Towards an Interoperability Framework for Metadata Standards

Mikael Nilsson
KMR Group, NADA, Royal Institute of Technology, Stockholm, Sweden
mini@nada.kth.se

Pete Johnston
Eduserv Foundation, United Kingdom
pete.johnston@eduserv.org.uk

Ambjörn Naeve
KMR Group, NADA, Royal Institute of Technology, Stockholm, Sweden
amb@nada.kth.se

Andy Powell
Eduserv Foundation, United Kingdom
andy.powell@eduserv.org.uk

1. Abstract

This paper presents a conceptual metadata framework intended to support the development of interoperable metadata standards and applications. The model rests on the fundamental concept of an "abstract model" for metadata, as exemplified by the DCMI Abstract Model, and is based on concepts and ideas that have developed over the years within the Dublin Core Metadata Initiative.

The model thus incorporates the concepts of metadata vocabularies, schemas, formats and application profiles into a single framework that can be used to analyse and compare metadata standards, and aid in the process of harmonization of metadata standards. In this paper, the model is used to compare the Dublin Core metadata standard and the IEEE LOM standard. Some fundamental differences between the two standards are discussed briefly, and important gaps in the current set of Dublin Core metadata specifications are noted.

2. Background

The publication of the DCMI Abstract model (DCAM) (Powell *et al*, 2005) in March 2005 marked a major milestone for the Dublin Core community and the DCMI. In developing the DCAM, the DCMI has shown its intention to gradually move away from dealing primarily with the "core" set of terms, moving instead to dealing primarily with community-specific application profiles, each defined within a common framework

(Baker, 2005). Within such a framework, metadata terms from different and independent communities can coexist, allowing for a controlled mix-and-match of community- and application-specific metadata constructs.

This framework of the DCMI is still not formalized, even if considerable experience and documentation regarding the necessary components of such a framework have been collected over the years. It is the intention of this paper to introduce an overarching model to describe the components of this framework, to serve as a possible basis for further formalization, and to highlight the strong and weak points of the current situation.

The model proposed in this paper is also intended to serve as a guide to understanding the conceptual relationships between the structures of the many different metadata standards currently in use. We will demonstrate this by using the model as a tool to compare the structure of the Dublin Core metadata framework with the IEEE LOM standard. Although the model has its origins in the Dublin Core metadata framework, we believe the model has a substantially more general applicability.

This attempt at designing a framework share some features with the Warwick framework (Lagoze, 1996), although that framework had more to do with packaging of metadata than with interoperability *within* metadata descriptions. The RDF suite of specifications, however, follow a more similar pattern to the framework presented here.

3. Components of the Framework

In this section, a set of components of a framework for Dublin Core metadata are presented: the *abstract model*, *metadata formats*, *metadata vocabularies*, the *vocabulary model*, *application profiles* and the *profile model*. Not all of these components correspond to existing formalized concepts or recommendations, but represent abstractions based on current usage of Dublin Core metadata and on current directions in metadata interoperability..

3.1 The abstract model

The abstract model specifies the concepts used in the framework, the nature of terms and how they combine to form a metadata description. The abstract model is the key used by a metadata application to unlock the secrets of a metadata expression given in a specific format, thus making it possible for a single standard, though expressed in several different formats, to still be understood in a uniform way by users and applications. An early effort to produce such framework for Dublin Core was presented in Bearman, Miller, Rust, Trant and Weibel (1999).

Basing metadata on an abstract model carries a number of important benefits

> Clear guidelines on how to create and maintain customized metadata vocabularies. There is currently some confusion on how to best produce vocabularies, much due to the differing fundamental principles for vocabularies in the different metadata standards.

Fine-grained control over relationships between terms from different standards, including refinement and partial mappings. Automation of interoperable metadata management systems will be greatly improved, and metadata vocabularies will be able to build upon each other.

