



HAL
open science

Metadata for the caENTI.

Sylvie Damy, Bénédicte Herrmann-Philippe

► **To cite this version:**

Sylvie Damy, Bénédicte Herrmann-Philippe. Metadata for the caENTI.. In International Conference of Territorial Intelligence, Besançon 2008., Oct 2008, Besançon, France. pp.10. halshs-00516253

HAL Id: halshs-00516253

<https://shs.hal.science/halshs-00516253>

Submitted on 28 Apr 2014

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.

Sylvie Damy

Maître de conférences en Informatique

sylvie.damy@univ-fcomte.fr

+ 33 81 66 20 72

Bénédicte Herrmann-Philippe

Maître de conférences en Informatique

benedicte.herrmann@lifc.univ-fcomte.fr

+ 33 81 66 66 54

Adresse professionnelle

LIFC- UFR Sciences et Techniques
16 Route de Gray- F-25030 Besançon cedex

Summary: In this article, a study to define metadata for the CAENTI resources is presented. The Dublin Core standard is the most know for metadata standards to describe resources. This standard is detailed and illustrated by an example. Then a discussion shows which refinements must be added to the Dublin Core to characterize the CAENTI resources.

Résumé : Dans cet article, une étude visant à définir des métadonnées pour la CAENTI est présentée. Le standard du Dublin Core est le plus connu des standards de métadonnées pour la description de ressources. Ce standard est détaillé et illustré par un exemple. Ensuite une discussion montre quels raffinements doivent être ajoutés au Dublin Core pour caractériser les ressources de la CAENTI.

Keywords: Metadata, Dublin Core, CAENTI

Mots clés : Métadonnées, Dublin Core, CAENTI

1. INTRODUCTION

The Territorial information systems TIS, and the Territorial information systems TICS produce a large editorial workflow (see diag. 1). In this editorial workflow, in one side users and applications must be able of characterizing a document, and in another side when a user obtain after a treatment some data, he needs to know the origin their data. Metadata seems to be a suitable response to these problems. A metadata is a data about data. In the context of the Dublin Core a metadata record consists of a set of attributes, or items, necessary to describe a resource [2].

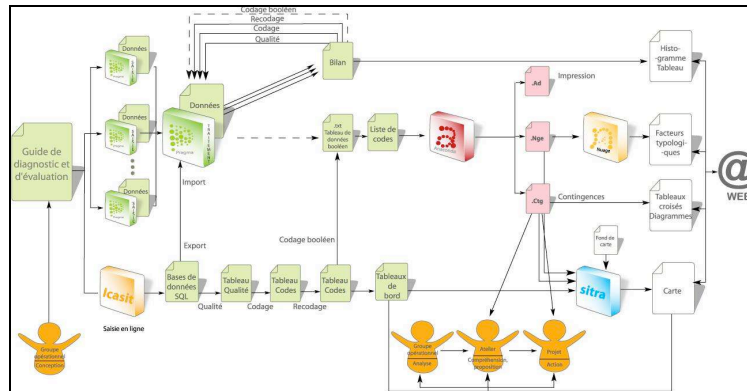


Diagram 1 : The editorial workflow

In a more precise way, metadata allow the characterization of resources in TICS, in indicators portal and in the territorial intelligence portal. Specific metadata needs appear in these three fields.

In TICS and TIS, metadata allow the management of the editorial workflow documents. This management must take account that: documents are computer files, there is a lot of kinds of documents (pictures, maps, ...), documents are changing (we need to have a good traceability of them), and documents are linked to spatial aspects.

In the indicators portal, users want to know which data they use, if these data exist and if they are update. Users also need to have a traceability of data, and elements of data comparison.

With the increase of Internet, and so the increase of the electronic publishing, the interest in metadata standards and practices has exploded, and many projects on metadata appeared. The Dublin Core Metadata Initiative (DCMI) is a major contribution in this domain.

In this article, we present our first studies about the design and the use of metadata in the CAENTI. For that, we describe in the second part, the main metadata standard for numerical resources, the Dublin Core Metadata Initiative and some metadata standards for geographical resources. In the third part, we discuss the creation of a metadata system for the CAENTI. Finally, we develop our perspectives for the construction of this system.

2. METADATA FORMATS

For the CAENTI metadata we study some metadata formats like the Dublin Core standard and some more specific formats which are interested in the description of geographical resources.

