

Metadata Standard Interoperability: Application in the Geographic Information Domain ¹

J.Nogueras-Iso, F.J.Zarazaga-Soria, J.Lacasta, R.Bejar, P.R.Muro-Medrano

*Department of Computer Science and System Engineering University of Zaragoza.
Mara de Luna, 1. 50018-Zaragoza (Spain)*

Abstract

The use of metadata expands on the opportunities for interoperability. By interoperability, it is meant the ability to develop conventions that enable data exchange and integration. Metadata descriptions from different domains are not semantically distinct but overlap and relate to each other in complex ways. As the number, size and complexity of the metadata standards grow, the task of facilitating metadata in different standards becomes more difficult and tedious. One possible solution for this problem is the creation of mechanisms that enable the translation of this information in order to make it conform to the different standards. These mechanisms are denominated "crosswalks" and the objective of this work is to present the process of "crosswalk-creation", which has been used by a research team at the University of Zaragoza in order to translate information among some of the most extended standards for geographic information metadata.

Key words: metadata interoperability, geographic metadata, crosswalks

1 Introduction

1.1 Metadata and Metadata Crosswalks

Most commonly defined as "structured data about data" or "data which describes attributes of a resource" or, more simply, "information about data", the concept of metadata is not new: map legends, library catalogue cards and business cards are everyday examples. Basically, metadata offers description of the content, quality, condition, authorship, and any other characteristics of the object or data. It also provides for standardized representation of information. That is, similar to a bibliographical record or map legend, it provides a common set of terminology to define the resource or data.

Networked knowledge organization systems typically contain objects which are described using a multitude of diverse metadata schemas. Hence machine understanding of metadata descriptions which conform to schemas from different domains is a fundamental requirement for access to information within networked knowledge organization systems (Hunter, 2001).

The use of metadata expands on the opportunities for interoperability. By interoperability, it is meant the ability to develop conventions so as data exchange and integration becomes possible. As a special kind of interoperability, semantic interoperability is the agreement about content description standards. In the Internet Commons, disparate description models interfere with the ability to search across discipline boundaries. Promoting a commonly understood set of descriptors that helps to unify other data content standards increases the possibility of semantic interoperability across disciplines. In this manner, by using an agreed set of terms, it is possible to search, locate and retrieve data with a high degree of accuracy while resting assured of its potential use and authenticity.

Metadata descriptions from different domains are not semantically distinct but

Email address: j.nog, j.vay, j.lacasta, r.bejar, p.muro@unizar.es
(J.Nogueras-Iso, F.J.Zarazaga-Soria, J.Lacasta, R.Bejar, P.R.Muro-Medrano).

URL: <http://i.aaa.cps.unizar.es> (J.Nogueras-Iso, F.J.Zarazaga-Soria, J.Lacasta, R.Bejar, P.R.Muro-Medrano).

¹ The basic technology of this work has been partially supported by the Spanish Ministry of Science and Technology through the projects TIC2000-1568-C03-01 from the National Plan for Scientific Research, Development and Technology Innovation and FIT-150500-2003-519 from the National Plan for Information Society; and by the Aragn Government through the project P089/2001. The work of J. Lacasta has been partially supported by a grant from the Aragn Government and the European Social Fund (ref. B139/2003).

overlap and relate to each other in complex ways. As the number, size and complexity of the metadata standards grow, the task of facilitating metadata in different standards becomes more difficult and tedious. In order to minimize the cost of time for the creation and maintenance of metadata and to maximize its usefulness to the wider audience of users, it should be desirable to use a unique metadata standard in storage labours and provide automated views of metadata in other related standards. Furthermore, other times the metadata interoperability is not uniquely a cross-domain problem. Within the same domain, a metadata describing an instance of an entity A can be derived from a set of metadata entries describing instances from an entity B. For instance, the bibliographic records describing a collection of books can be summarized to obtain the metadata which describes the entire collection. But once again, it should be desirable to maintain uniquely the source metadata entries and generate automatically the derived metadata.

Metadata records, each one describing a specific resource, are grouped in catalogues that provide users with the possibility of identifying the resources of their interest. Apart from the chosen metadata-standard, the metadata cataloguing systems must support (recognize) three forms of metadata (Nebert, eds): the implementation form (within a database or storage system), the export or encoding format (a machine-readable form designed for transfer of metadata between computers), and the presentation form (a format suitable to viewing by humans). For last two forms, there is a general consensus about the use of XML (eXtensible Markup Language (W3C, 2000)). First of all, it includes a capable mark-up language with structural rules enforced through a control file (Document Type Definition or DTD) to validate document structure, i.e. conformance with a metadata standard DTD. And secondly, through a companion specification (XML Style Language, or XSL (W3C, 2003)), an XML document may be used along with a style sheet to produce flexible presentations or reports of content according to user requirements.

According to this philosophy, the tendency of the current cataloguing systems is to interchange metadata in XML which conforms to a specific standard on user demand, that is to say, providing different views of the same metadata. In order to maintain this interoperability across related metadata standards, it is necessary the creation of software systems able "to speak several metadata dialects", that is to say, systems that provide crosswalks between metadata standards. According to the Dublin Core Metadata Glossary (DCMI, 2001): "A crosswalk is a table that maps the relationships and equivalencies between two or more metadata formats. Crosswalks or metadata mapping support the ability of search engines to search effectively across heterogeneous databases, i.e. crosswalks help promote interoperability".

