

# Metadata For the CAENTI

---

Sylvie Damy – Bénédicte Herrmann

[sylvie.damy@univ-fcomte.fr](mailto:sylvie.damy@univ-fcomte.fr), [benedicte.herrmann@lifc.univ-fcomte.fr](mailto:benedicte.herrmann@lifc.univ-fcomte.fr)

LIFC - UFC

Besançon 2008 - International Conference of Territorial  
Intelligence

October, 16th - 17th 2008

# Objectives

---

## Characterization of resources

- in TIS, TICS (editorial workflow)
- in indicators portal
- in the territorial intelligence portal

⇒ Metadata

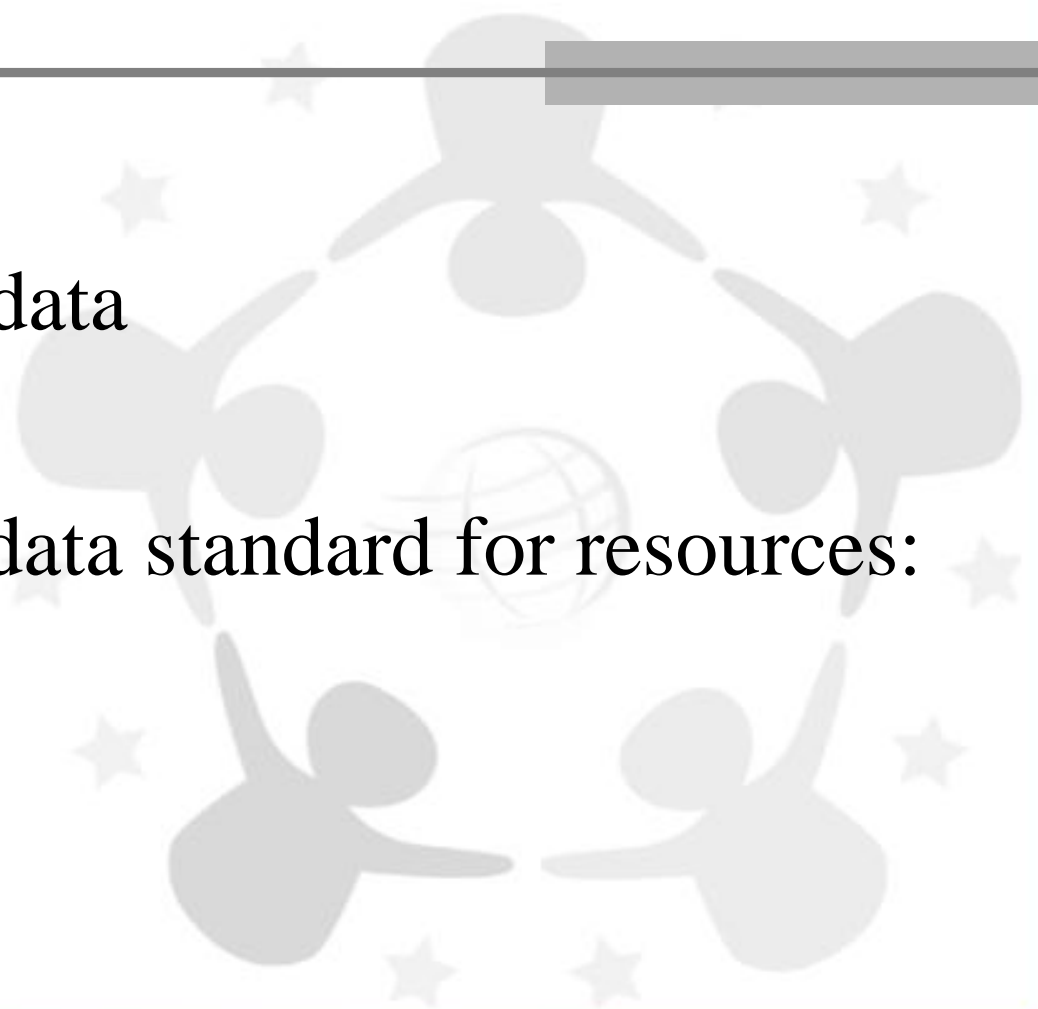
# Outline

---

- Presentation of Dublin Core Metadata standard
  - Presentation of Metadata geographical standard
  - Creation of CAENTI Metadata
- 

# What is a metadata ?

---

- Is a data about data
  - The main metadata standard for resources:  
Dublin Core
- 

