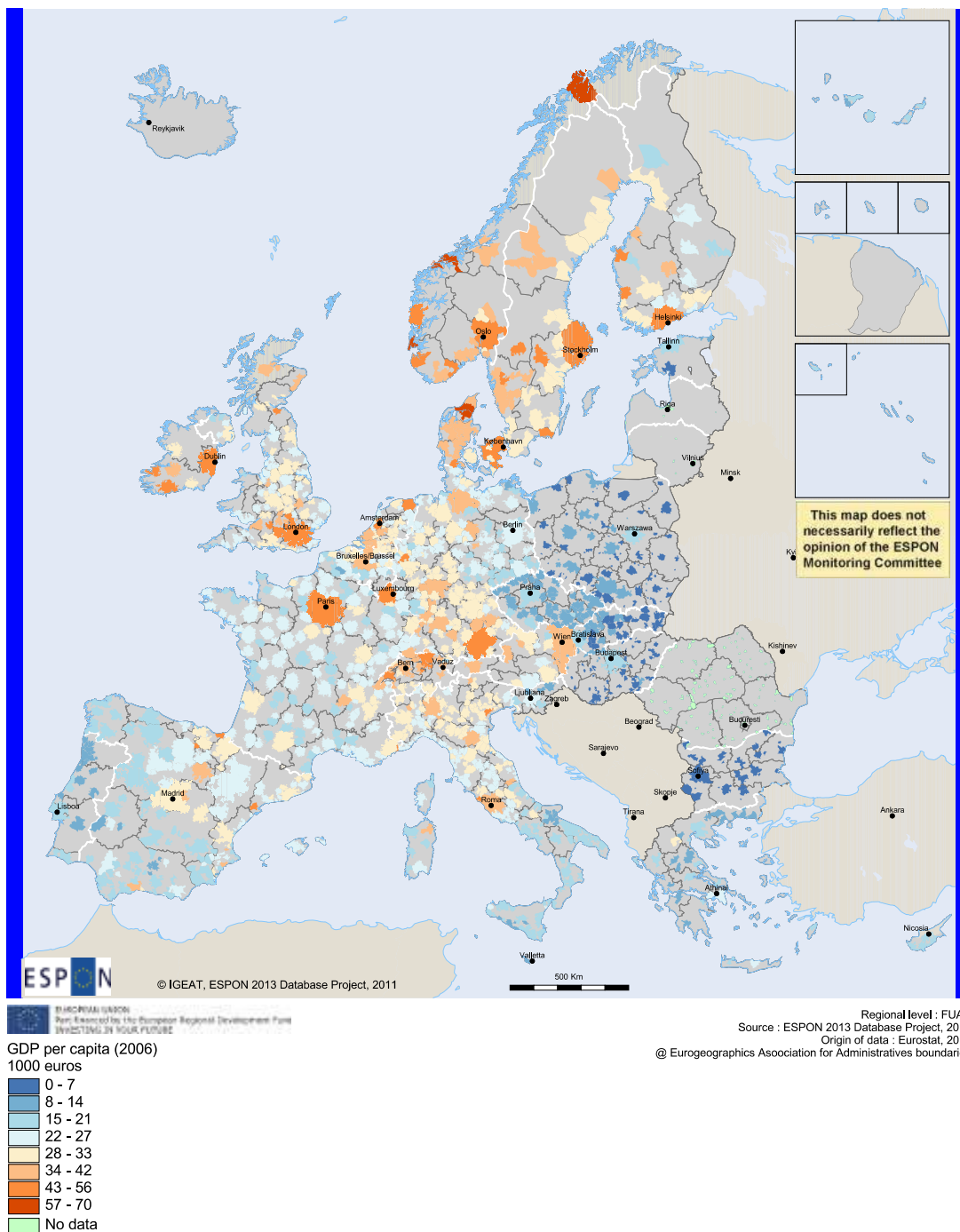
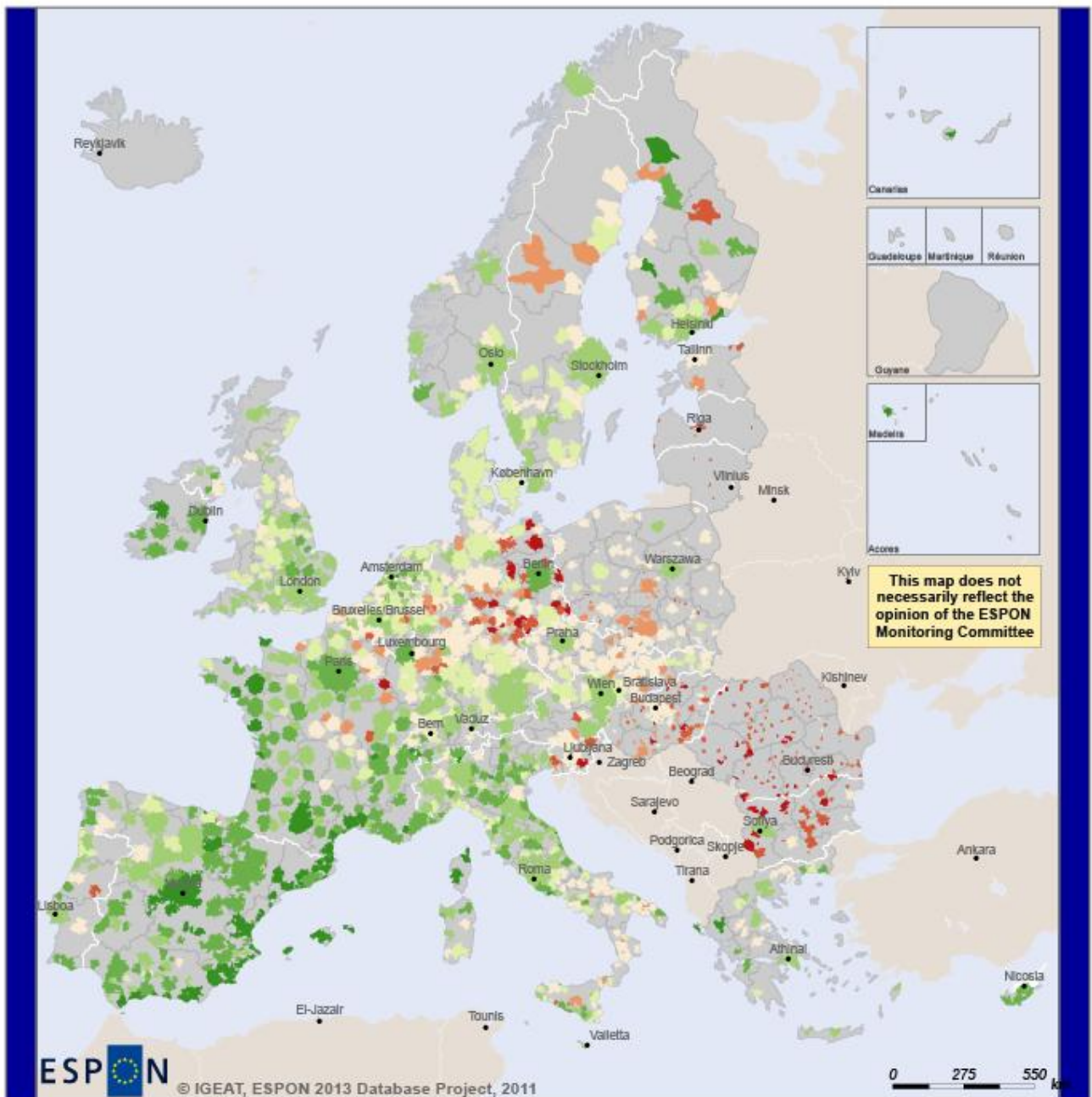