2.1 Dublin Core

“The Dublin Core Metadata Initiative is an organization dedicated to promoting the widespread adoption of interoperable metadata standards and developing specialized metadata vocabularies for describing resources that enable more intelligent information discovery systems”[1]. The aim of the Dublin Core Metadata Initiative is to provide simple standards to facilitate the finding, sharing and management of information. It does this by developing and maintaining international standards for describing resources, supporting a worldwide community of users and developers, promoting widespread use of Dublin Core solutions.

2.1.1 Dublin Core elements and qualifiers



The Dublin Core standard includes two levels: Simple and Qualified. The Simple level comprises fifteen elements; the Qualified one includes three additional elements and a group of element refinements or qualifiers. Each element is optional and may be repeated, and the set of elements is not ordered.

Some elements are interested in the contents of the resource, the others in its intellectual property and the last ones in the instance of the resource (Table 1).

Content	Intellectual property	Instanciacion
Title	Creator	Date
Subject	Publisher	Format
Description	Contributor	Identifier
Type	Rights	Language
Source		
Relation		
Coverage		

Table 1 : The different kinds of DC elements

One qualifier refines semantics of an element. There are two classes of qualifiers: the Element Refinement and the Encoding Schemes.

Below, we present the main elements of the Dublin Core. This presentation will be illustrated with an example. This example is this article (the file CaentiMetadata.doc).

The title

This element gives the name by which the resource is formally known.

Example: `Title = "Metadata for the CAENTI"`

It is possible to refine the title with the qualifier `Alternative`. The qualifier `alternative` precises other version of the title. These alternatives can be, for example, title with abbreviations or translation of the title.

Example: `Title.Alternative = "Métadonnées pour la CAENTI"`

For the CAENTI we can choice to have the title in English and the other translations are alternatives.

The subject

The `subject` element presents the topic of the content of the resource. Generally it is expressed as keywords, or key phrases, or classification codes selected from a controlled vocabulary or formal classification scheme.

Example: `Subject= "Dublin Core; Metadata; CAENTI "`

The description

This element is a text describing the content of the resource. Qualifiers of this element are: `tableOfContents`, `Abstract`.

Example: `Description = "In this article ..."`

The four precedent elements characterize the resource in a general way. To characterize in a unique way a resource, the `Identifier` element is used.

The identifier

An `Identifier` is an unambiguous reference to the resource within a given context. A resource is identified by using a string or number conforming to a formal identification system like `URI`, `URL`, `ISBN`, ...

An identifier can be refined by a `bibliographicCitation` which includes sufficient bibliographic detail to identify the resource as unambiguously as possible.

Example: `Identifier =« http:// ... »`



The source

The **source** is a reference to a resource from which the resource is derived. The reference to a resource is an **identifier**. This element and the element **relation**, allow a resource to be linked to its environment that is the other resources.

Example : `Source = " http://dublincore.org/documents/usageguide/elements.shtml"`

The relation

A **relation** element is a reference to a related resource. Like the **source** element, the reference to a resource is an **identifier**.

Example: `Relation = "http:// .../MetadataforTerritorialIndicatorsPortal.ppt"`

Most of the refinements of the element **relation** are reciprocal and may be used to link resources in two directions.

The qualifiers **isVersionOf** and its reverse **HasVersion** describe a resource like a version, edition, or adaptation of a resource.

The qualifiers **isReplacedBy** and **Replaces** indicate that the described resource is supplanted, displaced, or superseded by the referenced resource. When establishing a chain of versions, where only one version is valid, the use of **isReplacedBy** and **Replaces** allows the relationship to be expressed and the user directed to the appropriate version.

The qualifiers **isRequiredBy** and **Requires** indicate that the resource is required by the referenced resource, either physically or logically. In particular these qualifiers describe the relationships between software, applications, hardware and resources.

The qualifiers **isPartOf** and **HasPart** indicate that the described resource is a physical or logical part of a resource. These qualifiers establish "parent/child" relationships.

The qualifiers **isReferencedBy** and **References** specify that the described resource is referenced, cited, by another resource.

The qualifiers **isFormatOf** and **HasFormat** specify that the described resource is the same content of the referenced resource, but presented in another format. And the last qualifier **conformsTo** says that a resource conforms an established standard.

Example : `Relation.isFormatOf = "CaentiMetadata.pdf"Relation.Requires = "Word"`

`Relation.IsPartOf = "Besançon 2008, International Conference of Territorial Intelligence"`

The Coverage

The element **coverage** defines the extent or scope of the content of the resource. Typically it includes spatial location, temporal period or jurisdiction. Two qualifiers are defined for this element.