# Dublin core

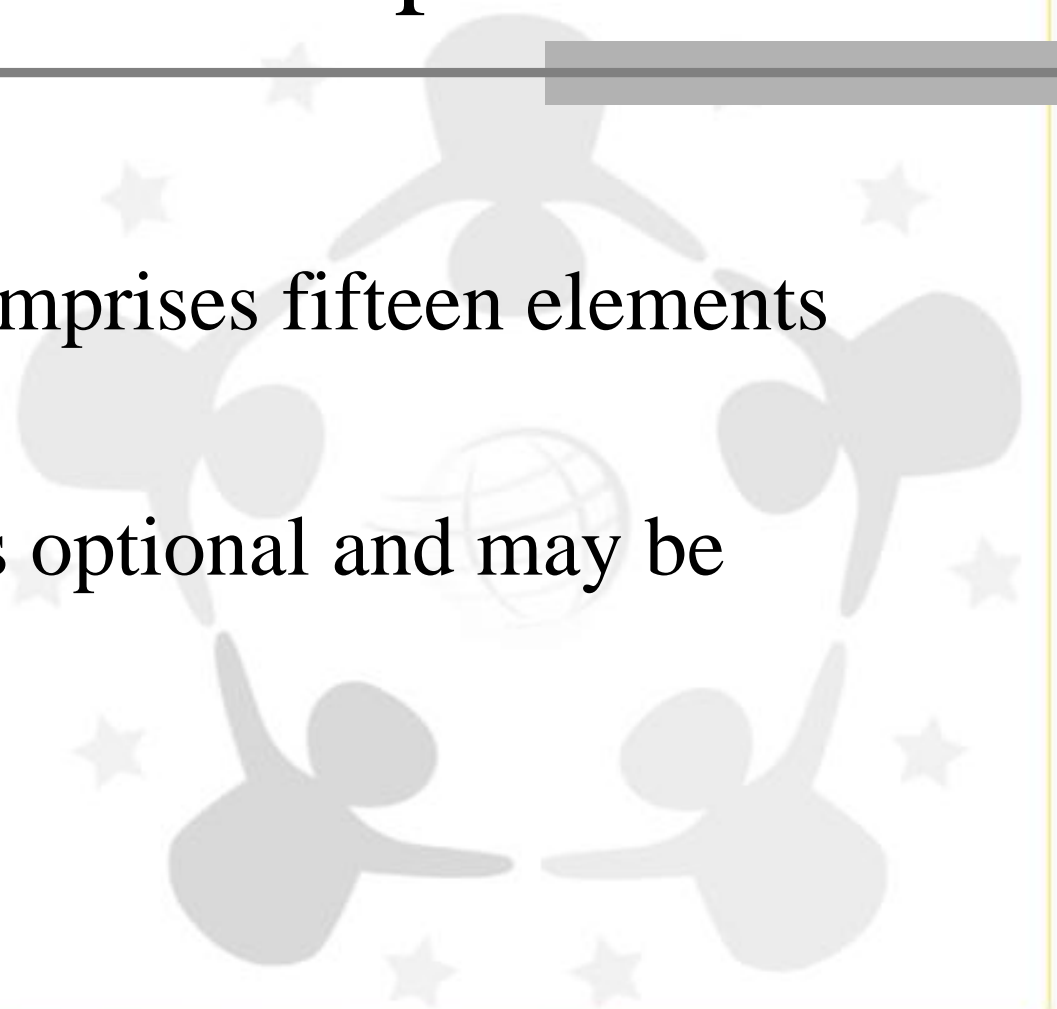
---

Facilitate the finding, the sharing and the management of information

⇒ Two levels : simple and qualifier

# Level : Simple

---

- Dublin Core comprises fifteen elements
  - Each element is optional and may be repeated
- 

# Level: Qualifier

---

Refines the semantics of the elements in ways that may be useful in resource discovery:

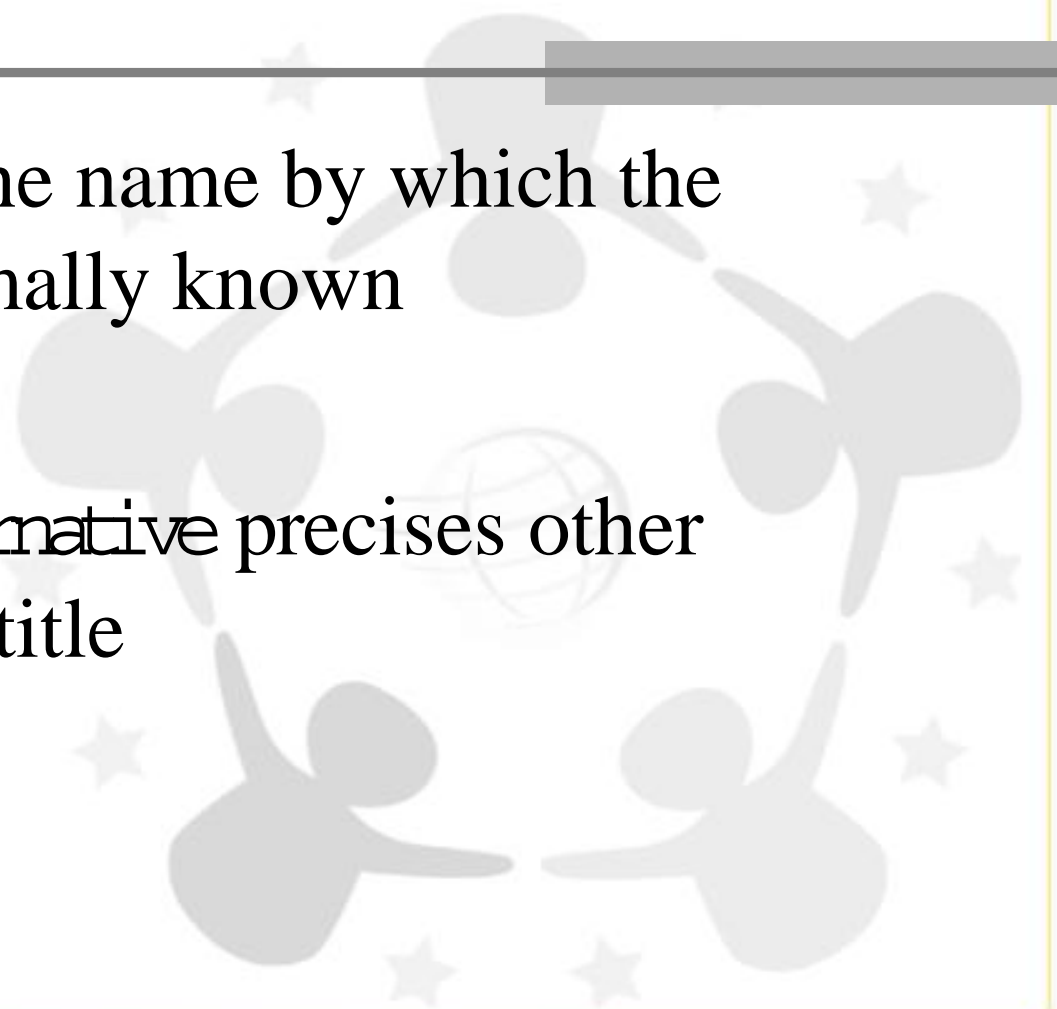
- **Element Refinement:** makes the meaning of an element more specific
- **Encoding Scheme:** describes the possible values of an element: controlled vocabularies, formal notations, parsing rules

# The elements

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

# Title

---

- *Description*: The name by which the resource is formally known
  - *Qualifier*: Alternative precises other versions of the title
- 

# Subject

---

*Description:* The topic of the content of the resource (expressed as keywords or key phrases or classification codes)

A faint background graphic featuring four stylized human figures in a circle, holding hands, with a globe in the center and several stars scattered around. The figures are in shades of gray and white.