Let us imagine a scenario where three different metadata-databases store meta-information that describes the elements from a library (books, reports

and other kinds of documents), events (movies, theaters, recitals, etc) and geographic data (maps, satellite images, etc) respectively. These databases can be used for providing specialized high-level services such as tourist information (events and publications can be linked with data for travelling to a tourist destination) or cultural information (publications can be linked to an event, and it could be useful to provide maps for accessing to the places where the event occurs). The problem is that standards used in each metadata-database belong to a distinct domain and it will be necessary to unify the metadata-access (search and retrieval) methods. Fig. 1 displays the scenario described above and the different databases that must be integrated.

Fig. 1. Crosswalk use cases

If we had to develop a system for the tourist information provider, this system should use a homogenous mechanism for querying and accessing the three databases. That is to say, the metadata schema of the tourist information provider system should be independent of the metadata representation used by the three databases. For instance, in the referred example, the tourism information provider could query the system and manage the information using Dublin Core, whereas the cultural information provider manages only MARC metadata. The aforementioned homogenous mechanism should be a crosswalk broker facilitating the integration and coordination of crosswalks when needed. This broker should consist of a repository of crosswalks (Dublin Core \leftrightarrow MARC, Dublin Core \leftrightarrow ISO 19115, and MARC \leftrightarrow ISO 19115 in the previous example) and the software for activating and processing these crosswalks when needed. Fig. 2 shows an example of the sequence of crosswalks applied in order to query the databases and obtain the results.

Presumably, given that de-facto standard for the exchange of metadata is XML, a formal implementation of crosswalks should be based on XSL technology (W3C, 2003) (Fig. 2 depicts the order of application of different XSL stylesheets). However, the construction of crosswalks between standards is much more than the use of a series of programming technologies. A crosswalk specifies the mapping between two related standards, thus enabling communities that use one standard to access the content of elements defined in another one. Unfortu-

Fig. 2. Crosswalks applied for the tourist information provider use case

nately, the construction of crosswalks constitutes a difficult and error-prone task that requires deep knowledge and vast experience with the standards. The knowledge required to construct a crosswalk is particularly problematic since each metadata standard has been developed frequently in an independent form and therefore different terminology, specialized methods and processes are used. Moreover, the maintenance of crosswalks between metadata standards which are not stable and subject to changes is even more problematic due to the additional requirement of adjusting crosswalks to historical versions. For that reason, the harmonization in the consistent specification of related metadata standards is vital to the development of crosswalks. Thanks to this harmonized specification, it is easier to match the metadata elements of the different standards. The objective of this work is to present the process followed to carry out a series of crosswalks that enable interoperation across some of the most relevant standards for geographic information metadata.

1.2 Geographic Information Metadata

Issues arisen in the development of metadata crosswalks are not constrained to a specific application domain. That is to say, similar problems must be solved for digital libraries metadata or for geographic information metadata. Therefore, given the independence of the application area and considering that the results of this work are applied to geographic information context, this subsection presents some geographic information metadata concepts.

The geographic information is the information that describes phenomena associated directly or indirectly with a location with respect to the Earth surface. This information is vital for decision-making and resource management in diverse areas (natural resources, facilities, cadastres, economy...), and at dif-

ferent levels (local, regional, national or even global) (Nebert, eds). Nowadays, large amounts of geographic data are gathered by different institutions and companies. In fact, it is recognized that around 80% of the databases used by the public administration contain some kind of geographic reference (postal codes, cartographic coordinates...). 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.

In order to extend the use and understanding of metadata through different communities of users, e.g. to enable distributed searches across a network catalog servers, it is necessary to use well-defined contents and thus adjust them to a metadata standard. In this way, there have been a lot of standard proposals to describe consistently a geographic resource, which have arisen at national or global level and with different scopes. Some of the most extended ones are:

- ² The "Content Standard for Digital Geospatial Metadata" (CSDGM) of the Federal Geographic Data Committee (FGDC) (Federal Geographic Data Committee(FGDC), 1998). This American initiative is the only one that has the rank of standard at this moment. It was carried out in the United States by the FGDC and approved in 1994. It is a national standard for spatial metadata development for give support to the construction of the United States Spatial Data National Infrastructure. This standard has been adopted in other countries like South Africa or Canada.
- ² The European voluntary norm prENV 12657 (European Committee for Standardization(CEN), 1998) from the European Committee for Standardization (CEN).
- ² Dublin Core (Dublin Core Metadata Initiative, <http://www.dublincore.org>). It is a metadata standard of general outreach, very popular in the world of digital-libraries, which is being adopted by geographic information world in order to enable compatibility with other cataloguing information systems.
- ² The international standard ISO/DIS 19115 (International Organization for Standardization (ISO), 2003). In 1992, the International Standard Organization (ISO) created the committee 211 (ISO/TC 211) with responsibilities in "geomatics". This committee has prepared a family of standards that are obtaining the rank as official international standard. One of these standards is the Nr. 19115, in charge of the standardization of geo-spatial metadata. The standardization process of ISO 19115 has been finally completed in April 2.003. Currently, this committee is preparing the XML implementation of this standard (the future ISO 19139).

Apart from these main standards, there are other metadata standard initiatives arisen at a regional, national or domain-specific level like: the Spanish norm for geographic information exchange known as MIGRA ("Exchange Mechanism for Relational Geographic Information constituted by Aggrega-

tion"); the UDK-metadata standard from the German Environmental data catalog; or GELOS ("Global Environmental Information Locator Service") from the G-7/G-8 Environment and Natural Resources Management project (<http://www.g7.fed.us/enrm/>).

The intention of the different organizations who had proposed these standards, many of them still drafts or pre-standards, was the harmonization of all the initiatives around ISO as soon as it was approved as international standard. This happened in April 2003, so it is expected that the convergence process will occur in the next future.

The rest of this paper is structured as follows. Next section presents the work related with the semantic interoperability with special interest in the geographic information domain. Section 3 proposes a general process to formalize metadata standards and construct crosswalks. Next, the results of developing several crosswalks using this process is explained. Finally, this work ends with a section of conclusions.

