(Draft) International Journal of Metadata, Semantics and Ontologies.
2006, vol. 1, nº 3, p. 183-188.

Semantic interoperability based on Dublin Core hierarchical one-to-one mappings

R. Tolosana-Calasanz*, J. Nogueras-Iso,
R. Béjar, P.R. Muro-Medrano, and
F.J. Zarazaga-Soria

Department of Computer Science and Systems Engineering,
University of Zaragoza, Zaragoza, Spain
E-mail: rafaelt@unizar.es E-mail: jnog@unizar.es
E-mail: rbejar@unizar.es E-mail: prmuro@unizar.es
E-mail: javy@unizar.es
*Corresponding author

Abstract: The tendency of current cataloguing systems is to interchange metadata in XML according to the specific standard required by each user on demand. Furthermore, metadata schemas from different domains are not usually semantically distinct but overlap and relate to each other in complex ways. As a consequence, the semantic interoperability has to deal with the equivalences between those descriptions. There exist two main approaches in order to tackle this problem: solutions on the use of ontologies and solutions based on the creation of specific crosswalks for one-to-one mapping. This paper proposes a hierarchical one-to-one mapping solution for improving semantic interoperability.

Keywords: metadata standards; metadata servers; catalogue servers; spatial data infrastructures; interoperability; semantic interoperability; crosswalks; XML; Dublin Core; RDF; ISO 19115.

Reference to this paper should be made as follows: Tolosana-Calasanz, R., Nogueras-Iso, J., Béjar, R., Muro-Medrano, P.R. and Zarazaga-Soria, F.J. (xxxx) 'Semantic interoperability based on Dublin Core hierarchical one-to-one mappings', *Int. J. Metadata, Semantics and Ontologies*, Vol. x, No. x, pp.xxx–xxx.

Biographical notes: R. Tolosana-Calasanz is a Lecturer at the Computer Science Department of the University of Zaragoza (Spain). His research interests include metadata management and information retrieval in the context of spatial data infrastructures.

J. Nogueras-Iso, PhD, is an Assistant Professor at the Computer Science Department of the University of Zaragoza (Spain). His research interests include metadata management, information retrieval and ontologies in the context of spatial data infrastructures.

R. Béjar is a Lecturer at the Computer Science Department of the University of Zaragoza (Spain). His research interests include service chaining and semantic interoperability in the context of spatial data infrastructures.

P.R. Muro-Medrano, PhD, is an Associate Professor at the Computer Science Department of the University of Zaragoza (Spain). His research interests include metadata management, service chaining and semantic interoperability in the context of spatial data infrastructures.

F.J. Zarazaga-Soria, PhD, is an Associate Professor at the Computer Science Department of the University of Zaragoza (Spain). His research interests include metadata management, information retrieval and ontologies in the context of spatial data infrastructures.

1 Introduction

The importance of interoperability among systems, the ability of two or more systems or components to exchange information and to use the information that has been exchanged (IEEE, 1990), has been progressively increasing over the last years. Interoperability is not simply a technical issue concerned with linking up computer networks. It goes beyond this to include the sharing of information between networks and the rearrangement of administrative processes.

Therefore, it allows organisations to re-use information both internally and with their business partners, and to cooperate in achieving agreed objectives. Thus, interoperability helps organisations, in both the public and the enterprise sectors, to be more effective in the achievement of their goals.

(Draft) International Journal of Metadata, Semantics and Ontologies. 2006, vol. 1, nº 3, p. 183-188.

One of its specific aspects, which is concerned with the ability to access, consistently and coherently, similar (though autonomously defined and managed) classes of digital data, objects and services distributed across heterogeneous repositories (Nogueras-Iso et al., 2004), is known as semantic interoperability.

With little doubt, the most obvious way to broaden the opportunities for interoperability is by making the information stored, which is known as metadata (or simply descriptors), exchangeable. This may be carried out by promoting a commonly understood set of descriptors that helps to unify other data content standards.

The tendency of current cataloguing systems is to interchange metadata in XML according to the specific standard required by each user on demand, that is to say, providing different views of the same metadata. Furthermore, metadata schemas from different domains are not usually semantically distinct but overlap and relate to each other in complex ways. As a consequence, the semantic interoperability has to deal with the equivalences between those descriptions.