A single set of format bindings. Contrast this with the current situation, which requires every metadata standard to have its own set of format bindings. This will make life easier not only for metadata standardization bodies, but also for applications that will only need to support one format.

A single framework for extending and combining metadata from different standards. This will enable standardized principles for the construction of interoperable application profiles.

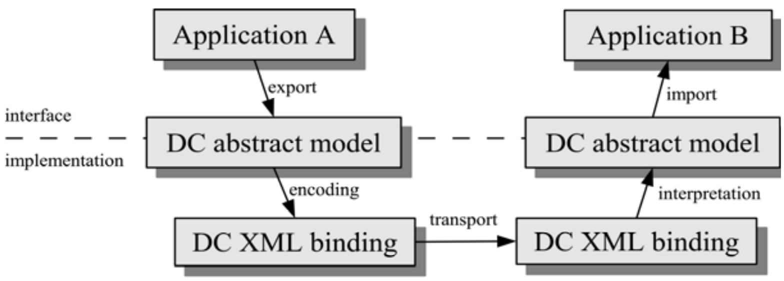

Figure 1. The process of encoding/interpretation of metadata within the framework of an abstract model.

A single storage and query model for very different types of data and vocabularies. For example, storing metadata from different specifications in the same database will become more straightforward.

Implementing searching that includes dependencies between metadata expressed in different schemas will be simplified.

The DCMI Abstract Model was published in March 2005, as a substantial reformulation and clarification of the "DCMI Grammatical Principles" (DCMI Usage Board, 2003).

3.2 Metadata Formats

Metadata exchange is usually performed using one of several *metadata formats* or *bindings*. A binding is constructed by specifying how each kind of concept in the abstract model is to be encoded in a particular format. Conversely, the binding also specifies how to interpret data given in a specific format in terms of the abstract model. For example, when interpreting a metadata record that uses the Dublin Core XML binding, an XML element called "dcterms:modified" used in a particular place in the XML document represents a property, and the value "dcterms:W3CDTF" of a particular XML attribute represents a syntax encoding scheme for the value string "2001-07-18" occurring as XML content in a particular position.

This fundamental process of *encoding/interpretation* is described in Figure 1. Application A uses the DCMI Abstract Model to represent some metadata about a resource. This metadata is encoded using the Dublin Core XML binding, and transferred to another application. Application B will use the rules of the Dublin Core XML binding to interpret the XML data in terms of the DCMI Abstract Model. This representation of the metadata can then be used in the application.

When two applications want to exchange Dublin Core metadata, they understand metadata through the lens of the abstract model. The abstract model functions as an opaque interface, an API, to the metadata. In practice, the exchange is realized using one of the Dublin Core bindings, but the details of the formats are of no interest to the applications, which instead analyse the metadata in terms of the interface given by the abstract model.

Note that it is possible to produce applications that process metadata without regard to the abstract model. Such *ad-hoc processing* of metadata records requires that the precise content of the records is well-known in advance, which is the case in many systems where extensibility, modularity and refinements are not design requirements. In contrast, the kind of *interoperable processing* based on the abstract model described above is necessary when an application needs to be prepared for metadata constructs that do not fall within the limits of such a precise, pre-conceived description. Thus, it should be clear that interoperable processing is a basic prerequisite for metadata interoperability.

3.3 Metadata Vocabularies

In a metadata standardization framework supported by a common abstract model, the work of defining new metadata terms is much reduced. As the "grammatical structure" of metadata descriptions is already laid down, the only thing needed is to fill the abstract model with specific terms.

The main benefit of developing vocabularies in a common framework is that reuse across standards will be much simpler. As an example, many elements in the LOM standard are not specific to learning, and have similar counterparts in other standards. If they were defined to fit a common framework, the LOM elements would be made into a fully-fledged element vocabulary capable of being extended, refined and semantically annotated. The semantic relationships to terms in these other standards can be made explicit and machine-processable.