2 Related work

There are two main approaches to handle the semantic interoperability problem between metadata standards: solutions that are based on the use of ontologies (i.e. establishing or inferring relationships between the metadata vocabularies employed by the different metadata standards); and the creation of specific crosswalks for one-to-one mapping. In next subsections, related work to both approaches is presented.

2.1 Ontology based semantic interoperability

The impact of the Internet as the biggest platform for the distribution of resources has motivated the birth of a great deal of initiatives that aim at solving the problem of semantic interoperability on the Web. And most of these approaches propose the use of ontologies and RDF technologies as the basis for information sharing (Pundt and Bishr, 2002). As mentioned in the introduction chapter, an ontology is defined as an explicit specification of some shared vocabulary or conceptualization of an specific subject matter, and it seems to be an adequate methodology that helps to define a common ground between different information communities. Furthermore, these approaches are closely related to a new conception of the Web: the Semantic Web. According to (Berners-Lee et al., 2001; W3C, 2001), "the Semantic Web is an extension of the current web in which information is given well-defined meaning, better

enabling computers and people to work in cooperation".

RDF (Resource Description Framework) (W3C, 1999) is a W3C recommendation for modelling and exchanging metadata. The major advantage of RDF is its flexibility. RDF is not really a metadata standard defining a series of elements. On the contrary, it can be considered as a meta-model that contains other metadata schemas or combinations of them. RDF uniquely defines a simple model for describing the interrelationships among resources in terms of named properties and values. But for the declaration and interpretation of those properties, a complementary technology of RDF is needed. This complementary technology is RDFS, which stands for RDF Schema although it has been recently renamed as RDF Vocabulary Description Language (W3C, 2002). RDFS provides a rich set of constructs to define and constrain the interpretation of vocabularies used in a certain information community. In fact, a RDFS document defines the ontology that is used to construct particular RDF documents in an information community. That is to say, RDFS can be used to define the semantic meaning of metadata elements contained in a metadata standard or schema, viewing the structure of metadata schemas as ontologies. In this sense, an instance of RDFS could be seen as the ontology of metadata elements used for a particular profile. A more general solution for interoperability should be based on the use of ontologies that define the semantic meaning of metadata elements contained in each metadata standard or schema.

Moreover, RDFS documents (defining ontologies) can reuse other ontologies that may be located and controlled in other places on the Internet. As a result, if different information communities define their domain ontologies by means of RDFS and publish their metadata in RDF, other information communities can check whether this metadata (including the semantics) is usable or not.

An example of this kind of approaches is the work presented in (Hunter, 2001). There, the ontology is implemented as a thesaurus, named MetaNet, whose objective is to provide the semantic knowledge required to enable machine understanding of equivalence and hierarchical relationships between metadata terms from different domains. The scope of this thesaurus is limited to the most significant metadata models/vocabularies used for describing attributes and events associated with resources and their life cycles. This encompasses metadata vocabularies from the bibliographic, museum, archival, record keeping and rights management communities. MetaNet has been developed by performing WordNet (an upper level ontology) searches of the core terms used in the different domains. In order to implement dynamically the interoperability, this work provides an RDFS representation of the MetaNet thesaurus together with an XML style sheet. This stylesheet parses an input metadata description and searches the MetaNet RDFS representation for the elements in the output metadata standard that are equivalent to the input element names.

As alternatives to RDF technologies for resource description and knowledge representation on the Web, there are some proposals like SHOE language, which can be found in (Hein and Hendler, 2000). This work remarks the fact that RDFS representation is more limited than most artificial intelligence ontologies because it does not possess any mechanisms for defining general axioms (rules that allow additional reasoning). On the contrary, SHOE is presented as an ontology-based knowledge representation language designed for the Web that permits the discovery of implicit knowledge through the use of taxonomies and inference rules. The syntax of this language is defined as an application of SGML that extends the HMTL DTD, primarily because XML was still evolving when SHOE was created. SHOE ontologies are made publicly available by locating them on web pages. Then, ordinary web pages (the resource itself) are extended with special tags to include instances of the entities defined by a referenced SHOE ontology. Finally, the interoperability in SHOE is through use of the ontology extension and renaming features (two categories are similar to the extent that they share the same supercategories).

As it has been seen, these approaches offer flexible solutions for interoperability. However, this ambitious aim of flexibility may also imply a lack of accuracy in the mappings performed. The ontology based solutions presented until now do not consider the local structural constraints imposed by the different specific domains, e.g. parent/child relationships; cardinality/occurrence constraints; datatyping, enumeration and formatting constraints on the element values. The SHOE approach even defines its own metadata encoding language. As it is stated in (Hunter, 2001): "the wider the targeted scope of interoperability, the more difficult it is to achieve accurate, precise mappings". For a small set of metadata standards, whose syntax and semantics are relatively fixed and constrained, hardwired crosswalks establishing the mapping between metadata terms (from specific standards) may result more adequate than ontology-based solutions. That is precisely the case in the geographic information context.

2.2 Crosswalk based semantic interoperability

There is a big experience in developing mappings among several standards and different domains. Interesting collections of links to metadata-crosswalk initiatives can be found at <http://www.sinica.edu.tw/metadata/tool/mapping-foreign.html> and <http://www.ukoln.ac.uk/metadata/interoperability/>. There, it is possible to find mappings from MARC 21 to Dublin Core, Dublin Core to USMARC, Dublin Core to EAD/GILS/USMARC, Dublin Core to FIN-MARC/GILS, Dublin Core to IAFA/ROADS templates, Dublin Core to UNIMARC, FDGC to GCMD DIF, FGDC to USMARC, and others.

