



Atelier- EXtraction de Connaissances à partir de
donnéEs Spatialisées (EXCES) – Nice 06/12/16

CSW_Harvester

une chaîne d'extraction et d'analyse d'informations à partir des services
web de catalogage (CSW) des Infrastructures de Données Géographiques
nationales et régionales en France



Mathias Rouan (mathias.rouan@univ-brest.fr)



Julie Pierson (julie.pierson@cnrs.fr)

Plan

1. Contexte : le projet GÉOBS
2. Définitions : ISO19115, CSW...
3. La chaîne de traitement
 - a. extraction et archivage
 - b. Analyse et visualisation
 - i. Extraction d'informations statistiques
 - ii. Extraction d'information spatiales
4. Conclusion

Contexte : le projet GÉOBS

- Projet de recherche (M NOUCHER / F GOURMELON)
- Les Infrastructures de Données Géographiques dans la gouvernance informationnelle de l'environnement
- Analyser les stratégies, usages et contenu des IDG
- Corpus de 45 IDG (16 nat. et 29 rég.) ← inventaire 2014



Définitions : ISO19115, CSW

- Métadonnées sur l'information géographique
- Norme ISO19115
- Catalogue et services web OGC
- CSW : Catalogue Service for the Web
 - *GetCapabilities* récupérer les capacités du service (description, opérations, critères de recherche, standards)
 - *GetRecords* réaliser une recherche de métadonnées dans le catalogue et récupérer les enregistrements répondants aux critères.



La chaîne de traitement : extraction et archivage

- Script Python 2.7
- Bibliothèques :
 - OWSLib 0.11.2 :
→ Web services OGC
 - Psycopg2 2.6.2
→ BDD PostgreSQL 9.3

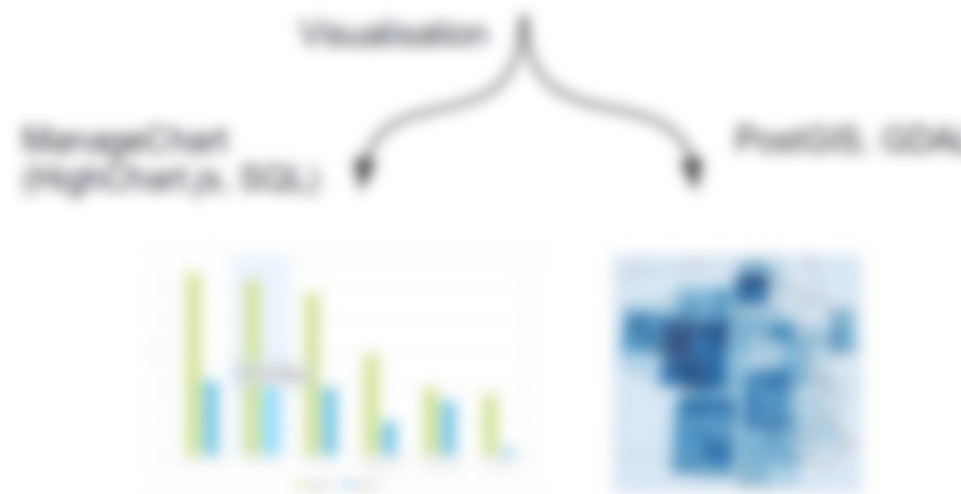
Liste des IDG avec CSW

Atlasante	http://www.atlasante.fr/geonetwork/srv/fre/csw?service=CSW&request=GetCapabilities
CARTOMER	http://cartographie.aieres-marines.fr/geosource/srv/fre/csw?service=CSW&request=GetCapabilities
Geolittoral	http://www.mongeosource.fr/geosource/1111/fre/csw?service=CSW&request=GetCapabilities
Data SHOM	http://services.data.shom.fr/csw/ISOAD?service=CSW&request=GetCapabilities

Python (OWSLib, Psycopg2)