# Description

---

*Description:* An account of the content of the resource

*Qualifier:* TableOfContents

A list of subunits of the content of the resource

*Qualifier:* Abstract

A summary of the content of the resource

# Identifier

---

The four precedent elements characterize the resource in a general way

⇒ To characterize in a unique way a resource the Identifier element is used

# Identifier

---

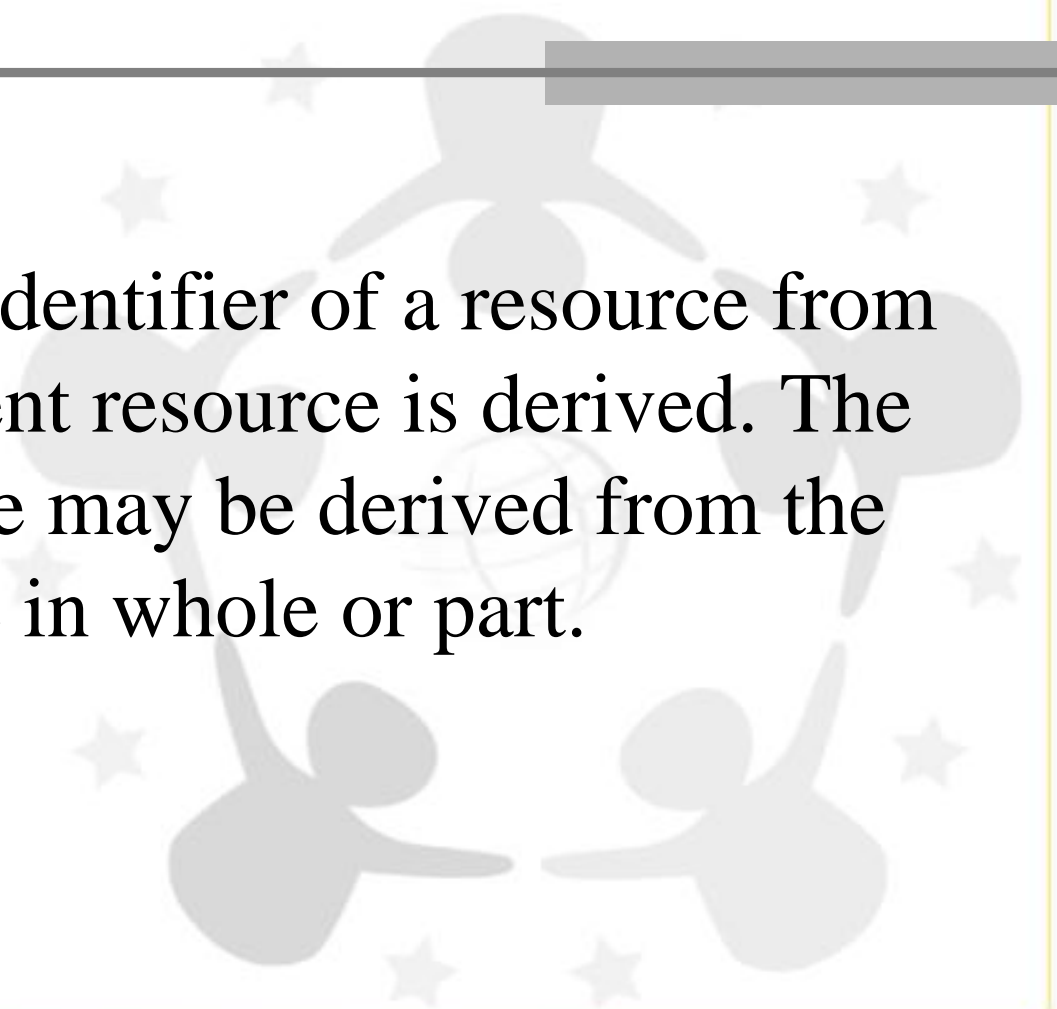
*Description:* An unambiguous reference to the resource within a given context: a string or number conforming to a formal identification system (URI, URL, DOI, ISBN)

*Qualifier:* bibliographicCitation is a bibliographic reference for the resource

# Source

---

*Description:* An identifier of a resource from which the present resource is derived. The present resource may be derived from the source resource in whole or part.

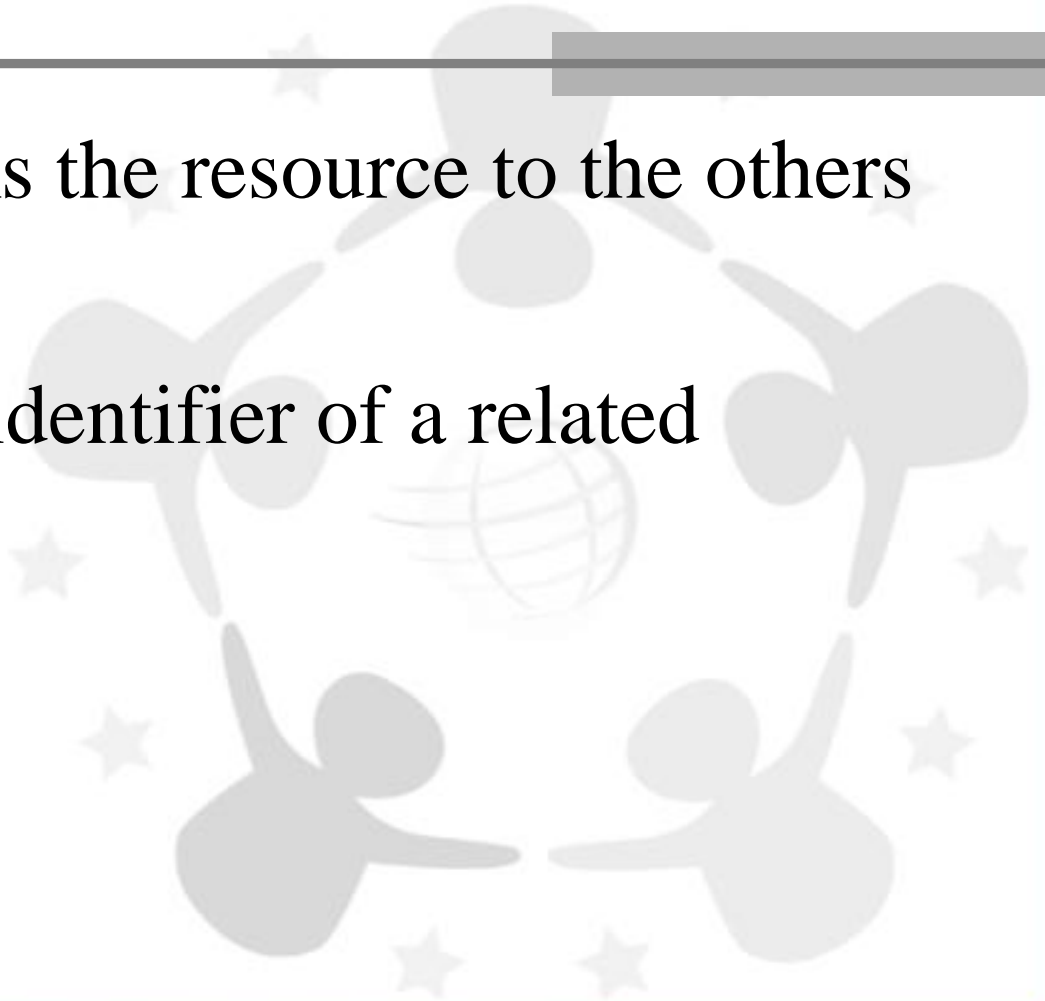


# Relation

---

This element links the resource to the others

*Description:* An identifier of a related resource



# Relation – Qual.

---

Most of the refinements of Relation are expressed as "reciprocals" and may be used to link resources in two directions

isVersionOf - hasVersion

isReplacedBy - Replaces

isRequiredBy - Requires

isPartOf - hasPart

isReferencedBy - References

isFormatOf - hasFormat

conformsTo

# Coverage

---

*Description:* The extent or scope of the content of the resource

*Qualified:* Spatial

Spatial characteristics of the content of the resource:  
geographic names, latitude/longitude, or other established georeferenced values attention to standard schemes and controlled vocabularies

*Qualified:* Temporal

Temporal characteristics of the content of the resource.

# Creator-Publisher-Contributor

---

are entities (a person, an organization, or a service) responsible for:

- designing the content of the resource: creator
- making the resource available: publisher
- making contributions to the content of the resource: contributor

# Rights

---

*Description:* Information about rights held in and over the resource

*Qualifier:* accessRights

Information about who can access the resource or an indication of its security status.