The Canadian Heritage Information Network (CHIN, <http://www.chin.gc.ca/>) (L.E.Sherwood, 1998) offers some links to crosswalks that may be of use to museums. Some examples of the links offered could be the "Crosswalk of Metadata Element Sets for Art, Architecture, and Cultural Heritage Information and Online Resources" (developed by the Getty Research Institute, its mapped standards include Categories for the Description of Works of Art, VRA Core Categories, Dublin Core, Object ID, the CIMI Access Points, the Guide to the Description of Architectural Drawings, as well as library and archival standards), or the Mapping from CHIN Natural Sciences Data Dictionary to Darwin Core (CHIN has completed a mapping between the Darwin Core and the CHIN Natural Sciences Data Dictionary so museums following the CHIN Natural Sciences Data Dictionary could use the same or similar mapping to Darwin Core). As long as the geographic information metadata is concerned, the MADAME project (<http://www.shef.ac.uk/scgisa/MADAMENew/faq.html>) developed a correspondence between Dublin Core and ISO19115.3. This correspondence, which can also be found at the ETeMII project document (Craglia, 2001), offers a table with the correspondence between the Dublin Core sections and the ISO 19115.3 sections, but it does not offer any automatic or semi-automatic tool for transforming from one to the other. It also provides a correspondence between prENV 12657 and ISO TC 211 /CD 19115.3 with similar limitations (<http://www.shef.ac.uk/scgisa/MADAMENew/cen2iso.pdf>).

The Canadian Geospatial Data Infrastructure has also developed a crosswalk between ISO19115 and FGDC standard (see (GeoConnections, 2001; Teng, 2000)). The discovery portal of this infrastructure (GeoConnections Discovery Portal at <http://geoconnections.ca>) offers data products catalogued in accordance with the FGDC CSDGM standard but plans to support the new ISO19115 in future versions.

Additionally, the DGIWG (Digital Geographic Information Working Group) Metadata Work Program, supported by NIMA (National Imagery and Mapping Agency of United States), offers a crosswalk between ISO19115 and FGDC standard (<http://metadata.dgiwg.org/ISO19115/related.htm>) too. This program is taking a leading role in developing an implementation model and XML schema of the ISO 19115 metadata standard (officially known as ISO 19139) and provides a Metadata Development Efforts Website to coordinate the metadata standardization efforts of several organizations.

On the other hand, the own FGDC organization provides a mapping between FGDC standard and Dublin Core. It is the *mp* tool (parser of formal metadata provided by FGDC) that generates an HTML output, where FGDC elements are mapped to Dublin Core elements in the META tags of the HEAD section. The intended use of META tags is to divulgate the content of a Web page, thus making this meta-information visible to search engines.

Inside the project "Cooperative Online Research Catalog (CORC)" a converter among FGDC, Dublin Core and MARC21 standards was developed as one of its goals (Chandler et al., 2000, 1999). One of the motivations of this work was the unsuccessful results (on average) obtained from queries directed at nodes of the FGDC Clearinghouse (Federal Geographic Data Committee (FGDC), 2003). Therefore, it was proposed to convert FGDC metadata into more widely used metadata standards for inclusion in systems other than the FGDC Clearinghouse.

Most specifically within the context of environmental geographic information, a mapping between ISO and GELOS has been built inside the project "ETC/CDS (EIONET): European Topic Centre on Catalogue of Data Sources" (<http://eionet.eu.int/>). Another work is the mapping between UDK-metadata standard and ISO. This mapping has been developed inside the project "UDK (Umwelt Data Katalog)" (<http://www.umweltdatenkatalog.de/>), German Environmental data catalog.

Most of these works do not include any other result apart from the table that maps the relationships and equivalencies among the standards. In some cases, any kinds of tools for automatic or semiautomatic translation are included. And almost no-one offers details about the process followed. In this sense, there are two interesting works that manage this problem. In (Woodley, 2000) some of the common misalignments in creating crosswalks are presented. The other interesting work is (Pierre and LaPlant, 1998). It provides many of the key issues involved in crosswalk development and identifies those areas in which harmonization can contribute. As the paper explains, its main contribution is the delineation of the general issues involved in the harmonization of metadata standards and in the development of crosswalks between related metadata standards. Many concepts and ideas presented in it has been used as a base for the development of the work presented in this paper.

3 Construction of crosswalks between metadata standards

This section presents the steps of the process that has been followed to construct a series of crosswalks between standards and that simplifies its implementation by means of the use of formal specifications and automated mechanisms. The process has the following steps:

- (1) Harmonization: This phase aims at obtaining a formal and homogeneous specification of both standards.
- (2) Semantic mapping: In order to determine the semantic correspondence of elements between the standards of metadata a deep knowledge of the origin and target metadata standards is required. As result of this phase,

- a mapping table is created.
- (3) Additional rules for metadata conversion. Apart from the mapping table, it should be necessary to provide additional metadata conversion rules in order to solve problems such as different level of hierarchy, data type conversions, etc.
 - (4) Mapping implementation: The last objective of the process is to obtain a completely automated crosswalk by means of the application of some type of tool. In this way, maintaining only one set of metadata, searches and views can be provided according to the different families from metadata.

The following subsections present further details of each one of these steps.

3.1 Harmonization

Many of the metadata standards use similar properties in the definition of their content elements. Some examples of similar properties could be: a unique identifier for each metadata element (for example: tag, label, identifier); a semantic definition for each element; the mandatory, optional or conditional character of each element; the multiplicity or allowed number of occurrences of an element; the hierarchical organization with respect to the rest of elements; or constraints on the value of an element (e.g. free text, numerical range, dates or a predefined code list). If the way to express those properties were fixed, every metadata standard could be described in a similar way. Consequently, similar processes could be applied to related metadata standards, thus simplifying not only standards implementation but also the development of new crosswalks between them.