According to scientific research, it seems that there exist two main approaches in order to tackle this problem: solutions that are based on the use of ontologies and solutions that are based on the creation of specific crosswalks for one-to-one mapping. This work will propose a hierarchical one-to-one mapping solution for improving the semantic interoperability.

The remainder of this paper is structured as follows: in Section 2, related work from the semantic interoperability domain is reviewed. In Section 3, our proposal for improving the semantic interoperability is described. In Section 4, a functional kernel which exploits this proposal with success is described.

2 Related work

Since the emergence of the internet, a great deal of effort has been invested in the development of metadata vocabularies to enable the exchange and discovery of information across different applications and domains. Metadata vocabularies such as Dublin Core (ANSI, 2001), MARC (US Library of Congress, 2004), FGDC (FGDC, 1998), provide standardised sets of descriptive elements to enable the exchange of resources for specific applications or domains. Although these standards enable interoperability within domains, they introduce the problem of incompatibility between disparate and heterogeneous metadata schemas or schemas across domains.

A literature survey reveals many different proposals for improving interoperability between domain-specific vocabularies, thesauri and ontologies in the context of information retrieval and exchange.

On the other hand, there are three main scenarios in which interoperability among metadata schemas is required, according to Hunter (2001):

- to enable a single search interface across heterogeneous metadata schemas
- to enable the integration or merging of descriptions which are based on complementary but possibly overlapping metadata schemas or standards
- to enable different views of the one underlying and complete metadata schema, depending on the user's interest, perspective or requirements.

In the next subsections, the two approaches, commented on above, will be reviewed in detail.

2.1 *The ontology-based approach for semantic interoperability*

In the information systems and knowledge representation field, the ontology concept denotes a knowledge model that represents a particular domain of interest. These kinds of solutions are based on these models since they may help to define a common ground between different information communities.

In this sense, the work developed in the OBSERVER system (Mena, 1998) provides an architecture for query processing in global information systems that supports interoperation using ontologies. Besides, each ontology defines the terms used in the concept of a specific data repository, i.e. the ontology compiles the terms which are later mapped to the specific data structures (names of entities and attributes). An interontology contains the relationships which relate the terms in the different ontologies, and which enable the translation of the user query to the specific ontology of each distributed repository.

Another interesting research work (Hunter, 2001) implements the ontology by means of a thesaurus (MetaNet) applied to the ABC model. Its main objective is to provide the semantic knowledge required in order 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 searches in WordNet of the core terms used in the different domains. Furthermore, its thesaurus has been implemented by using technology, RDF (Resource Description Framework, (Manola and Miller, 2004)) and RDFS (RDF Schema (Brickley and Guha, 2004)), borrowed from the semantic web field which is in fact a closely related concept. However, other proposals (Nogueras-Iso et al., 2005) outline the limitations of RDFS, since it does not provide mechanisms for specifying general axioms (rules that permit additional reasoning) which appear at most artificial intelligence ontologies. The result of this research is the SHOE language (Heflin and Hendler, 2000) which attempts to solve this lack of functionality.

(Draft) International Journal of Metadata, Semantics and Ontologies. 2006, vol. 1, nº 3, p. 183-188.

As stated in Nogueras-Iso et al. (2005), all these are flexible solutions for interoperability improvement. Nonetheless, this ambitious aim of flexibility may also imply a lack of accuracy in the mappings performed. As far as we know, current ontology-based solutions do not consider the local structural constraints imposed by the different specific domains (i.e., parent-child relationships, cardinality-occurrence, constraints, data typing, ...)

2.2 The crosswalk-based approach for semantic interoperability

This set of solutions uses software components which map the relationships and equivalences between two or more metadata schemas. These software components are called crosswalks. There is much experience in developing mappings among several standards and different domains, which it is really an important fact in order to establish the semantic equivalences properly. For instance, interesting collections of links to metadata crosswalk initiatives can be found through the websites of the UK Office of the Library and Information Networking and the Metadata Architecture and Application Team of the National Digital Archives Program in Taiwan. There, it is possible to find several mappings among the main metadata standards (specially those used for library metadata): from MARC standards to Dublin Core; from Dublin Core to EAD (Encoded Archival Description) (US Library of Congress, 1998); from Dublin Core to GILS (a Z39.50 metadata profile for the US Government Information Locator Service); or from Dublin Core to GCMD DIF (Directory Interchange Format (FGDC, 1998)).