Figure 3 – Gross Domestic Product per capita 2006 in the FUA delineation



ESPON

© IGEAT, ESPON 2013 Database Project, 2011

EUROPEAN UNION  
Part-financed by the European Regional Development Fund  
INVESTING IN YOUR FUTURE

Regional level: FUAs  
Source: ESPON 2013 Database Project, 2010  
Origin of data: Eurostat, 2010  
© EuroGeographics Association for administrative boundaries

**Evolution of population 2001-2006,  
annual growth rate (%)**

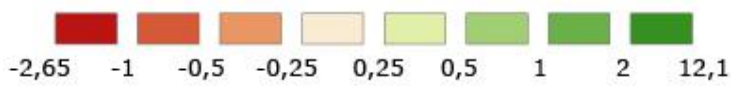


Figure 4 – Evolution of population 2001-2006 in the FUA delineation

### 4.3 Missing data issues

**Poland:** So far there are no commuting data in Poland. We received from Przemyslaw Sleszynski (Polish Academy of Sciences) a set of functional areas based on the last census data (2002) and other data from 2004 like the socio-cultural profiles of the population. We have included these data 'as is' in our database.

**Romania, Latvia and Lithuania:** No commuting data are provided for Romania, Latvia and Lithuania and we couldn't find any substitute, so we used the population data from Espon 1.4.3..

**LAU2 Population in 2006:** We don't have population numbers at the LAU2 level for Bulgaria, Cyprus, Denmark, United Kingdom, Lithuania, Latvia, Portugal and Romania. So we used the NUTS-3 replacement methodology described above to obtain the population of the FUA in 2006.