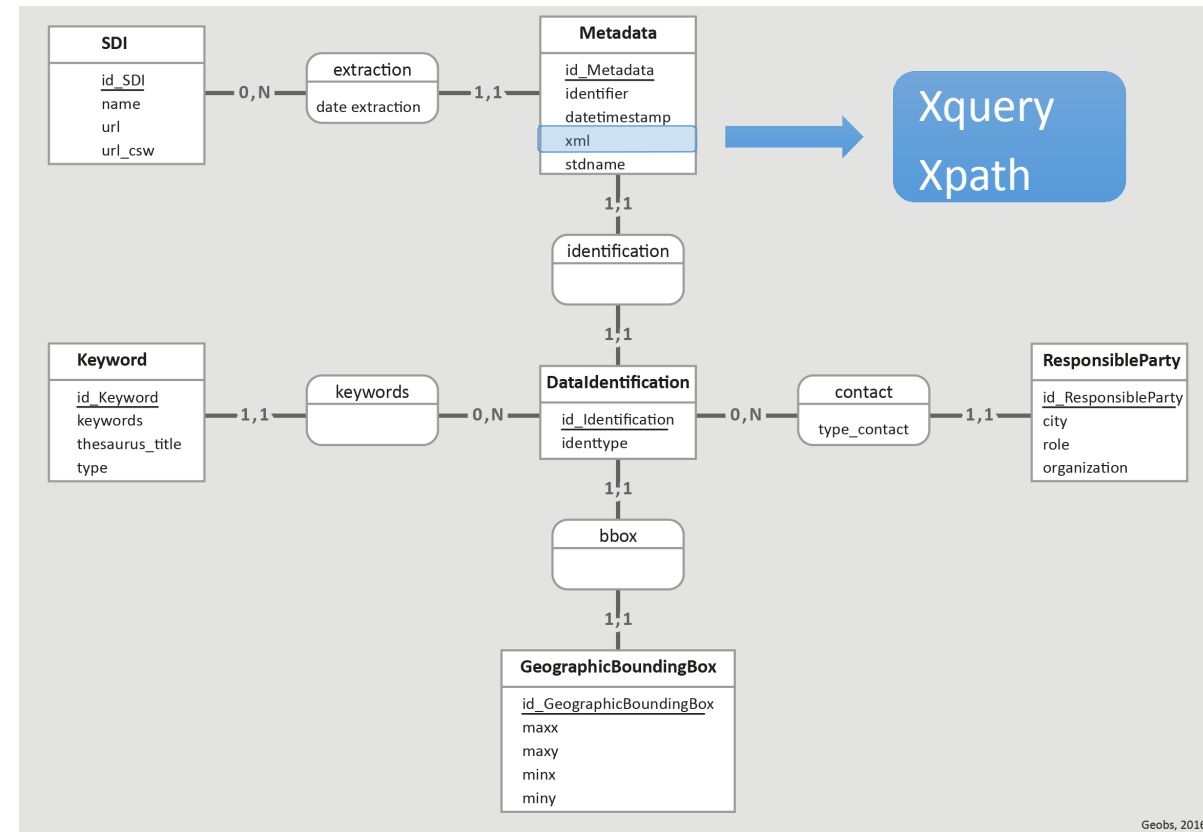
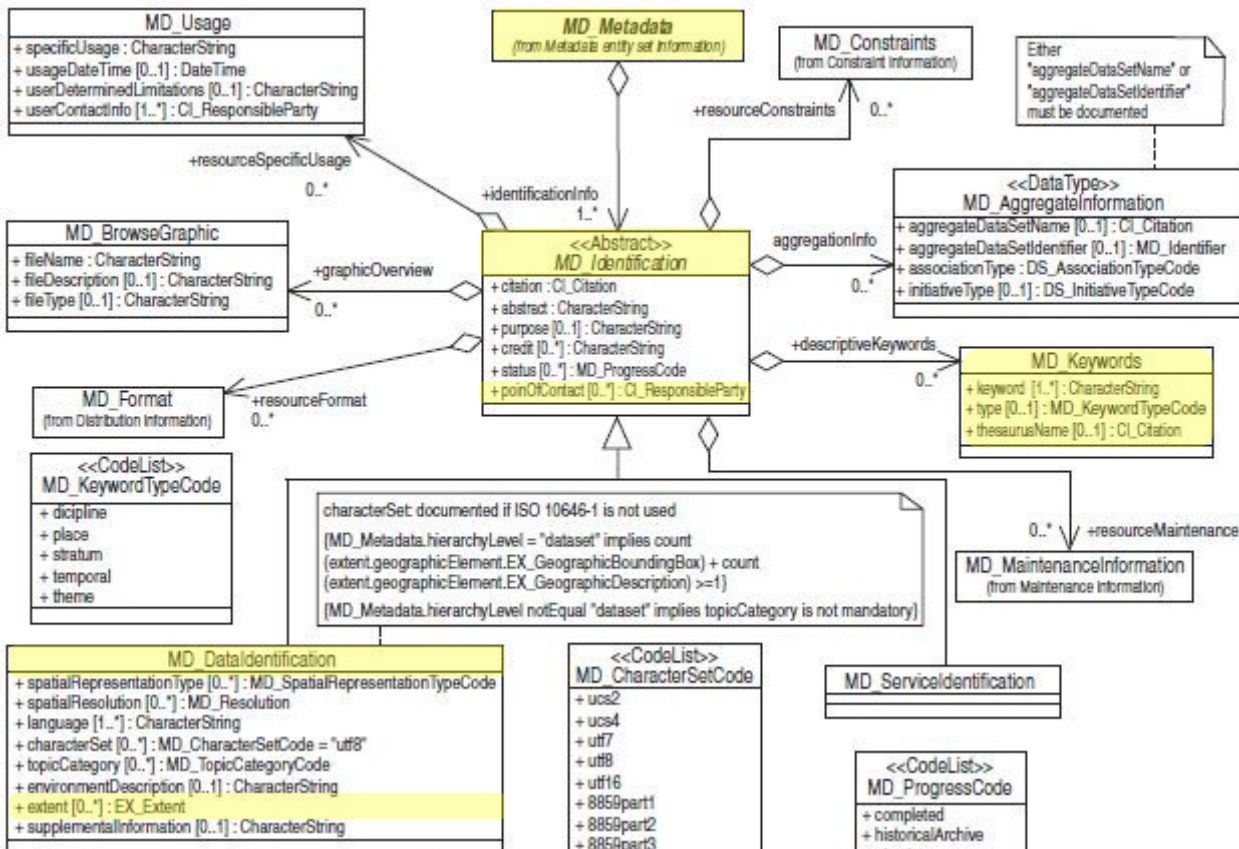
mots clés,
contact,
emprise...