The generalization and formalization in the specification of metadata standard properties are usually done by means of a canonical representation or a specification language. This procedure is analogous to the specification of a programming language syntax using the well-known notation Backus-Naur-Form (BNF (Naur, ed.)). In fact, thanks to the circumstance that most standards use XML as exchange and presentation format, they also provide a DTD or XML-Schema that describes formally their syntax.

Nevertheless, a mere syntactic description of a metadata standard is not enough to store all the information necessary to automate the development of crosswalks. For instance, a minimum set of data types must be defined as a basis to obtain from it the derived data types that are required to represent all the elements in the target standard. And in addition to this, as it happens with BNF, a metadata specification does not contain information about the semantics of elements. For that reason, in this step it is proposed the creation of a table (that could be implemented by means of the use of a Excel

sheet) describing the elements of each standard apart from the DTD available for each standard. In this table, each element of metadata will be defined by means of the following fields: number assigned in its own metadata standard according to its level in the hierarchy, "long" name assigned by the standard to this element (besides, it is recommended to indent sections and subsections of metadata in order to show the hierarchical structure of the standard), "short" name of the element (this "short" name usually corresponds with the tag used for XML encoding of metadata), multiplicity and mandatory constraints that the standard impose on this element, semantic definition of the element, and data type for the values of this element.

3.2 Semantic mapping

The most important task in the development of crosswalks is the one in charge of determining the semantic correspondence between the elements of the standards to be mapped (Pierre and LaPlant, 1998). This task implies the specification of a mapping between each element in the origin standard and the element that is semantically equivalent to this one in the target standard. For that purpose, it is very important to count on a clear and precise definition of each-standard elements.

Many metadata standards already provide a semantic mapping with standards of related metadata, frequently this mapping appears in form of a table in an annex of the standard. In the process that appears here, at the end of this phase, a mapping table is produced.

3.3 Additional rules for metadata conversion

A crosswalk is a set of transformations that applied to a set of elements in the source metadata standard produce, as a result, an equivalent content in the target standard, which has been properly modified and redistributed to meet the requirements of the analogous elements. Therefore, a completely specified crosswalk must consist of a table of semantic mapping accompanied by a metadata conversion specification. This specification contains the additional transformations required to convert the metadata document whose contents fulfil the source standard into a document whose contents fulfil the target standard. Following subsections present the different metadata conversion problems that may arise and which those additional rules must solve. These rules are usually included as descriptions in an additional column of the mapping table or in an annex document.

3.3.1 *Content Conversion*

Frequently, metadata standards restrict the contents of each element to a particular data type, range of values or controlled vocabulary. In some cases, two analogous elements in different standards may have different content restrictions. For example, it could happen that a text value must be transformed into a numerical value or a date value. Therefore specific rules are required to establish the correspondence between the initial element whose values may be specified as free text and a target element whose value is constrained to a controlled vocabulary. Moreover, when mapping two elements restricted to different controlled vocabularies, it is necessary to establish the relationship between values on one-to-one basis.

3.3.2 *Element to element mapping*

All metadata standards specify a number of properties associated with the definition of each element. For instance, some standards qualify each element as repeatable or non-repeatable and indicate additionally whether this element is mandatory or optional. Others, such as FGDC, incorporate both features into a single property containing a lower and upper bound number of occurrences. A lower bound of zero indicates an optional element, whereas a lower bound of one indicates that the element must occur at least once and thus is mandatory. For crosswalk development, these properties must be taken into careful consideration. The trivial case is the mapping between two elements that share identical properties, e.g. a mandatory non-repeatable element which matches with a mandatory non-repeatable element in target standard. The rest of cases can be classified in the following categories:

- ² One to many. In most cases, a one-to-many map is trivial; an occurrence of the source element maps to a single occurrence in the target element. However, there are cases where the mapping requires more explicit resolution. For example, the source standard may contain a non-repeatable "keywords" element and according to its definition the content of this element consists of one or more keyword values separated by commas. Nevertheless, this element should match with a repeatable element in the target standard, that is to say, an occurrence for each keyword value. In this case, the mapping requires specialized knowledge of the composition of the source element, and how it expands into multiple target elements. Another interesting case is the mapping of one source element to two unique target elements. For example, a crosswalk for Dublin Core to FGDC standard should map the Dublin Core "Rights" element to the "Access Constraints" and "Use Constraints" elements in FGDC. In this case, special rules must be provided to extract correctly the content of the source element and map it to the corresponding elements in FGDC.

- ² Many to one. The many-to-one map must specify what to do with the extra elements. If the solution adopted is to map all values of the source element to a single value in the target element, explicit rules are required to specify how concatenate the original values. Alternatively, if the solution is to map a unique value of the source element, with the consequent information loss, a rule must indicate the criteria for this value selection, e.g. the first value or the most recently added.
- ² Extra elements in source. Another problem arises when a source element does not have any equivalent element in the target standard. Since many metadata standards provide the ability to capture additional information or to define appropriate extensions, a rule must be established to precisely specify how these extra-elements element are handled. " Unresolved mandatory elements in target. In some cases, mandatory elements in the target standard may have no mapping in the source standard. Because the target requires a value for the mandatory elements, the crosswalk must provide a rule to fill these elements with appropriate values.

3.3.3 *Hierarchy*

Most metadata standards organize their metadata hierarchically (by means of sections and subsections). The crosswalk must consider the possible differences between the hierarchies of the source and target standards. In the process presented, the mapping table itself shows the elements organized hierarchically in every standard, although it excludes the infinite mapping of those sections, which are recursively defined (e.g. Citation section of FGDC) and make the depth of the hierarchy unlimited.