The notion of a metadata "vocabulary" is somewhat ambiguous and is used differently in different standards. In Dublin Core metadata, a vocabulary can be one of two things:

1. A *value vocabulary*, consisting of concepts from a controlled set as specified by a vocabulary encoding scheme. For example, the "dcterms:LCSH" vocabulary encoding scheme refers to the vocabulary formed by the set of Library of Congress subject headings.
2. An *element vocabulary*, consisting of a set of metadata properties together with their definitions. For example, the Dublin Core Element Set, consisting of the 15 original Dublin Core elements (*dc:title*, *dc:subject*, etc.), is such a vocabulary.

Element vocabularies and value vocabularies have fundamentally different characteristics. While value vocabularies are used to construct taxonomies and thesauri that describe relationships between concepts in terms of broader/narrower, containment etc, element vocabularies are used to construct application profiles, schemas and ontologies that describe how metadata instances are to be constructed.

3.4 The Vocabulary Model

In order to do define machine-processable vocabularies for use in the framework, a language for describing metadata vocabularies is necessary. Such a vocabulary language and its corresponding *vocabulary model* specifies how element and value vocabularies should be described in order to conform to the abstract model.

The Dublin Core Vocabulary model has not yet been formalized, but embryos such as Baker (2003) exist. DCMI has a history of using RDF Schema (Brickley *et al* 2004) as a basis for its machine-readable term declarations, which is useful for describing both element and value vocabularies.

3.5 Application Profiles

Implementers of metadata standards should be able to assemble the components that they require for some particular set of functions - and if that means drawing on components that are specified within different metadata standards, that should be possible – safe in the knowledge that the assembled whole can be interpreted correctly by independently designed applications. Duval et al (2002) employ the metaphor of the Lego set to describe this process: an application designer should be able to “snap together” selected “building blocks” drawn from the “kits” provided by different metadata standards to build the construction that meets their requirements, even if the kits that provide those blocks were created quite independently.

Heery and Patel (2000) present a compelling vision of metadata implementers “mixing and matching” “data elements”, constructing application profiles by selecting from the sets of “data elements” provided by metadata standards and by other implementers. Such application profiles are fundamental to a modern metadata framework.

Application profiles may contain specifications of the related descriptions of several kinds of related resources, such as a collection, the items it consists of and the associated contributors. Thus, such a specification is a multi-layered structure of some complexity, that can not, in general, be captured by a flat list of properties.

3.6 Profile Model

A common model for expressing application profiles will be a necessary building block for the construction of reusable application profiles. The model must not be tied to a specific metadata format, but must operate at the level of the abstract model, so that the application profile can be reused in all metadata formats.

Promising work on machine-processable application profiles can be seen in, e.g., “Guidelines” (2005). There are also other initiatives for such frameworks, but none are yet in widespread use.

3.7 Model diagram

The relationships between these concepts are depicted in Figure 2.

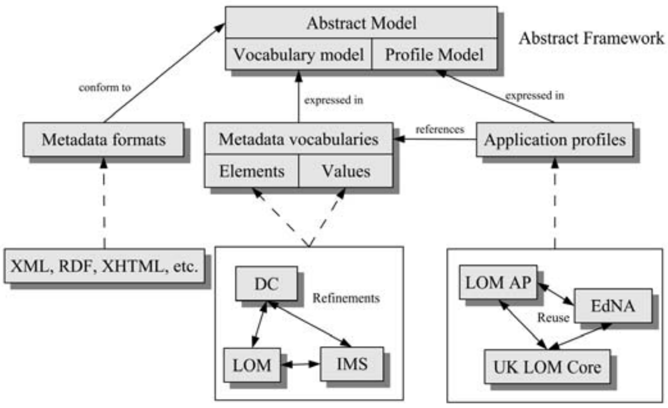

Figure 2. A model of the Dublin Core metadata framework.