La chaîne de traitement : extraction et archivage

ISO 19115

CSW-HARVESTER



Geobs, 2016.

La chaîne de traitement : analyse et visualisation

- Informations statistiques :
 - ManageChart
 - framework PHP Symfony
 - librairie JavaScript Highcharts.js
- Informations spatiales
 - PostGIS, Python

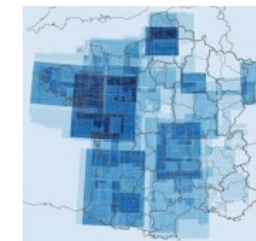
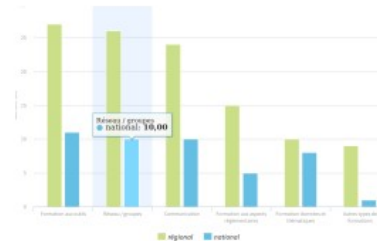


mots clés,
contact,
emprise...

Visualisation

ManageChart
(HighChart.js, SQL)

PostGIS, GDAL



La chaîne de traitement : analyse et visualisation

extraction informations statistiques, exemple de l'accessibilité des données

- Requête :
 - compter le nombre de métadonnées incluant les mots-clefs « données ouvertes » ou « open data » et d'en faire le ratio par rapport au nombre total de métadonnées par IDG
- *Pattern matching* :
 - (keywords LIKE '%donn%e%ouverte%' OR keywords LIKE '%open%data%')

commune ; bassin de vie ; donnée ouverte

donnees ouvertes

données ouvertes

donnée ouverte

donnée ouvertes

données ouvertes

données ouvertes ? en fonction du choix du producteur

données ouvertes, France Métropolitaine

données ouvertes.

données ouvertes;

données ouvertes

La chaîne de traitement : analyse et visualisation

extraction informations statistiques, exemple de l'accessibilité des données

ManageChart Bases de données Requêtes SQL Graphiques Utilisateurs mathias.rouan@univ-brest.fr