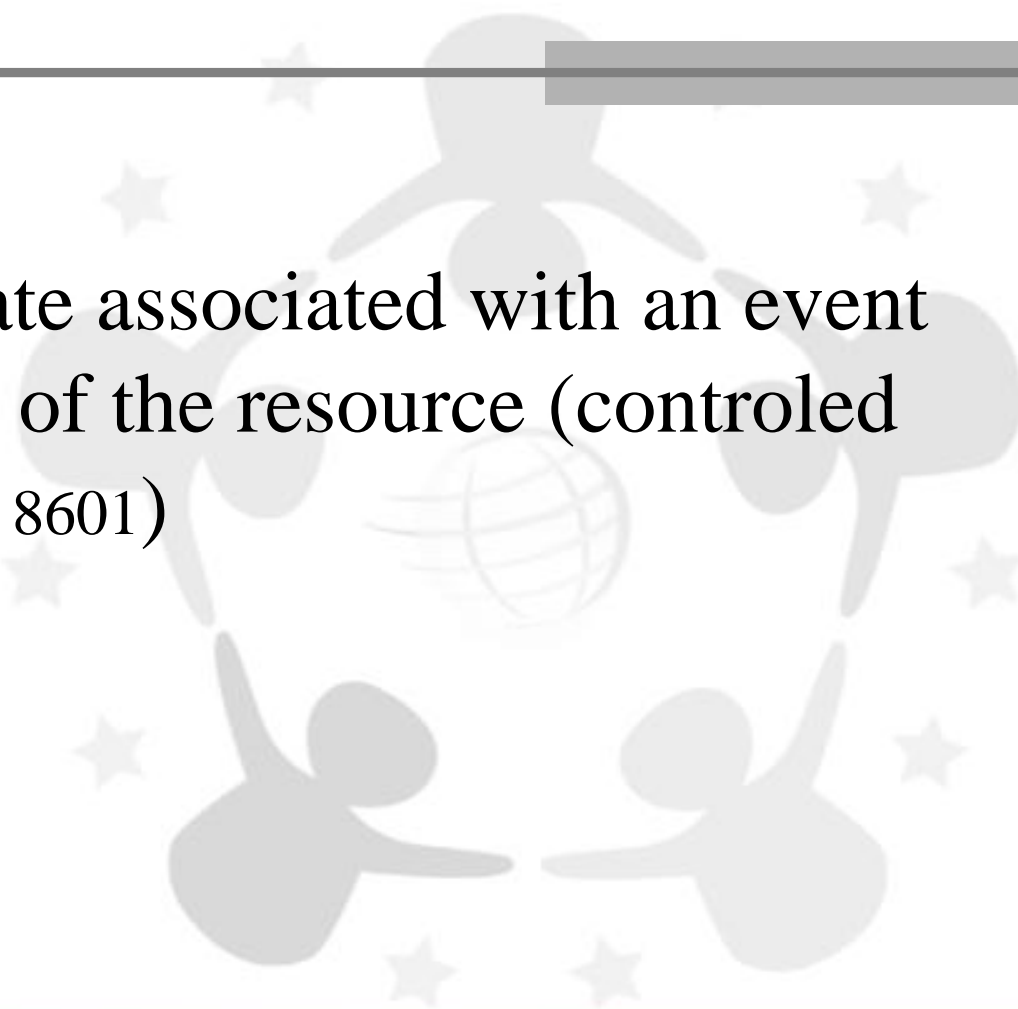
*Qualifier:* license

A legal document giving official permission to do something with the resource. Recommended best practice is to identify the license using a URI.

# Date

---

*Description:* A date associated with an event in the life cycle of the resource (controlled vocabulary: ISO 8601)



