Domain Engineering in GIS: Metadata Standard Interoperability

Anne Elise Weiss

April 18, 2012

Introduction

- Earlier in the course: Discussed a system storing information about different types of things (concerts, restaurants, train schedules, etc.)
- Would mean: Combining information from many different domains, utilizing it in the GIS domain
- Need: metadata interoperability

Metadata

Different definition proposals (by ISO):

- Data describing and documenting data
- Data about datasets and usage aspects of it
- Data about the content, quality, condition, and other characteristics of data

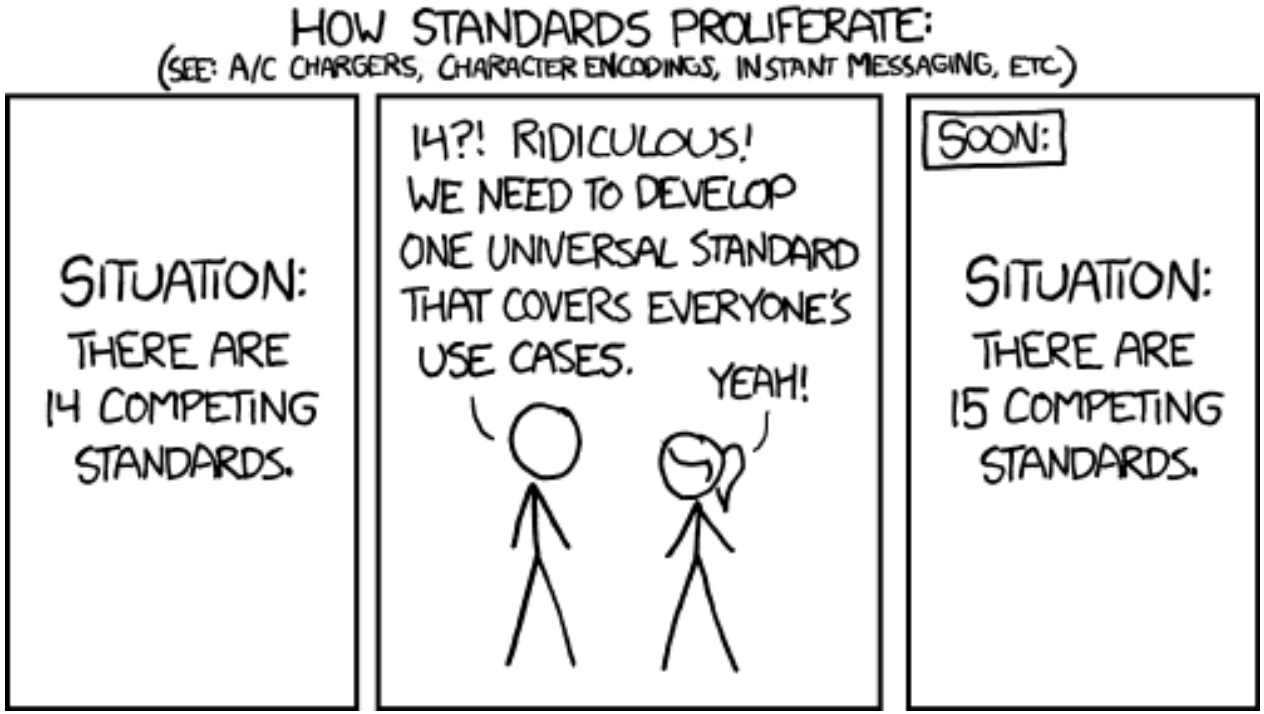
In a GIS context: Metadata defined for geographical information mainly has a documenting role. The information which is given to the user is related to the dataset structure and to its contents.

Interoperability

- Interoperability: The ability to develop conventions that enable data exchange and integration.
  - Semantic interoperability: agreement about content description standards.
- Growing number and complexity of metadata standards $\rightarrow$ more difficult and tedious to handle information in different standards.

Metadata standards

- Ideally: one unique metadata standard



xkcd.com/927

Metadata standards

- One possible solution: Create a mechanism enabling translation in order to make it conform to different standards – *crosswalks*.

Example of a cross-domain application

- We have three different databases store metadata of different types:
  - Library items (books, reports, other documents)
  - Events (movies, theatre, recitals, etc.)
  - Geographic data (maps, satellite images, etc.)
- Together, they can be used for e.g. tourist information
- The standard used in each metadata database belongs to a distinct domain, so they are different standards.

Example of a cross-domain application

- The tourist information provider system should use a homogeneous mechanism for querying and accessing the databases.
- I.e: the metadata schema of the system should be independent of the metadata representation used by the different databases.
- The system uses one standard for querying and managing information, e.g. Dublin Core.
- The homogeneous mechanism acts as a crosswalk broker
  - A repository of crosswalks
  - The software for activating and processing the crosswalks when needed

Example of a cross-domain application

- A crosswalk specifies the mapping between two related standards
- Problems with crosswalks:
  - the construction is difficult and error-prone
  - the construction requires deep knowledge and lots of experience
  - the standards have mostly been developed independently
  - maintenance
- Therefore: harmonization is vital, makes it easier to match the metadata elements

Domain Engineering and metadata crosswalks

The problems associated with developing metadata crosswalks are not constrained to a specific application domain.