3.4 *Automated implementation of crosswalks: the use of style sheets*

Taking into account that most metadata standards presented in the Introduction chapter use XML as exchange and presentation format, it has been considered that the most suitable technology to carry out the implementation of crosswalks is by means of XSL (eXtensible Stylesheet Language (W3C, 2003)), whose purpose is precisely the manipulation and transformation of XML. XSL is a language for expressing style sheets that integrates two related languages: a transformation language (XSL Transformations or XSLT); and a formatting language (XSL Formatting Objects) of XML documents, which is comparable to the language CSS (Cascading Style Sheets) for HTML pages. The transformation language (XSLT) provides elements that define rules to transform an XML-document into another XML-document. This second document can use the same set of elements that the original document (it is associated to the same DTD) or can use a completely different set of elements. Therefore, the

method to make transformations will consist of constructing the style sheet that applied to the original XML-document (in agreement the corresponding standard of metadata) generates as a result an XML-document whose elements fulfil the target standard, and that contains the same information represented in the input document. Next it is detailed the general methodology that has been followed during the construction of style sheets that implement the crosswalks between the different metadata standards. The followed methodology is based on the successive transformation of each section applying the mapping tables that have been defined previously. In particular, the following steps are followed to complete the style sheet:

- 2 Establish the document type declaration that will appear in the output document, and that will include the route (URL) of the DTD corresponding to the target standard.
- 2 Next, for each section to match in the target standard:
 - ⌚ A template will be created (based on the mapping table) whose pattern is the element (name of section or subsection) in the source standard that generates the corresponding elements in the target. In this template the necessary transformation rules will be applied in order to fulfil the specification with respect to the properties and content in the target standard.
 - ⌚ Once the first version of the style sheet has been built, it is applied to a XML document that conforms to the source standard, and contains values for all the elements belonging to the section previously matched. The style sheet processor (e.g. Java XML parser provided by Oracle at <http://technet.oracle.com/>) generates as a result a new document. Although this document will not probably validate the DTD corresponding to the target standard (it only contains the sections mapped until this moment), it must be verified that the transformations have been made correctly. By means of a XML edition tool it is possible to visualize the XML document as a tree of nodes, which correspond to the sections, subsections or PCDATA tags. Therefore, this tree of nodes is used to check: the absence of a mandatory element; the order of generated elements; and the content constraints. In case of detecting some error, the template must be revised.
 - ⌚ Additionally, it should be verified that there is not information loss in case the inverse style sheet were applied to the target document. Usually, a crosswalk and the inverse crosswalk are developed in parallel. If there exists some difference between the initial document and this new generated document, the mapping table should be verify the cause of the problem. It may be due to a problem of extra-elements in source standard that has not been resolved by any rule. But if this circumstance does not take place, the XSL template should be checked again.
 - ⌚ Once it has been proved that the transformation of the last section has been done correctly, the process must be started again for the next section in the source standard until the crosswalk is completely implemented.

4 Putting the method to work: Transformation between ISO 19115 Core and Dublin Core

The Dublin Core Metadata Initiative (<http://www.dublincore.org>), created in 1995, is an organization which promotes the widespread adoption of interoperable metadata standards and the development of specialized metadata vocabularies that enable more intelligent information discovery systems. The Dublin Core metadata element set is a standard for the description of cross-domain information resources. This set consists of 15 basic descriptors which are the result of an international and interdisciplinary consensus. Nowadays, the Dublin Core metadata element set has become an important part of the emerging infrastructure of the Internet. Moreover, since 4th April 2003, the Dublin Core Metadata Element Set standard has been adopted as ISO standard (ISO 15836). This approval is the culmination of an incremental process to bring the Dublin Core metadata element set into a worldwide audience.

Table 1

Dublin Core - ISO 19115 Core mapping

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.topicCategory)
DESCRIPTION	Abstract describing the dataset (M) (MD_Metadata.identificationInfo.abstract)
PUBLISHER	Dataset responsible party (O) (MD_Metadata.identificationInfo.pointOfContact) (when role="publisher") Metadata point of contact (M) (MD_Metadata.contact)
DATE	Dataset reference date (M) (MD_Metadata.identificationInfo.citation.date) Metadata date stamp (M) (MD_Metadata.dateStamp)
TYPE	Spatial representation type (O) (MD_Metadata.identificationInfo.spatialRepresentationType)
FORMAT	Distribution format (O) (MD_Metadata.distributionInfo.distributionFormat)
IDENTIFIER	On-line resource (O) (MD_Metadata.distributionInfo.transferOptions.onLine.linkage)
SOURCE	Lineage (O) (MD_Metadata.dataQualityInfo.lineage.source.description)
LANGUAGE	Dataset language (M) (MD_Metadata.identificationInfo.language)
COVERAGE SPATIAL	Geographic location of the dataset (by four coordinates or by geographic identifier) (C) (MD_Metadata.identificationInfo.extent.geographicElement)
COVERAGE TEMPORAL	Additional extent information for the dataset (vertical and temporal) (O) (MD_Metadata.identificationInfo.extent.temporalElement.extent)

On the other hand, the standardization process of ISO 19115 has been nearly completed. This standard defines the schema required for describing

geographic information and services. It provides information about the identification, the extent, the quality, the spatial and temporal schema, spatial reference, and distribution of digital geographic data. And although ISO 19115 is mainly oriented to the description of digital data (geographic datasets, dataset series or individual geographic features), its principles may be also extended for many other forms of geographic data such as maps, charts, and textual documents as well as non-geographic data. Another remarkable aspect concerning ISO 19115 is that despite defining an extensive set of metadata elements, in practice only a subset of these elements is used. However, it is essential to maintain a basic minimum number of metadata elements for describing geographic datasets. For this purpose, the standard defines a small list of core metadata elements. These core metadata elements facilitate interoperability because they allow users to understand without ambiguity the geographic data and metadata provided by either producers or distributors. Furthermore, ISO 19115 enforces all application profiles of this standard to include these core elements. Dublin Core does not aim at displacing other metadata standards. Instead, it is intended to co-exist (frequently Dublin Core descriptors form part of broader resource descriptions) with metadata standards that offer other semantics. In fact, the potential of Dublin Core is to provide the visibility of a collection of resources across different subject domains and at a low cost. Therefore, for standards like ISO 19115 which do not describe geographic information as general-purpose data, the interoperability with Dublin Core results very appealing. The tool to facilitate this interoperability is the definition of automatic-crosswalks between these standards.