# Date – Qual.

---

Created

Valid

Available

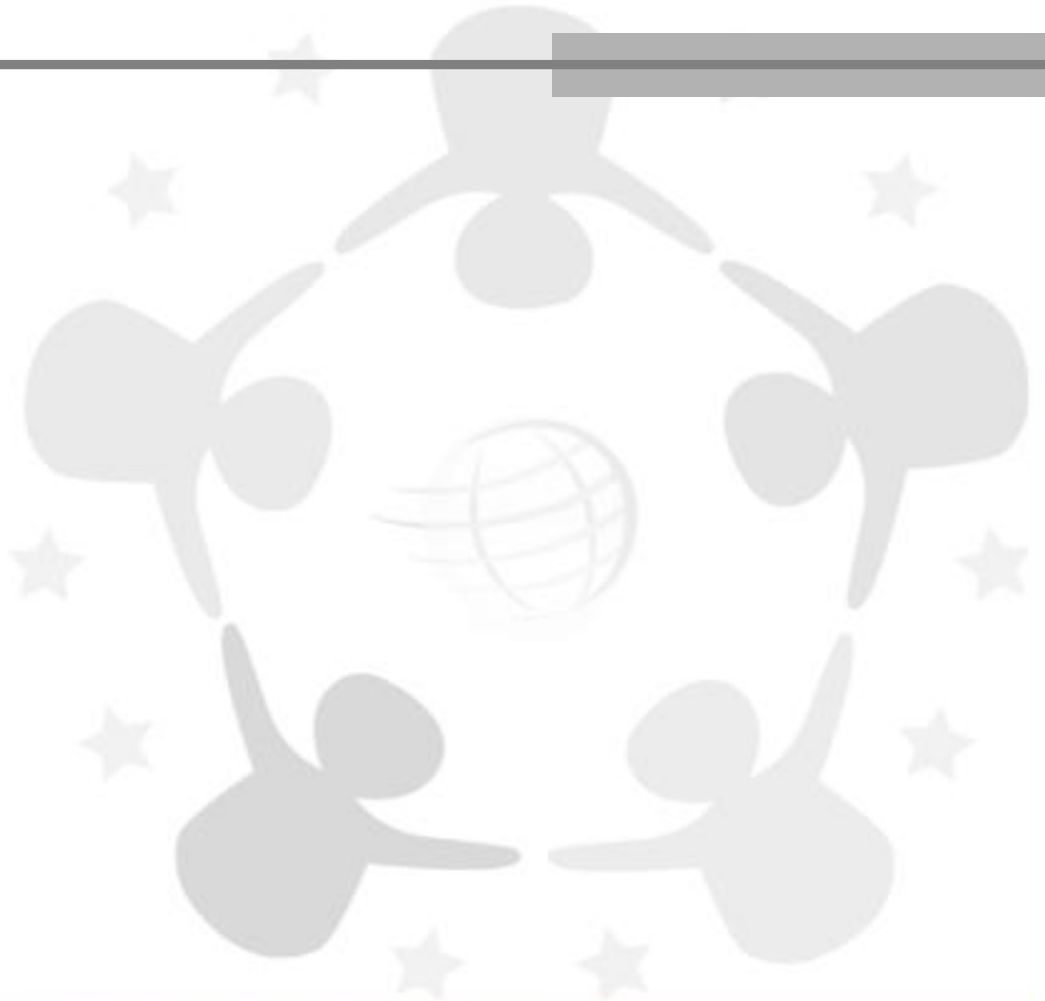
Issued

Modified

DateAccepted

DateCopyrighted

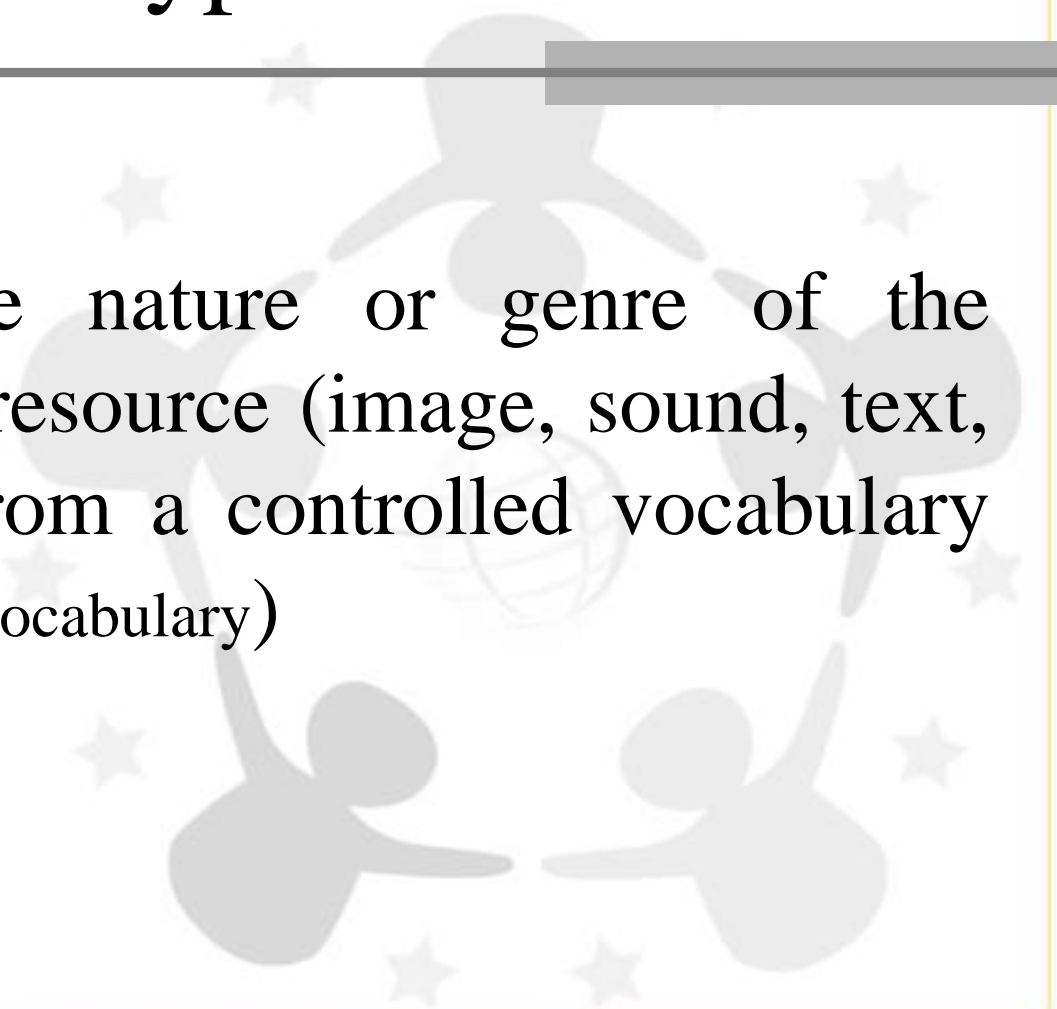
DateSubmitted



# Type

---

*Description:* The nature or genre of the content of the resource (image, sound, text, etc) selected from a controlled vocabulary (ex: DDCMI Type Vocabulary)



# Format

---

*Description:* more precise than the element type. Describes the file format of the resource. Format may be used to determine the software, hardware or other equipment needed to display or operate the resource.

*Qualifier Extent* precises the size or duration of the resource

*Qualifier Medium* gives the material or physical carrier of the resource.

# Language

---

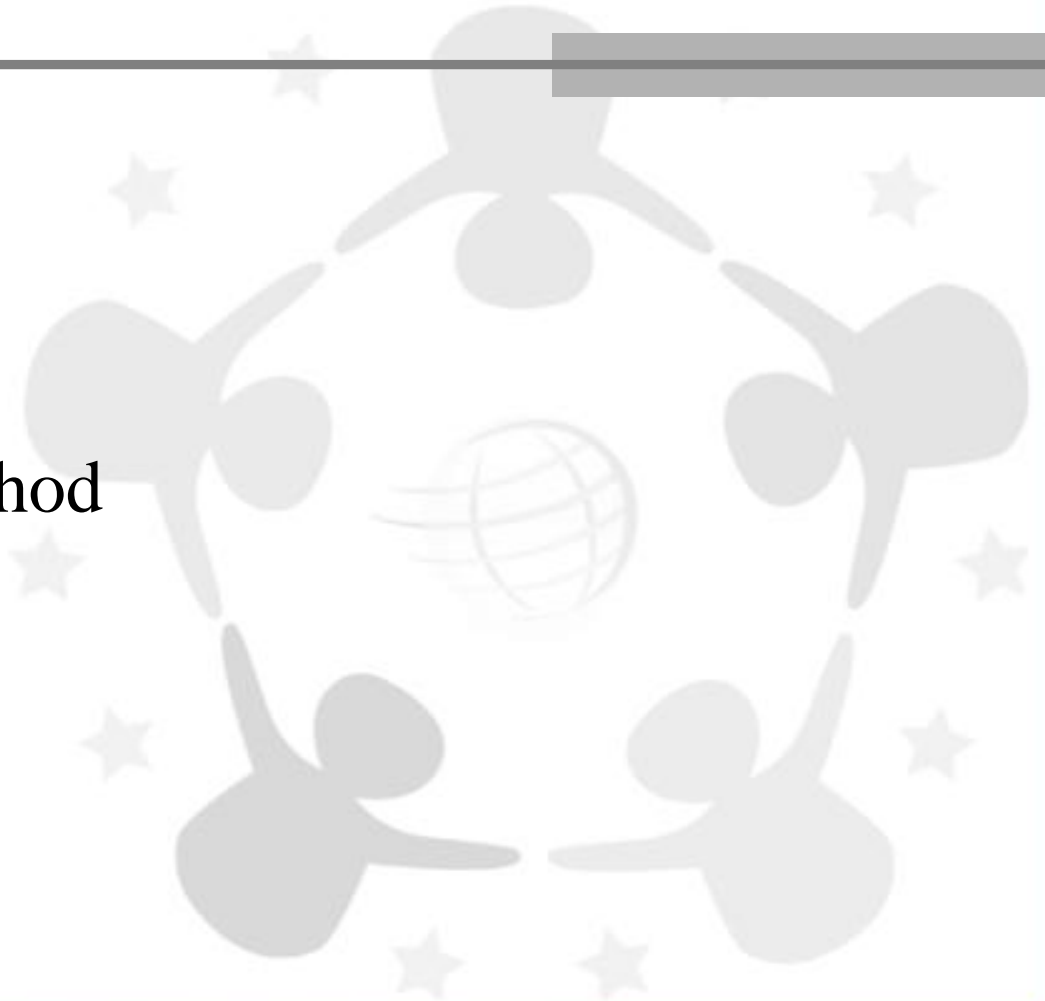
*Description:* language of the intellectual content of the resource (RFC 3066)