4. Implications for Metadata Standards

In light of the presented model, it seems clear that the current use of the term "metadata standard" or "metadata schema" will need refinement. These terms are often used interchangeably to describe one of the following:

The overarching abstract model standard. This will also include a specification for how to express the semantics of vocabularies adhering to the abstract model (the vocabulary model) as well as a specification for how to express application profiles in a machine-processable way (the profile model).

Metadata format specifications. These will include bindings of the abstract model to a set of formats and systems, including XML, database layouts, programming languages, etc., as well as translations or mappings to other knowledge representation systems such as RDF. Such specification are closely tied to the abstract model.

Metadata vocabularies. These will include metadata terms from different communities. The Dublin Core terms, the LOM elements and so on are examples of metadata element vocabularies, and a large set of value vocabularies also fit into this category.

Application profiles. These will specify usages of metadata vocabularies in complex combinations.

4.1 Reusing "Elements" Across Metadata Standards

The CORES Resolution (Baker and Dekkers, 2002), which has been signed by both the IEEE LTSC and the Dublin Core Metadata Initiative, encouraged the owners of metadata standards to assign URI references to their "elements", the "units of meaning comparable and mappable to elements of other standards", but it did not specify what "comparable and mappable" meant. As a consequence the owners of different standards assigned URIs to "elements" that are created within different abstract models and uses metadata formats that rely on those incompatible abstract models for their meaning and interpretation. The assignment of a URI to an "element" means that it can be unambiguously cited, but it does not change the nature of the "element": and it does not mean that it is meaningful to use a URI for a LOM element as, e.g., a property URI in a Dublin Core metadata description. Similar incompatibilities have been noted between, e.g., RDF and MPEG-7 (van Ossenbruggen, Nack and Hardman, 2004 and Nack, van Ossenbruggen and Hardman, 2005).

The analysis in Nilsson *et al* (2006) shows that we must not confuse the components used in a metadata format and the constructs in the abstract model. The components in a metadata format, such as “element URIs” may seem to be similar and compatible, but in reality they belong to completely different frameworks that might not be compatible.

Thus, the notion of reusing “elements” between metadata standards and formats using incompatible frameworks is fundamentally flawed. While assigning URIs for the component parts of a metadata standard is clearly a worthwhile effort in other ways, this does not really address the fundamental issue when creating interoperable metadata standards, namely the compatibility of their respective frameworks, and in particular, their abstract models.

In conclusion, we see that in order to reuse components of different standards in a machine-processable way, the following criteria must be met:

1. The components must be unambiguously identified, so that components from different sources can be clearly distinguished and their origins can be separated. This is addressed by the CORES resolution.
2. The components must adhere to compatible abstract models. There is currently no resolution to address this, although the Dublin Core – IEEE Memorandum of Understanding (“Memorandum”, 2000) points in this direction.
3. A metadata format must be used that allows for consistent interpretation of the components with respect to their respective abstract models.

It should be clear that using a common metadata framework like the one described above would fulfil these requirements.

5. Applying the framework to LOM, Dublin Core and the Semantic Web

Trying to understand what parts of the presented model exists in Dublin Core today, and comparing that to IEEE LOM, we arrive at the following table. Note that by “Dublin Core framework” we refer to the complete set of DCMI specifications, and similarly for LOM.

Framework concept	*Dublin Core framework*	*LOM framework*	*Semantic Web framework*
Abstract Model	DCMI Abstract Model	Implicit in LOM Data Model	RDF Concepts and Abstract Syntax
Metadata Formats	XML, RDF and HTML bindings	XML binding	RDF/XML syntax, N-triples, etc.
Metadata Element Vocabularies	DCMES, large set of external properties and encoding schemes	LOM Data Model includes element vocabulary, various extensions to LOM	Many external element vocabularies