**GDP in Switzerland and Norway:** The GDP for Swiss and Norwegian FUAs are from 2005, due to missing data in Eurostat for the NUTS-3.

**Unemployment:** At the NUTS-3 level the unemployment values are not available in 2006 for Denmark, Sweden, Ireland, Switzerland, Sachsen-Anhalt and two Italian provinces (Sassari and Cagliari). See the synthesis table below for the replacement years.

Indicators	Countries	Substituting Years
Commuting data	Poland Romania Latvia Lithuania	-
2006 LAU2 population	Bulgaria Cyprus Denmark Lithuania Latvia Portugal Romania United Kingdom	NUTS-3 replacement
GDP	Switzerland Norway	2005
Unemployment	Denmark Sweden Sachsen-Anhalt	2007
	Ireland	2004
	Switzerland	2005
	Sassari province (ITG25) Cagliari province (ITG27)	2008
	Northern Ireland	2003

Table 4 - synthesis of the missing data issues

## 4.4 Synthesis of the problems encountered to the attention of the Espon MC national representatives

For every country except France it should be very appreciated to have the commuters working in a foreign country, at the LAU2 level, so that we could delineate transnational Urban Areas.

It would be also interesting to enrich to local data by providing population figures for more years (2010 ?) than 2006 in order to be able to make time series, and of course any local (LAU2 level) indicators would be useful, such as economical, employment and social indicators.

Country	Issue	Description	Request
<b>Bulgaria</b>	Commuting data at LAU1 level only		Commuting and active population data at LAU2 level
	Missing population numbers for 2006		Population numbers for 2006
<b>Cyprus</b>	Missing population numbers for 2006		Population numbers for 2006
<b>Germany</b>	Active population at LAU2 level in SIRE are wrong	At least 50 % of the LAU2s have active population (much) lower than the numbers of commuters, which is impossible.	Correct active population at LAU2 level.
	Many LAU2s reshaping	The SIRE LAU2s are quite different from the GISCO LAU2s because of municipality merging in former east Germany regions. This leads to incoherences between the databases	Commuting statistics at LAU2 level according to the current LAU2s delimitations
	The origin of the commuting data.	We heard that the german commuting data are provided by the insurance companies, which probably biases the figure.	Commuting data at LAU2 level based on better sources.
<b>Denmark</b>	Many LAU2s reshaping	The SIRE LAU2s are quite different from the GISCO LAU2s because of municipality merging and Nuts reshaping. This leads to incoherences between the databases.	Commuting statistics and active population numbers at LAU2 level according to the current LAU2s delimitations
	Missing population numbers for 2006		Population numbers for 2006
<b>Estonia</b>	Apparently insufficient commuting data	The FUAs of Tartu and Tõrva seem to be placed away from these cities, which probably comes from inaccurate commuting data. We don't know the origin of this problem.	Better commuting and active population data at LAU2 level
<b>Hungary</b>	Apparently insufficient commuting data	Globally the Hungarian FUAs don't appear to be well delineated, probably because of inaccurate commuting and/or active population data.	Better commuting and active population data at LAU2 level
<b>Lithuania</b>	No commuting data.		Commuting and active population data at LAU2 level
	Missing population numbers for 2006		Population numbers for 2006

<b>Latvia</b>	No commuting data.		Commuting and active population data at LAU2 level
	Missing population numbers for 2006		Population numbers for 2006
<b>Netherlands</b>	Many LAU2s reshaping	The SIRE LAU2s are quite different from the GISCO LAU2s because of municipality merging and Nuts reshaping. This leads to incoherences between the databases.	Commuting statistics and active population numbers at LAU2 level according to the current LAU2s delimitations
<b>Poland</b>	No commuting data.		Commuting and active population data at LAU2 level
<b>Portugal</b>	Commuting data at LAU1 level only in the urban areas.		Commuting and active population data at LAU2 level everywhere
	Missing population numbers for 2006		Population numbers for 2006
<b>Romania</b>	No commuting data.		Commuting and active population data at LAU2 level
<b>Slovenia</b>	Commuting data at LAU1 level only		Commuting and active population data at LAU2 level
<b>Slovakia</b>	Many active population at LAU2 level in SIRE are wrong	Some LAU2s have active population (much) lower than the numbers of commuters, which is impossible.	Correct active population at LAU2 level.
<b>United Kingdom</b>	Many LAU2s reshaping	The UK LAU2s change very frequently but the commuting data are old. This leads to incoherences between the databases, especially in Scotland.	Commuting statistics and active population numbers at LAU2 level according to the current LAU2s delimitations
	Missing population numbers for 2006		Population numbers for 2006

*Table 5 - synthesis of the problems encountered*