Table 2

Dublin Core elements that must be mapped to ISO19115 Comprehensive (no match with ISO19115 core)

DC element	DC reñement	ISO 19115 Comprehensive
CONTRIBUTOR		MD_Metadata.identificationInfo.credit
RELATION	isVersionOf, replaces, isPartOf, references, isFormatOf, < otherwise >	MD_Metadata.identificationInfo.aggregationInfo
RELATION	isPartOf	MD_Metadata.identificationInfo.citation.series.name
RIGHTS	< none >	MD_Metadata.identificationInfo.resourceConstraints
RIGHTS	accessRights	MD_Metadata.identificationInfo.resourceConstraints.accessConstraints
AUDIENCE	< none >, educationLevel	MD_Metadata.identificationInfo.purpose
AUDIENCE	mediator	MD_Metadata.distributionInformation.distributor.distributorContact (when role=distributor)

Following the process described above, the crosswalk between both standards has been built. The main component of this crosswalk is the mapping between the standards. The ISO 19115 Core compiles the 22 elements that minimally describe a geographic resource. And therefore, the main focus of this work item was to obtain the crosswalk between Dublin Core and these basic elements of

ISO. Table 1 presents the mapping table between Dublin Core and ISO 19115 Core.

As it can be observed in table 1, there are four elements of Dublin Core that have no correspondence with any element of the Core version of the ISO 19115. These four elements are "CONTRIBUTOR", "RELATION", "RIGHTS" and "AUDIENCE". Nevertheless, all of them have a correspondence with one or more elements of the ISO 19115 Comprehensive profile. The ISO 19115 Comprehensive profile fully defines the complete range of metadata required to identify, evaluate, extract, employ, and manage geographic information. In fact, it almost includes all the metadata entities defined in ISO 19115 document. Tables 2 and ?? show this mapping to the ISO elements contained in the Comprehensive profile.

The lack of mapping between these last 4 DC elements and ISO 19115 Core metadata could justify the expansion of the ISO 19115 Core Metadata to include the comprehensive elements appearing in table 2. This way, a full mapping DC ! ISO 19115 Core would be possible. The aim of Dublin Core is to compile the minimum elements that describe a resource and thus ISO Core should include at least these Dublin Core elements to be really "Core".

Additionally, another deficiency in the mapping that can be observed is that there are some elements from the Core version of the ISO 19115 having no direct correspondence with elements from Dublin Core (the full information about this mapping and the solutions proposed for these deficiencies can be found in (CEN/ISSS Workshop - Metadata for Multimedia Information - Dublin Core, 2003c), (CEN/ISSS Workshop - Metadata for Multimedia Information - Dublin Core, 2003b), (CEN/ISSS Workshop - Metadata for Multimedia Information - Dublin Core, 2003a)).

5 Conclusions

This work has presented the process followed to carry out the construction of a series of crosswalks that enable interoperability between some of the most used standards for geographic information metadata, illustrating it with a concrete example of one of the made crosswalks.

Nowadays, most organizations in charge of cataloguing geographic metadata (in accordance with standards like CSDGM (Federal Geographic Data Committee(FGDC), 1998) or CEN/TC 287 prENV 12657 (European Committee for Standardization(CEN), 1998)) aim at migrating towards the international ISO standard. Apart from that, they are usually asked to provide a more generic description of their resources, that is to say, they are asked to provide

a summary view of their specific geographic metadata understandable by general public. This summary view could be the one defined by Dublin Core, a de facto standard that is having great acceptance in public administration or in the description of web resources. Under these requirements, the use of different editors to maintain same metadata in each standard does not prove to be the best option. On the contrary, a more sensible option for an institution would be the maintenance of metadata in accordance with a unique standard and produced by a stable cataloguing tool. Then when other views of metadata are required, crosswalks would be applied to obtain the metadata conforming to the demanded standard. Nevertheless, it must be taken into account that these crosswalks must be constructed by means of formalized methods, which enable verification of information transformations and minimize the possible loss of information.

As it has been mentioned before, the process presented has been used for the elaboration of a series of crosswalks, which allow the interoperation between some of the most popular standards in geographic information metadata. There is no notice about the availability of any crosswalks between these standards, neither free nor paying. In fact, some contacts have been established with FGDC and ISO in order to contribute in the creation of the official crosswalk between both standards. In the same way, this research team has collaborated in an European project whose one of their objectives is to create a geographic application profile for Dublin Core standard and its mapping to ISO 19115. This project is founded by the European Committee for Standardization (Comit European de Normalisation, CEN, <http://www.cenorm.be>), from the European Union, and provide style sheets to transform metadata between ISO XML representation and Dublin Core RDF representation. The results of this project can be found in <http://www.cenorm.be/iss/Workshop/MMI-DC/>.