Framework concept	*Dublin Core framework*	*LOM framework*	*Semantic Web framework*
Metadata Value Vocabularies	DCMIType vocabulary. Many external value vocabularies	LOM Data Model includes several basic value vocabularies, many external vocabularies	Many external value vocabularies
Vocabulary Model	Not formalized, but see Baker (2003)	Not formalized	RDF Vocabulary Description Language
Application Profiles	Some published by DCMI, many external application profiles	LOM Data Model includes basic application profile, many external application profiles.	Many in the form of ontologies
Profile Model	Not formalized, CWA...	Not formalized	Possibly OWL, the Web Ontology Language

A few comments on this table:

Not all parts are formalized. The DCMI is slowly progressing towards formalizing the complete abstract framework, including abstract model, vocabulary model and profile model. Similar efforts are not under way in LOM.

The most mature parts are certainly value vocabularies, where many external sources exist. Dublin Core metadata element vocabularies are also relatively mature. To some extent, and to some extent application profiles have some maturity, even though there is still a certain amount of confusion in the community regarding the precise nature of an application profile.

In spite of the existence of many application profiles and metadata vocabularies, no formal model is usually followed in their design.

LOM has a very weak notion of element vocabularies, as noted in Nilsson *et al* (2006), that does not support URI identification of elements.

The LOM Data Model defines, in a single standard, both an abstract model (implicitly, at least), a metadata element vocabulary, a set of metadata value vocabularies, and a basic application profile. This is one way of expressing the well-known "monolithic" nature of the LOM standard.

In short, the above table can be used to analyse and compare metadata standards, and understand how they relate to different aspects of the Dublin Core universe.

6. Looking Forward

We have presented an overarching interoperability framework for metadata standards, based on the implicit structure of Dublin Core metadata standardization. By applying the framework to LOM and Dublin Core metadata, we learn about differences between the metadata standards and deficiencies in their respective frameworks.

The authors believe that the Dublin Core Metadata Initiative would be greatly helped by applying this understanding to improve its documentation and vision of metadata interoperability. In particular, a high-level framework for Dublin Core metadata has not been proposed since the Warwick framework, and it is now time to revisit the overall structure of metadata standardization. Luckily, as the analysis shows, there is some coherency in the current set of DCMI specifications, though much of it remains implicit.

Another issue is that of interoperability with other metadata standards. The authors have little hope that deep integration between metadata standards can be made a reality unless they adhere to a common framework. Unfortunately, a thorough analysis shows (Nilsson *et al,* 2006) that there are fundamental incompatibilities between frameworks such as the LOM framework and that of Dublin Core. On the other hand, the framework of RDF and the semantic web share many features with Dublin Core, and advanced interoperability between those frameworks has already been demonstrated.

The authors therefore argue that the long-term solution is to proceed towards a *shared* metadata framework. Having all metadata standards expressed using a common abstract model, or at least using compatible abstract models, would greatly increase interoperability in several ways. It would also create a natural separation between the specification of the structure of metadata descriptions and the declaration of metadata terms used within that structure, so that both LOM vocabularies and Dublin Core vocabularies would appear as metadata vocabularies within that one structure.

There are already initiatives to develop a common abstract model that covers both LOM and Dublin Core, but unfortunately it seems to be impossible to arrive at such a model without re-engineering at least one standard to retrofit it to the new abstract model, which naturally is a major undertaking. But it seems clear that this is the only long-term solution to the interoperability problems we have seen here. Reaching out to embrace the other important metadata standards, such as MODS, MPEG-7 and the IMS set of standards is then the logical next step. In addition, great care must be taken to ensure that such an abstract model does not conflict with the emerging metadata format for the Web: RDF.

The basis of the envisioned metadata standardization framework is the abstract model. The incompatibilities of abstract models are the most significant stumbling blocks for metadata interoperability. The development of a common abstract model for metadata is therefore of central importance if we are ever going to experience true metadata interoperability.