Other works, such as the CORC (Cooperative Online Research Catalogue) project (Chandler et al., 2000), have also proposed the conversion of CSDGM to more generic standards like MARC or Dublin Core.

On the other hand, there exists an interesting proposal (Nogueras-Iso et al., 2005) in which the attention is focused in the crosswalk process creation.

3 Our hierarchical solution

The three required scenarios, mentioned in the previous section, guided our proposal for the interoperability improvement. Its main idea consists in defining or choosing a general purpose metadata schema as a common core.

Next, one-to-one semantic correspondences between the system metadata schemas and such a general description have to be established. The system, at this point, can build as much functionality as possible through this common core view. Therefore, new metadata schemas introduced in the system do not necessarily have to involve code rewriting, providing that appropriate correspondences can be established between those new schemas and the common core. On the other hand, new metadata schemas do not have to be directly mapped to the common core, but to other schemas which were introduced previously. The related set of metadata schemas can be seen as a metadata profile hierarchy.

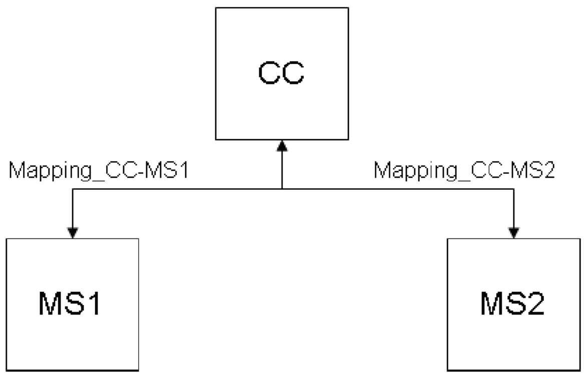
The following subsection will try to explain in detail this technique through a simple example. Then, a metadata profile hierarchy, based on the Dublin Core, will be proposed.

3.1 A simple example

Let us suppose a system with two metadata schemas, MS1 and MS2, in which MS1 presents a very rich description with hundreds of elements and MS2 presents a more general description with a considerably smaller number of elements. Our solution would consist of:

- Defining the Common Core (CC).
- Establishing semantic correspondences between the existing metadata schemas and this new general one (*Mapping CC-MS1, Mapping CC-MS2*) (see Figure 1).
- The CC has to be concrete enough in semantics in order to provide the system with a useful data view. Besides, for this same reason, it is desirable that the schemas of MS1 and MS2 are semantically richer than the information of the CC. Were it not the case, the services of the system through the common core view would probably not provide the quality results desired.

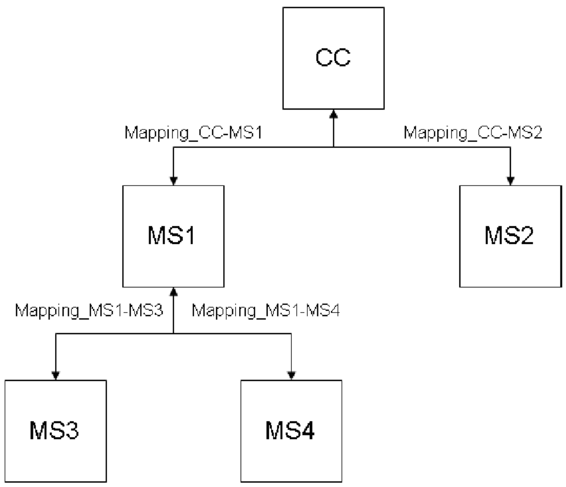
Figure 1 Correspondences among the metadata schemas



Let us suppose now that two new metadata schemas MS3 and MS4 have to be integrated into the system and that they are based on the metadata schema MS1, since they extend its elements. As a consequence, MS3 and MS4 might be mapped directly to MS1 and thus, the metadata schema hierarchy represented in Figure 2 is obtained.

(Draft) International Journal of Metadata, Semantics and Ontologies. 2006, vol. 1, nº 3, p. 183-188.

Figure 2 A simple metadata profile hierarchy



Another remarkable aspect is the way in which the semantic correspondences can be implemented. They will clearly depend on the metadata schemas involved. There are, however, two main possibilities:

- As reviewed in the previous section, there are a lot of experienced techniques in mapping some to others by using crosswalks. This may be helpful when dealing with heterogeneous metadata schemas which may represent the most frequent situation.
- In case that some bottom-hierarchy metadata schemas extend a top-hierarchy schema, there will be no need for crosswalks.