Once these crosswalks have been developed, next step is to prove their utility in the construction of search applications that perform queries against geographic information catalogs. These crosswalks will allow the establishment of restrictions and presentation of results in accordance with a standard selected by the application user on demand. The search client will access the gateway of a network of distributed catalogs, each of them providing metadata according probably to a different standard. However, thanks to the availability of crosswalks, the gateway will be able to translate user requests to the adequate format and present results according to the standard required by each client.

References

Berners-Lee, T., Hendler, J., Lassila, O., May 2001. The semantic web. Scientific American.

- CEN/ISSS Workshop - Metadata for Multimedia Information - Dublin Core, July 2003a. Draft Spatial Application Profile (ISSS/WS-MMI-DC/071). <http://www.cenorm.be/iss/Workshop/MMI-DC/>.
- CEN/ISSS Workshop - Metadata for Multimedia Information - Dublin Core, July 2003b. Guidance Material for Mapping Between Dublin Core and ISO in The Geographic Information Domain (ISSS/WS-MMI-DC/073). <http://www.cenorm.be/iss/Workshop/MMI-DC/>.
- CEN/ISSS Workshop - Metadata for Multimedia Information - Dublin Core, July 2003c. Mapping Between Dublin Core And The ISO Standard DIS 19115, "Geographic Information - Metadata (ISSS/WS-MMI-DC/072). <http://www.cenorm.be/iss/Workshop/MMI-DC/>.
- Chandler, A., Foley, D., Hafez, A., 4 October 1999. Dc-7 case study: Mapping essential federal geographic data committee (fgdc) metadata into dublin core. available at <http://eeirc.nwrc.gov/pubs/crosswalk/dc7casestudy.htm>.
- Chandler, A., Foley, D., Hafez, A., January 2000. Mapping and converting essential federal geo-graphic data committee (fgdc) metadata into marc21 and dublin core. towards an alternative to the fgdc clearinghouse. D-Lib Magazine 6 (1), also at <http://webdoc.sub.gwdg.de/edoc/aw/d-lib/dlib/january00/chandler/01chandler.html>.
- Craglia, M., 2001. Towards a european approach to metadata for geographic information. Etemii project document, ETeMII, <http://www.ec-gis.org:8080/wecgis/docs/F25765/D421-METADATA.PDF>.
- DCMI, Feb. 24 2001. Dublin core metadata glossary, final draft. Available online at <http://library.csun.edu/mwoodley/dublincoreglossary.html>.
- European Committee for Standardization(CEN), 1998. The CEN/TC 287 Geographic Information European Prestandards, ENV 12657. Euro-norme Voluntaire for Geographic Information - Data description - Metadata. <http://forum.afnor.fr/afnor/WORK/AFNOR/GPN2/Z13C/PUBLIC/WEB/ENGLISH/pren.htm>.
- Federal Geographic Data Committee(FGDC), 1998. Document FGDC-STD-001-1998. Content Standard for Digital Geospatial Metadata, version 2.0. Metadata Ad Hoc Working Group.
- Federal Geographic Data Committee(FGDC), 2003. Homepage of the National Geospatial Data Clearinghouse. <http://www.fgdc.gov/clearinghouse/clearinghouse.html>.
- GeoConnections, 2001. The access technical services manual version 1.1. <http://www.geoconnections.ca/>.
- Heo'in, J., Hendler, J., 2000. Semantic interoperability on the web. In: Association, G. C. (Ed.), Proceedings of Extreme Markup Languages 2000. Alexandria, VA, pp. 111{120, available at <http://www.cse.lehigh.edu/heo'in/pubs/extreme2000.pdf>.
- Hunter, J., 2001. Metanet - a metadata term thesaurus to enable semantic interoperability between metadata domains. Journal of Digital information 1 (8).
- International Organization for Standardization (ISO), 2003. Iso 19115:2003.

- geographic information - metadata.
- L.E.Sherwood, 1998. Standards for access to museum content: practical solutions or technical chimeras? *Computer Standards & Interfaces* 20, 111{115.
- Naur(ed.), P., 1963. Revised report on the algorithmic language algol 60. *Comm. ACM* 6 (1), 1{17.
- Nebert(eds), D., May 2001. Developing spatial data infrastructures: The sdi cookbook v.1.1. Global Spatial Data Infrastructure (<http://www.gsdi.org>).
- Pierre, M. S., LaPlant, W., 1998. Issues in crosswalking content metadata standards. Tech. rep., National Information Standards Organisation, <http://www.niso.org/press/whitepapers/crswalk.html>.
- Pundt, H., Bishr, Y., 2002. Domain ontologies for data sharing. an example from environmental monitoring using field gis. *Computer & Geosciences*.
- Teng, Y., June 2000. Use of xml for web-based query processing of geospatial data. Master's thesis, University of New Brunswick, master Thesis available at http://www.cs.unb.ca/tech-reports/files/TR00_135.pdf.
- W3C, February (22) 1999. Resource description framework (rdf) model and syntax specification. <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>.
- W3C, 2000. Extensible markup language (xml) 1.0 (second edition). W3C Recommendation, <http://www.w3.org/TR/2000/REC-xml-20001006>.
- W3C, 2001. The semantic web. W3C Technology and Society Domain, Semantic Web Activity, <http://www.w3.org/2001/sw/>.
- W3C, November (12) 2002. Rdf vocabulary description language 1.0: Rdf schema. W3C Working Draft , <http://www.w3.org/TR/2002/WD-rdf-schema-20021112/>.
- W3C, 2003. The extensible stylesheet language (xsl). <http://www.w3.org/Style/XSL/>.
- Woodley, M., 7 May 2000. Crosswalks: the path to universal access? Getty Research Institute, Getty Standards and Digital Resource Management, http://www.getty.edu/research/institute/standards/intrometadata/2_articles/woodley/index.html.