# Other elements

---

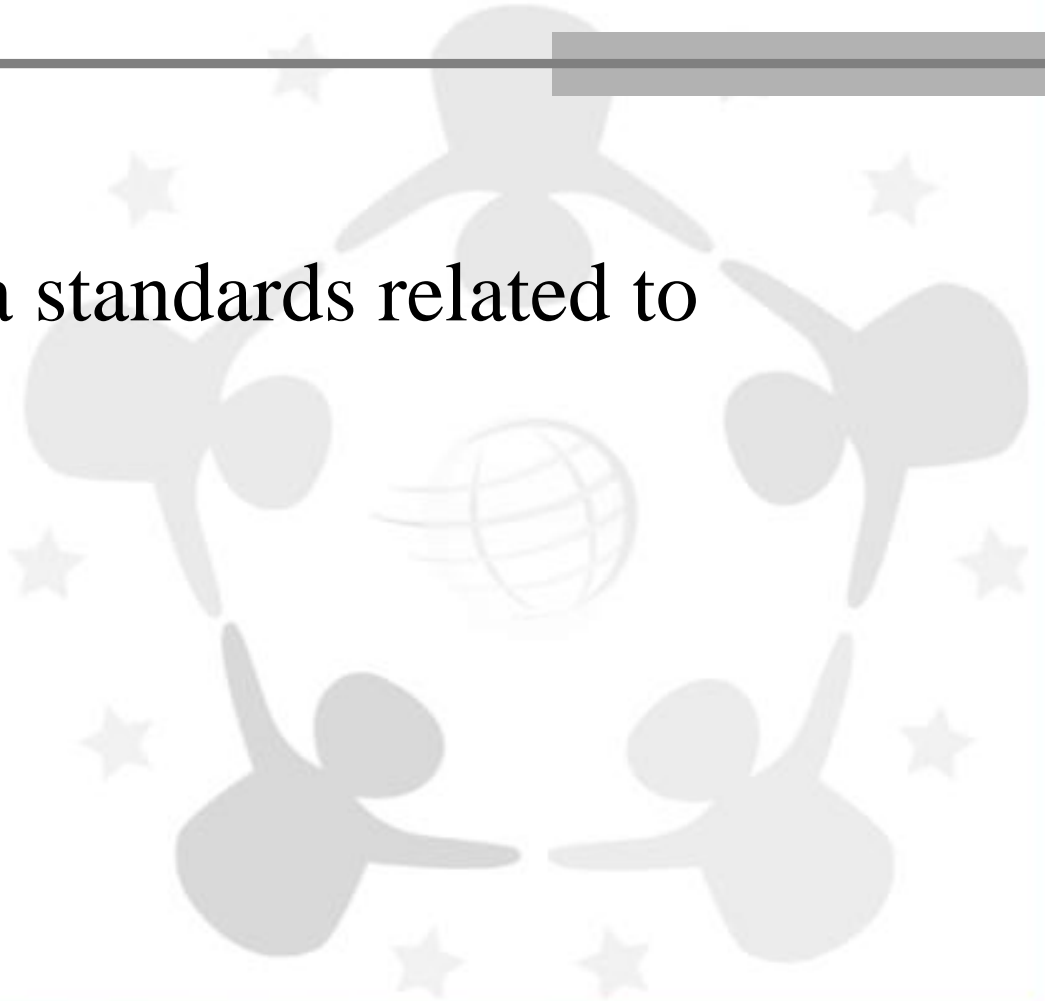
- Audience
- Provenance
- RightsHolder
- Instructional Method
- Acrual Method
- Periodicity
- Policy



# Others metadata formats

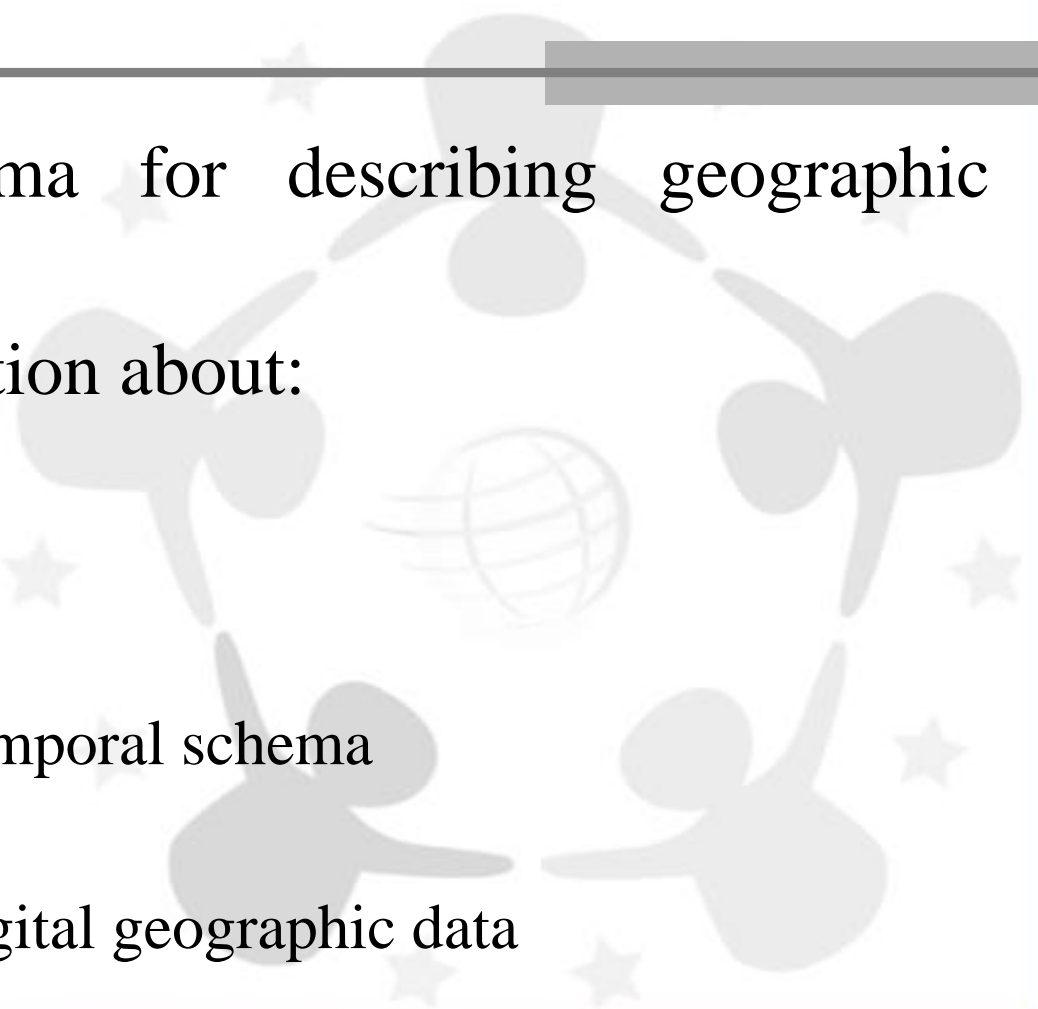
---

Specific metadata standards related to  
geography



# ISO 19115

---

- Defines a schema for describing geographic resources
  - Provides information about:
    - the identification
    - the extent
    - the quality
    - the spatial and temporal schema
    - spatial reference
    - distribution of digital geographic data
- 