By establishing those semantic levels, a metadata profile hierarchy was obtained. In this hierarchy, top schemas are semantically more general than bottom ones.

The benefits of such an approach are considerable, regarding the requirements which guided the design of our proposal:

- The single interface across heterogeneous metadata schema could be obtained by designing the search interface through the common core schema.
- The integration of descriptions that are based on complementary but possible overlapping metadata standards are achieved by introducing the schemas into the hierarchy.
- The different views of the information may be achieved by designing specific crosswalks across standards. Nevertheless, it can be somehow useful the fact that all the metadata schemas have the general description in common.

3.2 A metadata profile hierarchy based on Dublin Core

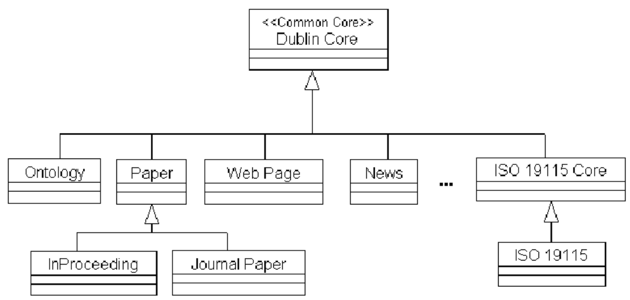
Dublin Core seeks to promote a commonly understood set of descriptors to help facilitate interoperability across disciplines (DCMI, 2004; ISO, 2003). Some of its well-designed features are its simplicity and extensibility as well as its objective of facilitating discovery of electronic resources. These features make Dublin Core an ideal candidate to be incorporated in our metadata profile hierarchy as the common core.

Certain mechanisms provided by the Dublin Core Metadata Initiative (DCMI) such as application profiles permit describing a wide range of heterogeneous resources by extending and adapting the semantics of Dublin Core.

However, in certain situations, there is no possibility to extend the semantics and the correspondence has to be carried out by using a crosswalk.

Figure 3 shows a possible metadata profile hierarchy based on Dublin Core. UML was used to represent it graphically and the inheritance relation should be interpreted in terms of semantic description. Dublin Core is the standard at the top of the hierarchy. Then, as can be seen, several DC application profiles have been defined to describe web pages, papers, news and even ontologies. All these schemas were created by means of the mechanisms provided by Dublin Core (application profiles). However, in order to establish the mapping between Dublin Core and ISO 19115 core a crosswalk had to be designed and implemented, since those standards are extremely different. Several guidelines from Zarazaga-Soria et al. (2003a, 2003b, 2003c), were followed in this process.

Figure 3 A Dublin Core-based hierarchy



4 Exploiting the hierarchical solution

This section illustrates an actual example in which our metadata profile hierarchy has been tested with excellent results in the Spatial Data Infrastructures environment. There are several SDI components which benefit from this solution, though the most relevant ones are related to CatServer, a server of metadata.

4.1 CatServer: a functional kernel

CatServer is a functional kernel which provides catalogue services for XML-coded metadata. It is being used for the Spatial Data Infrastructures (SDIs) development. It manages any XML-coded metadata, providing that the standard of the metadata can be semantically introduced in a metadata profile hierarchy based on Dublin Core.

CatServer's design and implementation have been based on the General Catalogue Interface Model specifications by the Open Geospatial Consortium (OGC) (Nebert and Whiteside, 2004). According to OGC, "the General

(Draft) International Journal of Metadata, Semantics and Ontologies. 2006, vol. 1, nº 3, p. 183-188.

Catalogue Interface Model provides a set of abstract service interfaces that support the discovery, the access, the maintenance and the organisation of catalogues of geospatial information and related resources". Consequently, our system provides management services – which allow the administrator to maintain and organise the metadata – and discovery services – which permit querying the information. In addition, a two-operation session component (initialise and close) is supplied for the interaction activity between the server and a client.

A more detailed description of CatServer's functionality and several techniques developed for performance improvement can be found at Tolosana-Calasanz et al. (2005).

4.2 CatServer: a SDI component

Several SDI architectural elements are built over metadata retrieval services of extremely high importance. For this reason, the design of a metadata retrieval kernel will have strong influence on upper-layer components. As discussed above, CatServer was designed over the Dublin Core metadata profile hierarchy and, therefore, its core functionality is built through the view provided by Dublin Core. CatServer simplifies the design of upper-layer components, foments software reusability and simplifies the software management, since it has the interoperable hierarchy schemas to lean on.