Geographic information metadata concepts:

- Geographic information: describes phenomena associated directly or indirectly with a location
- The geographic metadata: describes the content, quality, condition and other characteristics of the data that allow a person to locate data and to understand them
- There is a range of different existing standards

Different approaches

Main approaches to the semantic interoperability problem:

1. Solutions using ontologies
2. Create specific crosswalks for one-to-one mapping

Ontology-based semantic interoperability

- There exist many initiatives that aim towards solving the semantic interoperability problem on the Web
- Most of these propose using ontologies and
  - RDF (Resource Description Framework): simple model for describing the interrelationships among resources – named properties and values
  - RDFS (RDF Schema/RDF Vocabulary Language) for the declaration and interpretation of those properties
- Closely related to the Semantic Web

Ontology-based semantic interoperability

- There exist alternatives to RDF technologies
- These approaches offer flexible solutions for interoperability.
- Ambitious aim of flexibility $\rightarrow$ lack of accuracy in the performed mappings?
- No local structural constraints considered

Ontology-based semantic interoperability

- "the wider the targeted scope of interoperability, the more difficult it is to achieve accurate, precise mappings"
- For a small set of metadata standards, hardwired crosswalks may result more adequate than ontology-based solutions
- In the geographic information context, the set of metadata standards is small and syntax and semantics are relatively fixed

Crosswalk-based semantic interoperability

Many different mappings have been made:

- MARC 21 to Dublin Core
- Dublin Core to USMARC
- Dublin Core to EAD/GILS/USMARC
- ...

Mostly, the only result included is the mapping table, almost no one offers details about the process.

Construction of crosswalks between metadata standards

1. Harmonization – obtain a formal and homogeneous specification of both standards

2. Semantic mapping – mapping table to determine the semantic correspondence of elements between the standards

3. Additional rules for metadata conversion – to solve problems like differing hierarchy levels, data type conversions, etc.

4. Mapping implementation – obtain a completely automated crosswalk (application of some tool)

Harmonization

- Standards often have properties that are very similar
- If had a fixed way of describing these properties:
  - every metadata standard could be described in a similar way
  - similar processes could be applied to related metadata standards
  - standards implementation would be simplified
  - *development of new crosswalks* between them would be simplified

Harmonization

- Generalization and formalisation: by means of a canonical representation or a specification language
- Because most standards use XML as exchange and presentation format, they also provide a DTD or XML Schema formally describing the syntax
- But: a mere syntactic description is not enough to store the necessary information to automate development of crosswalks
- Therefore: propose to create a table describing the elements of each standard apart from the available DTD

Semantic mapping

- Specification of a mapping between each element in the origin standard and the semantically equivalent element in the target standard
- Need a clear and precise definition of the elements
- Many metadata standards already provide a semantic mapping with related standards
- A mapping table should be produced at the end of this phase

Additional rules for metadata conversion

- Crosswalk: set of transformations
- $\rightarrow$ a completely specified crosswalk:
  a table of semantic mappings + a metadata conversion specification

Additional rules for metadata conversion

Content conversion

- Elements are frequently restricted to contain a particular data type, range of values or controlled vocabulary
- Analogous elements in different standards may have different content restrictions
- $\rightarrow$ specific rules are required to establish the correspondence
- It is necessary to establish the relationship between values on a one-to-one basis.

Additional rules for metadata conversion

Element-to-element mapping

- Properties specified with each element: whether repeatable or not, mandatory or not. Possibly non-trivial cases for crosswalks:
  - One-to-many: Trivial in most cases, but not all
  - Many-to-one: Must specify what to do with the extra elements.
    E.g. explicit concatenation rules, or rules for which value to select.
  - Extra/unresolved elements in source/target

Additional rules for metadata conversion

Hierarchy

- Most metadata standards organize their metadata hierarchically
- Crosswalks must consider the possible differences between source and target
- The mapping table itself shows the elements organized hierarchically in every standard

Mapping implementation

Automated implementation of crosswalks: the use of style sheets

- Most of the mentioned metadata standards use XML
  $\rightarrow$ most suitable technology to carry out implementation of crosswalks: XSL (eXtensible Stylesheet Language).
  - Purpose of XSL: manipulation and transformation of XML.
- Transformation method: constructing the style sheet applying to the original XML-document
- Each section is transformed by applying the previously defined mapping tables

Example transformation

Dublin Core (ISO 15836): 15 basic descriptors (result of an international and interdisciplinary consensus)

ISO 19115: defines the schema required for describing geographic information and services

DC element	ISO-CORE element
TITLE	Dataset title (M) MD_Metadata.identificationInfo.citation.title)
CREATOR	Dataset responsible party (O) (MD_Metadata.identificationInfo. pointOfContact) (when role="originator")
SUBJECT	Dataset topic category (M) (MD_Metadata.identificationInfo. topic-Category)
DESCRIPTION	Abstract describing the dataset (M) (MD_Metadata.identificationInfo.abstract)

Example transformation

- Main component of the crosswalk: the mapping between the standards
- Some DC elements cannot be mapped to ISO 19115 Core, but to ISO 19115 Comprehensive.
- Some elements from ISO 19115 Core that have no direct correspondence with elements from DC

Conclusions

- Organizations aim at migrating towards ISO
- Also asked to provide a generic description
- Would be more sensible to maintain metadata in accordance with a unique standard and produced by a stable cataloguing tool
  - Crosswalks would be applied when other views are required
- Crosswalks must be constructed by means of formalized methods
- Next step: prove the utility in the construction of search applications

References

- Metadata standard interoperability: application in the geographic information domain (2004)
- Integrating geographic and non-geographic data search services using metadata crosswalks (2003)
- Metadata And GIS: A Classification of Metadata for GIS (1998)