Edition

Nom : geobs - accessibilité d

Crédits : GEOBS

Titre : Accessibilité des donn

Uri Crédits : http://www-ium.univ-1

Sous-titre : Filtrage des métadonn

Légende :

Titre axe-X :

Export impression :

Unité axe-X :

Export CSV :

Type axe-X : Catégorie

Inverser les axes :

Axe-Y n° 1 :

Titre : Accessibilité (%)

Type : Linéaire

Série n° 5 :

Titre : IDG régionales (été 21

Unité :

Requête : geobs - test

Paramètre : régional

Type : Colonne verticale

Couleur : #CBDE88

Marqueur :

Style de ligne : Ligne

Accessibilité des données

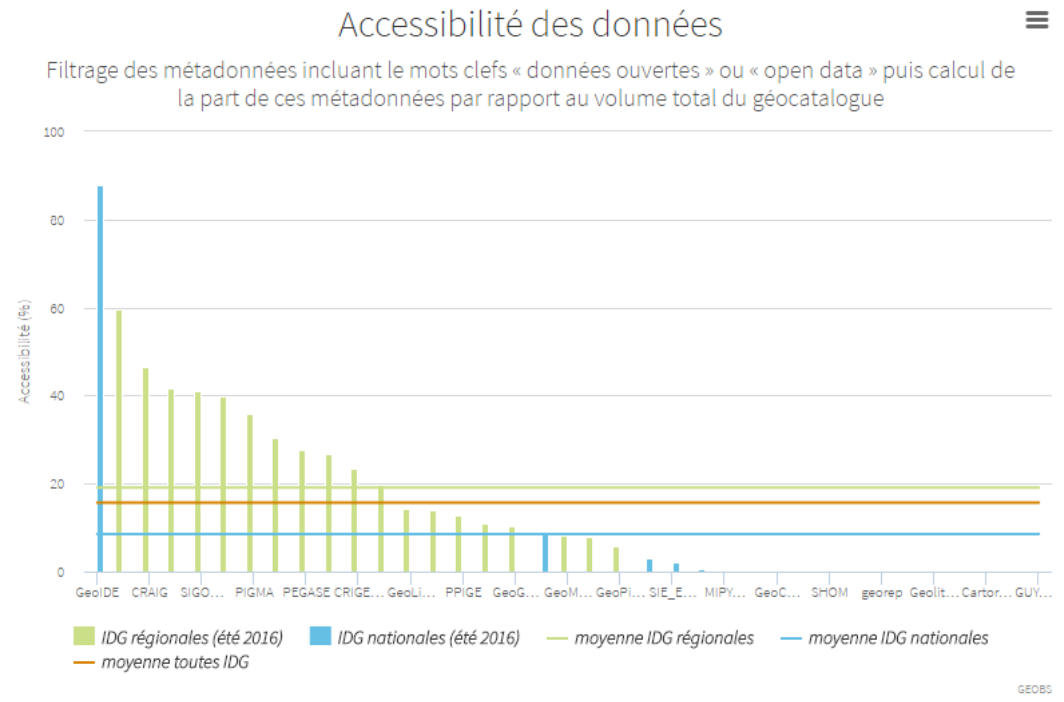
Filtrage des métadonnées incluant le mots clefs « données ouvertes » ou « open data » puis calcul de la part de ces métadonnées par rapport au volume total du géocatalogue

Ajouter un axe-Y **Enregistrer**

Ajouter toutes les séries **Ajouter une série**

Ajouter les séries **Ajouter une série**

Annuler **Annuler**



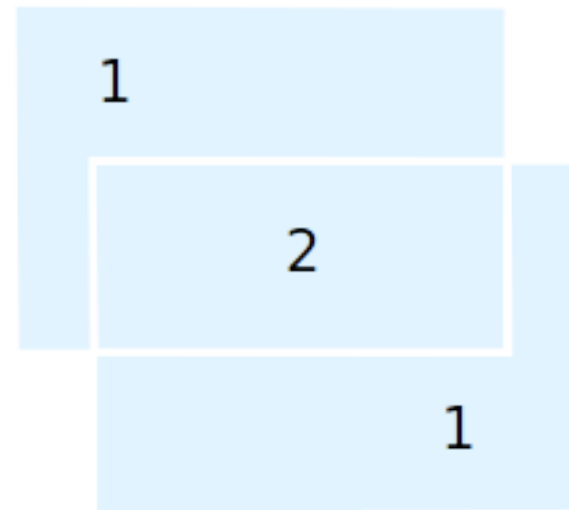
La chaîne de traitement : analyse et visualisation