Consequently, CatServer could be integrated in some SDI components such as geographical dataset catalogues, services catalogues, geocoders, web Feature Servers and gazetteers.

The following enumeration reflects that briefly:

- According to OGC a catalogue is a component that supports the ability to publish and to search collections of metadata for data, services and related information objects. If this definition is restricted to geographical metadata, we have a geographical metadata catalogue which stores descriptions of the geographical information in the SDI. Two metadata catalogues which have been built using CatServer's technology are the geographical catalogue of the Spanish SDI accessible at http://www.idee.es, which stores about thirty thousand metadata, and the geographical catalogue of the Zaragoza City Council accessible at http://idezar.unizar.es with about two hundred metadata.
- A catalogue of services stores descriptions of services supplied somewhere. The Spanish SDI services catalogue (http://www.idee.es) uses CatServer as well.
- A Gazetteer service is a network-accessible service that retrieves the known geometries, for one or more features, given their associated well-known feature identifiers (text strings). The Spanish SDI gazetteer (accessible at http://www.idee.es) is built upon CatServer's technology and manages about one million metadata records.
- A Geocoder Service is a network-accessible service that transforms a description of a feature location, such as a place name, street address or postal code into a normalised description of the location, which includes a coordinate geometry. An example of a Geocoder which currently has CatServer is the Zaragoza City Council Geocoder (http://idezar.unizar.es) which stores about six thousand metadata.
- A portal is a website on the internet which people use to search. Sometimes portals access database repositories to obtain the information. That is the case of the Spanish SDI portal which (http://www.idee.es) provides information about news, web pages and organisations by means of CatServer.

Additionally, CatServer will be used in the development of new components such as a web Ontology Server (Lacasta et al., 2005), a catalogue of images and other components in which retrieval services are the base of the functionality.

5 Conclusions

This paper has described a technique for the improvement of metadata semantic interoperability in systems with heterogeneous metadata schemas. This technique consists of erecting a semantically general metadata standard, called common core, and building the most general functionality of the system through this general view. Then, semantic correspondence is established between other metadata schemas, which have to be integrated and managed by the system, and the common core. Those schemas, at the same time, can be general descriptions for other new schemas and so a metadata profile hierarchy is generated.

Besides, since Dublin Core is a metadata standard which promotes a commonly understood set of descriptors and is simple and extensible; a metadata profile hierarchy, in which the root schema is Dublin Core, was proposed. Systems storing heterogeneous metadata schemas obtain excellent advantages when using the Dublin Core hierarchy. This is mainly because this technique simplifies the design of upper-layer components, facilitates software reusability and simplifies the software maintenance, since it has the interoperable hierarchy schemas to lean on.

Over a Dublin Core metadata profile hierarchy, a set of technological components has been developed. These components try to exploit the above mentioned benefits from such a hierarchy. Undoubtedly, one of the most relevant components among these is CatServer. Its utilisation in industrial purpose products, like the systems described here, demonstrates the viability of the work presented in this paper in the software industry domain.

Acknowledgements

The basic technology of this work has been partially supported by the Spanish Ministry of Science and

(Draft) International Journal of Metadata, Semantics and Ontologies. 2006, vol. 1, nº 3, p. 183-188.

Technology through the project TIC2003-09365-C02-01 from the National Plan for Scientific Research, Development and Technology Innovation.

References

ANSI (2001) *The Dublin Core Metadata Element Set*, ANSI/NISO Z39.85-2001, American National Standards Institute, 4733 Bethesda Avenue, Suite 300, Bethesda, MD 20814, USA, http://www.niso.org/standards/resources/Z39-85.pdf.

Brickley, D. and Guha, R.V. (Eds.) (2004) *RDF Vocabulary Description Language 1.0: RDF Schema*, W3C, W3C Recommendation 10 February 2004. http://www.w3.org/TR/2004/REC-rdf-schema-20040210/.

Chandler, A., Foley, D. and Hafez, A.M. (2000). 'Mapping and converting essential federal geographic data committee (FGDC) metadata into MARC21 and dublin core. Towards an alternative to the FGDC clearinghouse', *D-Lib Magazine*, Vol. 6, No. 1, http://www.dlib.org/dlib/january00/chandler/01chandler.html.