The qualifier **Spatial** characterizes the intellectual content of the resource, it may include geographic names, latitude/longitude, or georeferenced values, attention to standard schemes and controlled vocabularies.

The qualifier **Temporal** includes the aspects of time related to the intellectual content of the resource and not its lifecycle.

The creator, the publisher and the contributor

A **creator** is an entity (a person, an organization, or a service) primarily responsible for designing the content of the resource. A **publisher** is the entity responsible for making the resource available. A **contributor** is an entity responsible for making contributions to the content of the resource. Example : `Creator = "Damy, Sylvie"`

`Creator = "Herrmann, Bénédicte"`

The rights

The **rights** element gives some information about rights related the resource. It can contain a rights management statement for the resource, or a reference to a service providing such information. The qualifier **accessRights** precises who can access the resource or gives an indication of its security status.

The qualifier **license** refers to a legal document giving official permission to do something with the resource. A license can be identify using a URI.



The Date

The `date` element is a date associated with an event in the life cycle of the resource (creation or availability of the resource). Eight qualifiers are defined for the element `date`. The qualifier `Created` gives the date of creation of the resource, the `Valid` one gives the date of validity of the resource, `Available` gives the date where the resource will become or did become available, `Issued` gives the date of formal publication of the resource, `Modified` gives the date on which the resource was changed, `DateAccepted` gives the date of acceptance of the resource, `DateCopyrighted` gives the date of a statement of copyright and `DateSubmitted` gives the date of submission of the resource. Example : `Date.Created = "2008-06-15"`

The type

This element gives the nature of the content of the resource. For example a resource can be: Image, Sound, Text, These values come from the controlled vocabulary: *DCMI Type Vocabulary*.

For the CAENTI all the resources are numerical resources, and generally of text type or Image (map).

Example : `Type = "text"`

The format

The `format` element is closed to the element `type`. This element describes the file format, physical medium, or dimensions of the resource. Format may be used to determine the software, hardware or other equipment needed to display or operate the resource.

The qualifier `Extent` precises the size or duration of the resource, and the qualifier `Medium` gives the material or physical carrier of the resource.

Example : `Format = "doc"`

The language

This element specifies the language of the content of the resource. The values of the `Language` element are defined by *RFC 3066*.

Example : `Language = "en"`

Others elements have been defined, we just present here the `RightsHolder` one.

The RightsHolder

The `RightsHolder` element specifies a person or organization which owns or manages rights over the resource.



2.1.2 Example in RDF

Here is the description in RDF of this article according to the Dublin Core.

```
<RDF>
<Description xml:lang="en" about="...">
<DC:Title> Metadata for the CAENTI </DC:Title>
<DC:Title.Alternative> Métadonnées pour la CAENTI </DC:Title.Alternative>
<DC:Creator> Damy, Sylvie </DC:Creator>
<DC:Creator> Herrmann, Bénédicte </DC:Creator>
<DC:Subject> Dublin Core; Metadata; CAENTI </DC:Subject>
<DC:Description> In this ... </DC:Description>
<DC:Identifier> http:// ... </DC:Identifier>
<DC:Publisher> CAENTI </DC:Publisher>
<DC>Date.Created DC:Scheme="ISO8601"> 2008-06-15 </DC.Date. Created >
<DC>Type> Text </DC>Type>
<DC:Source> http://dublincore.org/documents/usageguide/elements.shtml </DC:Source>
<DC:Relation> http:// .../MetadataforTerritorialIndicatorsPortal.ppt </DC:Relation>
<DC:Relation.isFormatOf> CaentiMetadata.pdf </DC:Relation.isFormatOf>
<DC:Relation.Requires> Word </DC:Relation.Requires>
<DC:Relation.IsPartOf> Besançon 2008, International Conference of Territorial
Intelligence </DC:Relation.IsPartOf>
<DC:Format> doc </DC:Format>
<DC.Relation DC:SCHEME="URI"> http:// .../MetadataforTerritorialIndicatorsPortal.ppt
</DC.Relation>
<DC:Language> en </DC:Language>
</Description>
</RDF>
```

2.2 Others metadata formats

We just quote here some specific metadata standards related to geography.

ISO 19115 [4] defines a schema for describing geographic resources. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data.

ArcGIS [3] is an integrated family of GIS software products for building a complete GIS. It provides metadata format following the *FGDC CSDGM* or *ISO 19115* standards.

The European INSPIRE project (Infrastructure for Spatial Information in the European Community) [5] works on the design of spatial metadata for the European Community. It takes account of other international standards like Dublin Core or ISO 19115.