extraction informations spatiales, exemple du recouvrement des emprises

- Objectif : couche de rectangles → compter le nombre de recouvrements



2 polygones partiellement superposés



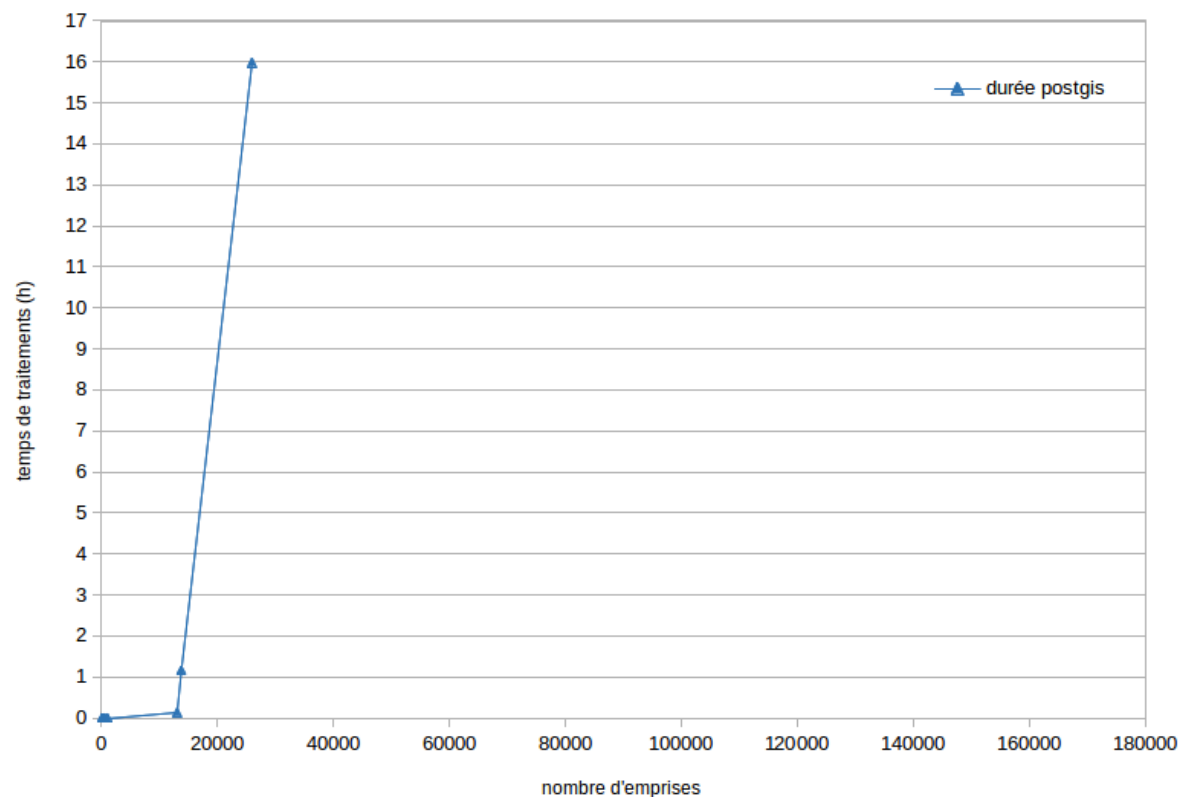
3 polygones sans superposition, avec le nombre de recouvrements en attribut

La chaîne de traitement : analyse et visualisation

extraction informations spatiales, exemple du recouvrement des emprises

1ère solution : requêtes PostGIS

- Simple d'utilisation...
- ... mais temps de calcul trop important quand le nombre d'emprises augmente