# ArcGIS

---

- 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
- works on the design of spatial metadata for the European Community
- takes account of other international standards like Dublin Core or ISO 19115

# Creation of the CAENTI Metadata

---

Dublin Core is suitable to describe in a general way all the documents, and all the data of the CAENTI

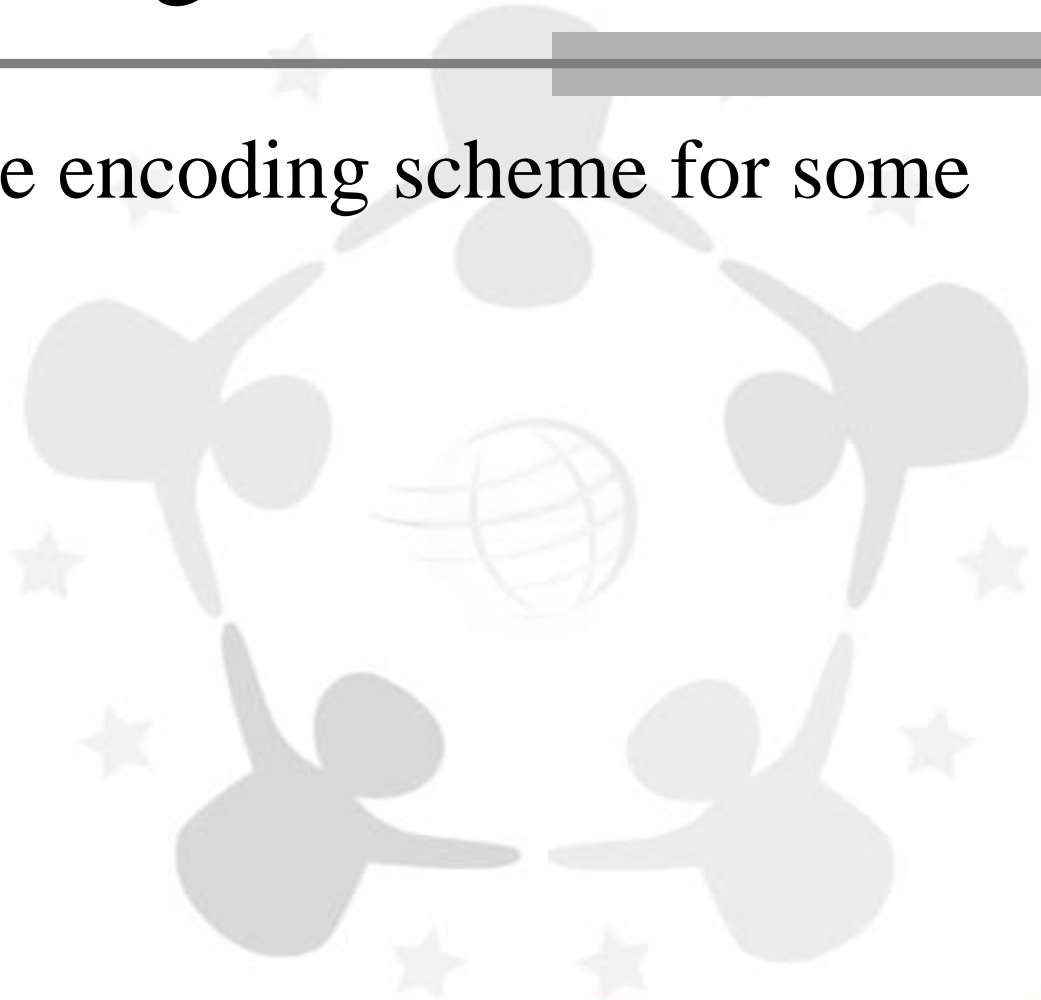
⇒ extend Dublin Core to the specificities of the CAENTI

# Encoding schemes

---

We have to choice encoding scheme for some elements:

- Date
- Identifier
- Subject ...



# New refinements

---

In some cases Dublin Core elements and qualifiers are not specific enough to express some data in the CAENTI:

- Geographical data (coverage)
- CAENTI rights (rights)

# Storage of Metadata

---

- May take two forms. Metadata may be:
  - Outside the resource (for many kinds of numerical resources such as *excel* files for example)
  - In the resource (in the Web pages for example)
- Must remain transparent to the final users: users must always see the resources with their metadata.

# 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 but RDF query language allows to do more requests

# Tools

---

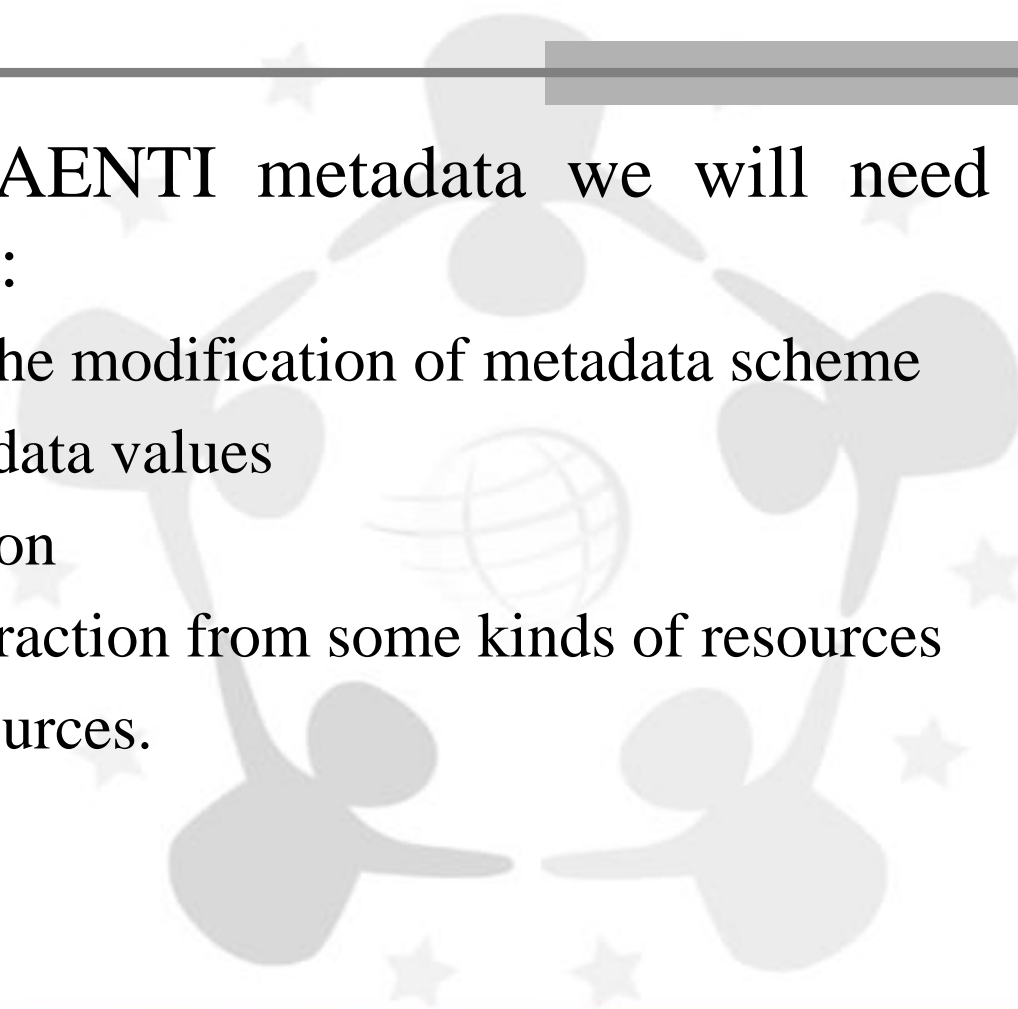
Actually some tools around the Dublin Core metadata exist and they are closely related to the Web:

- DC-Dot automatically creates Web Pages metadata.
- Other tools manage metadata for specific resources such as photography (Photo RDF-Gen) or articles to publish them on the Web (Mathematics Metadata Markup Editor).

# Tools

---

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
  - the search of resources.
- 

# Link between metadata and resources

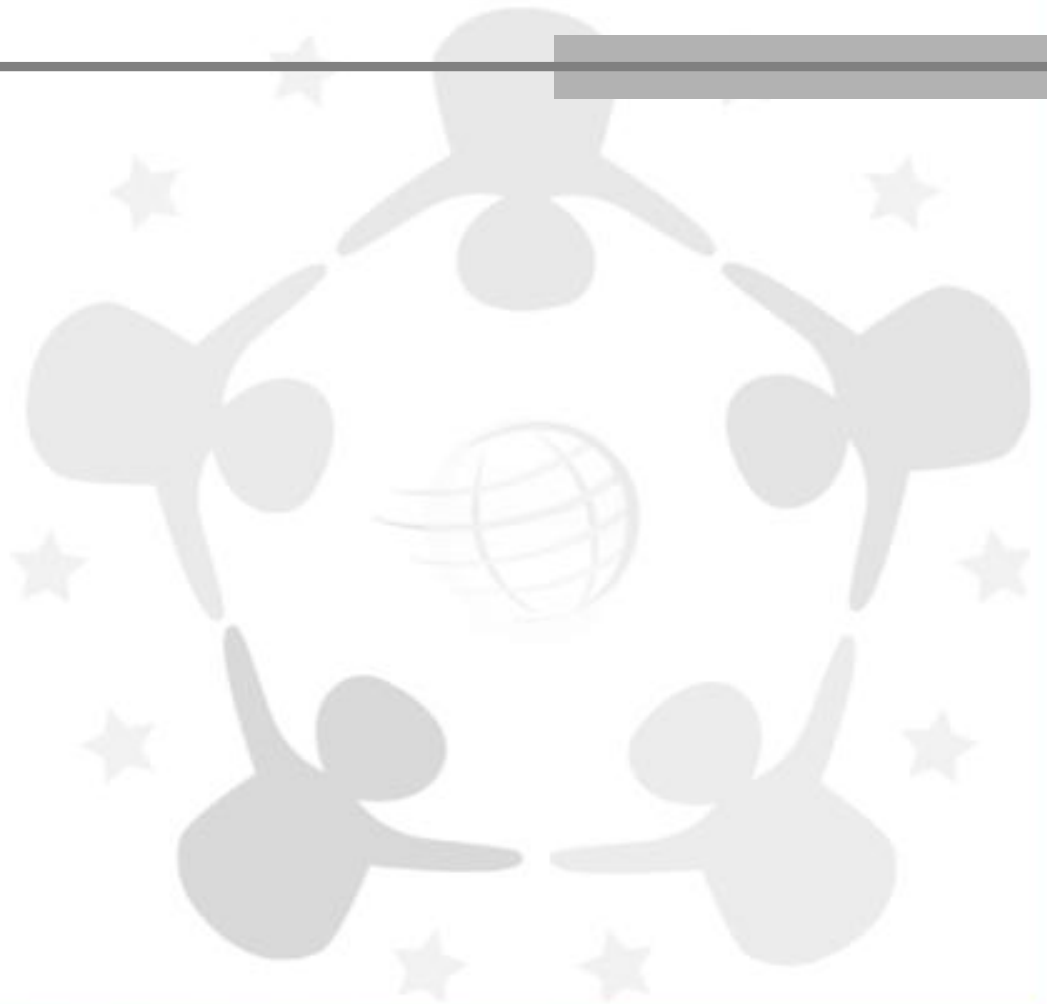
---

- 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.

# Conclusion

---

- With some new refinements, Dublin Core can meet the CAENTI's needs.
- To do:
  - for each kind of CAENTI resources, characterize the useful elements and the refinements
  - define or choose tools to manage the CAENTI metadata.



# Relation – Qual.

---

Most of the refinements of Relation are expressed as "reciprocals" and may be used to link resources in two directions

## **isVersionOf - hasVersion**

The described resource is a version, edition, or adaptation of the referenced resource. Changes in version imply substantive changes in content rather than differences in format.

## **isReplacedBy - Replaces**

The described resource is supplanted, displaced, or superseded by the referenced resource.

# Relation – Qual.

---

## **isRequiredBy - Requires**

The described resource is required by the referenced resource, either physically or logically. This relationship is most often seen in relationships between software and documents or applications and hardware and/or software requirements.

## **isPartOf - hasPart**

The described resource is a physical or logical part of the referenced resource.

# Relation – Qual.

---

## **isReferencedBy - References**

The described resource is referenced, cited, or otherwise pointed to by the referenced resource

## **isFormatOf - hasFormat**

The described resource is the same intellectual content of the referenced resource, but presented in another format

## **conformsTo**

A reference to an established standard to which the resource conforms

# Date – Qual.

**Created** : Date of creation of the resource.

**Valid** : Date (often a range) of validity of a resource.

**Available** : Date (often a range) that the resource will become or did become available.

**Issued** : Date of formal publication of the resource.

**Modified** : Date on which the resource was changed.

**DateAccepted** : Date of acceptance of the resource (e.g. of thesis by university department, of article by journal, etc.).

**DateCopyrighted** : Date of a statement of copyright.

**DateSubmitted** : Date of submission of the resource (e.g. thesis, articles, etc.).