3. ELEMENTS FOR THE CREATION OF THE CAENTI METADATA

The Dublin Core standard proposes a set of fifteen elements describing a wide range of numerical resources. With its fifteen elements, the Dublin Core is suitable to describe in a general way all the documents, and all the data of the CAENTI.

3.1 Encoding schemes

In Dublin Core, an encoding scheme is a qualifier which defines the possible values of an element. A value expressed according an encoding scheme either comes from a controlled vocabulary or matches specific format.

For example, Dublin Core recommends the use of *ISO 8601* format for the *date*, the use of *URI* to capture *identifier* or *source*, and the used of controlled vocabularies like *MeSH* or *DDC* to define a *subject* element.

As part of the CAENTI metadata, we have to choice encoding scheme for some elements.

For example, in elements such as *creator*, *publisher* and *contributor* we need to identify entities, which can be persons or organization. Usually to identify these entities we use their name. A *creator* can be one or several entities. So we have to explain how represent one entity and how represent several entities.



To represent one entity, we propose to define a controlled vocabulary composed of set of names. This set contains the names of the CAENTI members and of all persons or organizations in contact with the CAENTI.

To add a new entity in the controlled vocabulary we need to define a format. A person could be described with its name and first name and an organization with its name.

When a `creator` corresponds to several entities we can choice to repeat the `creator` element.

3.2 Definition of new refinements

In some cases Dublin Core elements and qualifiers are not specific enough to express some data in the CAENTI. For some geographical data, to define new qualifiers for the element `coverage`, we can use geographical metadata standards as *ISO 19115*. We must also think about refinements of the element `rights` specific to the CAENTI rights.

3.3 Storage of Metadata

The storage of a metadata record may take two forms. Metadata may be contained in a record separate from the item, or may be embedded in the resource itself. Usually the metadata are embedded in the Web pages, but this is not possible for many kinds of numerical resources (such as *excel* files for example). The storage of metadata must remain transparent to the final users. Users must always see the resources with their metadata.

3.4 Choice of language

Different languages are used to manage and store the Dublin Core metadata: XML, RDF, and HTML. XML and RDF allow the use of schemes and formats. Query languages exist to do request on data stored with XML and RDF. The powers of these query languages are different: RDF query language allow to do more requests.

The choice of a language leads to the repetition of the declarations of an element when there are several values for this element, or the definition of a unique declaration with all the values in it. The choice of the tool can also have influence on these repetitions.

3.5 Tools

Actually some tools around the Dublin Core metadata exist and they are closely related to the Web. For example, DC-Dot [6] automatically creates Web Pages metadata. Other tools manage metadata for specific resources such as photography [8] or articles [7] to publish them on the Web.

For manage the CAENTI metadata we will need tools which allow the creation and the modification of metadata scheme, the input of metadata values, the web publication, the automatic extraction from some kinds of resources and the search of resources.

3.6 Link between metadata and resources

The Dublin Core metadata characterize numerical resources. But they are not related to the data contained in the numerical resources. In the CAENTI, we need to define metadata for specific data like territorial indicators.

4. CONCLUSION

In this article we have presented the Dublin Core metadata standard. With some new refinements, this standard can meet the CAENTI's needs.

The first job will be to identify all the CAENTI kinds of resources in collaboration with the actors of the CAENTI.

Then for each kind of resources, it will be necessary to characterize the useful elements and the refinements. In particular, we'll study the metadata geographical standards in depth.

After the definition of the metadata, we will work to define tools to manage the CAENTI metadata.

Web references

[1] "Dublin Core Metadata Initiative", Official site of Dublin Core, <http://dublincore.org/about/>

[2] "Using Dublin Core ", Diane Hillman, 2005, <http://dublincore.org/documents/2005/11/07/usageguide/>



- [3] “ESRI Site”, Official site of GIS and mapping software, <http://www.esri.com/>
- [4] “ISO norme Site “, http://www.iso.org/iso/iso_catalogue
- [5] “INSPIRE Project”, Site of the European INSPIRE directive, <http://inspire.jrc.ec.europa.eu/>
- [6] “DC-dot”, <http://www.ukoln.ac.uk/metadata/dcdot/> dc-dot
- [7] “Mathematics Metadata Markup Editor”, <http://www.mathematik.uni-osnabrueck.de/cgi-bin/MMM3.2.cgi>
- [8] “Photo RDF-Gen”, http://www.webposible.com/utilidades/photo_rdf_generator_en.html