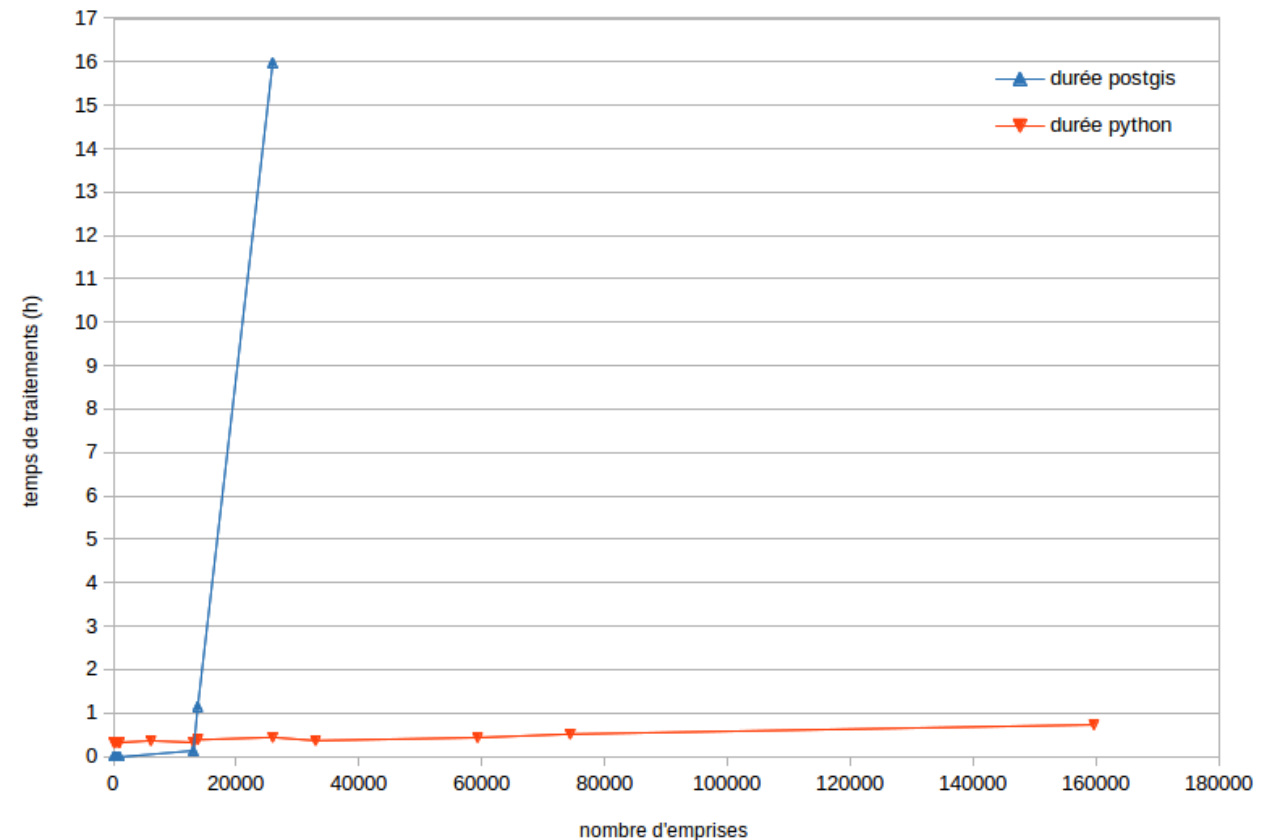


La chaîne de traitement : analyse et visualisation

extraction informations spatiales, exemple du recouvrement des emprises

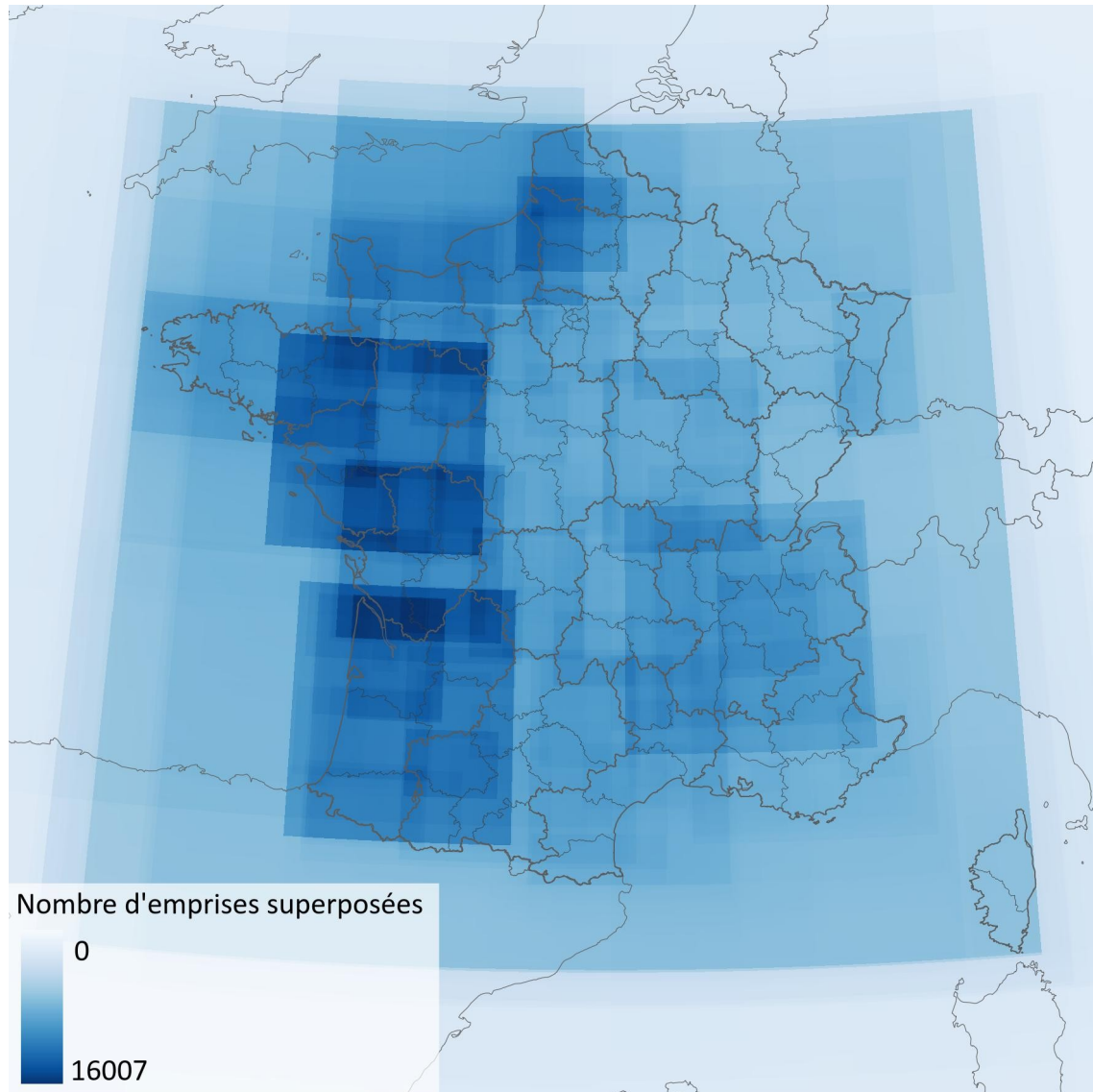
2ème solution : Python

- simple d'utilisation
- nécessite de fixer une résolution
- temps de calcul ok



La chaîne de traitement : analyse et visualisation

extraction informations spatiales, exemple du recouvrement des emprises



Résultat :

- Une carte montrant l'hétérogénéité de la couverture des métadonnées
- Méthode facilement reproductible donc simple à mettre en jeu pour des sous-jeux de données (ex. : métadonnées opendata)

Conclusion

- chaîne testée à l'été 2016 sur 37 IDG → 160603 fiches
- a permis d'alimenter en données et en visualisations exploratoires les quatre axes d'analyse déclinés dans GEOBS :
 1. l'accessibilité des données
 2. la mise en réseau des outils (interopérabilité)
 3. la mise en réseau des acteurs (géocollaboration)
 4. la couverture spatiale des données
- été 2017: relancer l'analyse pour traiter des dynamiques informationnelles de ces IDG et engager ainsi des analyses diachroniques sur ces quatre dimensions



Observer, visualiser et analyser l'information géographique numérique

Projet GEOBS : <https://www-iuem.univ-brest.fr/pops/projects/geobs>
<http://geobs.cnrs.fr/>

csw-harvester : <https://github.com/LETG/csw-harvester>

superpositions d'emprises : <https://github.com/UMR-PASSAGES/metadata-extents>