DCMI (2004) *Homepage of the Dublin Core Metadata Initiative (DCMI)*, http://www. dublincore. org.

FGDC (1998) *Content Standard for Digital Geospatial Metadata, Version 2.0. Document FGDC-STD-001-1998*, *Federal Geographic Data Committee*, Metadata Ad Hoc Working Group, U.S. Geological Survey, 590 National Center Reston, Virginia 20192, USA, http://www.fgdc.gov/metadata/csdgm.

Heflin, J. and Hendler, J. (2000) 'Semantic interoperability on the web', in *Proceedings of Extreme Markup Languages 2000. Graphic Communications Association*, Alexandria, VA, pp.111–120, http://www.cs.umd.edu/projects/plus/SHOE/pubs/extreme2000.pdf.

Hunter, J. (2001). 'MetaNet – a metadata term thesaurus to enable semantic interoperability between metadata domains', *Journal of Digital information*, Vol. 1, No. 8, http://jodi.ecs.soton.ac.uk/Articles/v01/i08/Hunter/.

IEEE (1990) *IEEE Standard Computer Dictionary: A Compilation of IEEE Standard Computer Glossaries*, Institute of Electrical and Electronics Engineers, New York, NY.

ISO (2003) *Information and Documentation – The Dublin Core Metadata Element Set*, ISO 15836:2003, International Organization for Standardization.

Lacasta, J., Muro-Medrano, P., Nogueras-Iso, J. and Zarazaga-Soria, F. (2005) 'Web ontology service, a key component of a spatial data infrastructure', *Proceedings of the 11th EC GI & GIS Workshop, ESDI Setting the Framework*, Alghero, Italy.

Manola, F. and Miller, E. (Eds.) (2004) *RDF Primer*, W3C, W3C Recommendation 10 February 2004. http://www.w3.org/TR/2004/REC-rdf-primer-20040210/.

Mena, E. (1998) *OBSERVER: An Approach for Query Processing in Global Information Systems based on Interoperation across Pre-existing Ontologies*, PhD Thesis, University of Zaragoza, Maria de Luna 1, 50018 – Zaragoza, Spain.

Nebert, D. and Whiteside, A. (Eds.) (2004) *OpenGIS – Catalogue Services Specification (version: 2.0)*, Open Geospatial Consortium Inc. (Open GIS Consortium Inc), OpenGIS Project Document 04-021r2.

Nogueras-Iso, J., Zarazaga-Soria, F.J. and Muro-Medrano, P.R. (2005) *Geographic Information Metadata for Spatial Data Infrastructures – Resources, Interoperability and Information Retrieval*, Springer-Verlag, New York, Inc. Secaucus, NJ, USA.

Nogueras-Iso, J., Zarazaga-Soria, F.J., Lacasta, J., Béjar, R. and Muro-Medrano, P.R. (2004) 'Metadata standard interoperability: application in the geographic information domain' *Computers, Environment and Urban Systems*, Vol. 28, No. 6, pp.611–634.

Tolosana-Calasanz, R., Portols-Rodrguez, D., Nogueras-Iso, J., Muro-Medrano, P.R. and Zarazaga-Soria, F. (2005) 'Catserver: a server of GATOS', *Proceedings of AGILE Conference 2005*, Estoril, Portugal, ISBN 972-8093-13-6, pp.359–366.

US Library of Congress (1998) *Encoded Archival Description (EAD) DTDi*, http://lcweb.loc.gov/ead/.

US Library of Congress (2004) *MARC Standards, Network Development and MARC Standards Office*, http://www.loc.gov/marc/.

Zarazaga-Soria, F.J., Nogueras-Iso, J. and Ford, M. (2003a) *Dublin Core Spatial Application Profile. CWA 14858, CEN/ISSS Workshop*, Metadata for Multimedia Information, Dublin Core.

Zarazaga-Soria, F.J., Nogueras-Iso, J. and Ford, M. (2003b) *Guidance Material for Mapping between Dublin Core and ISO in the Geographic Information Domain. CWA 14856, CEN/ISSS Workshop*, Metadata for Multimedia Information, Dublin Core.

Zarazaga-Soria, F.J., Nogueras-Iso, J. and Ford, M. (2003c) *Mapping between Dublin Core and IS 19115 Geographic Information Metadata. CWA 14857, CEN/ISSS Workshop*, Metadata for Multimedia Information, Dublin Core.