7. References

Baker, T. (2003), DCMI Usage Board Review of Application Profiles. Retrieved July 1 2005, from http://dublincore.org/usage/documents/profiles/

Baker, T. & Dekkers, M., (2002), CORES Standards Interoperability Forum Resolution on Metadata Element Identifiers. Retrieved July 1, 2005, from http://www.cores-eu.net/interoperability/cores-resolution/

Baker, T. (2003), Usage Board Application Profile (draft). Retrieved April 30, 2006, from http://dublincore.org/usage/meetings/2003/06/Usage-ap.html

Baker, T. (2005), Diverse Vocabularies in a Common Model: DCMI at ten years, Keynote speech, DC-2005, Madrid, Spain. Retrieved April 30, 2006, from http://dc2005.uc3m.es/program/presentations/2005-09-12.plenary.baker-keynote.ppt

Bearman, D., Miller, E., Rust, G., Trant, J. & Weibel, S. (1999), A Common Model to Support Interoperable Metadata, *D-Lib Magazine*, January 1999. Retrieved July 1, 2005, from http://www.dlib.org/dlib/january99/bearman/01bearman.html

Brickley, D. & Guha, R. V. (2004), RDF Vocabulary Description Language 1.0: RDF Schema, *W3C Recommendation 10 February 2004*. Retrieved July 1, 2005, from http://www.w3.org/TR/rdf-schema/

DCMI Usage Board (2003), DCMI Grammatical Principles. Retrieved April 30, 2006 from http://dublincore.org/usage/documents/principles/

Dublin Core Application Profile Guidelines (2003), CEN Workshop Agreement CWA 14855. Retrieved July 1, 2005, from ftp://ftp.cenorm.be/PUBLIC/CWAs/e-Europe/MMI-DC/cwa14855-00-2003-Nov.pdf

Duval, E., Hodgins, W., Sutton, S. & Weibel, S. L. (2002), Metadata Principles and Practicalities, *D-Lib Magazine*, April 2002. Retrieved July 1, 2005, from http://www.dlib.org/dlib/april02/weibel/04weibel.html

Friesen, N., Mason, J. & Ward, N. (2002), Building Educational Metadata Application Profiles, *Dublin Core - 2002 Proceedings: Metadata for e-Communities*: *Supporting Diversity and Convergence.* Retrieved July 1, 2005, from http://www.bncf.net/dc2002/program/ft/paper7.pdf

Godby, C. J., Smith, D. & Childress, E. (2003), Two Paths to Interoperable Metadata, *Proceedings of DC-2003: Supporting Communities of Discourse and Practice – Metadata Research & Applications,* Seattle, Washington (USA). Retrieved July 1, 2005, from http://www.siderean.com/dc2003/103_paper-22.pdf

Guidelines for machine-processable representation of Dublin Core Application Profiles (2005), CEN Workshop Agreement CWA 15248. Retrieved July 1, 2005, from ftp://ftp.cenorm.be/PUBLIC/CWAs/e-Europe/MMI-DC/cwa15248-00-2005-Apr.pdf

Heery, R. & Patel, M. (2000), Application Profiles: mixing and matching metadata schemas, *Ariadne Issue 25*, September 2000. Retrieved July 1, 2005, from http://www.ariadne.ac.uk/issue25/app-profiles/

Heflin, J. (2004), OWL Web Ontology Language – Use Cases and Requirements, *W3C Recommendation 10 February 2004*. Retrieved July 1, 2005, from http://www.w3.org/TR/webont-req/

IMS Global Learning Consortium (2004), IMS Meta-data Best Practice Guide for IEEE 1484.12.1-2002 Standard for Learning Object Metadata. Retrieved July 1, 2005, from http://www.imsglobal.org/metadata/mdv1p3pd/imsmd_bestv1p3pd.html

Johnston, P., (2005a), XML, RDF, and DCAPs. Retrieved July 1, 2005, from http://www.ukoln.ac.uk/metadata/dcmi/dc-elem-prop/

Johnston, P., (2005b), Element Refinement in Dublin Core Metadata. Retrieved July 1, 2005, from http://dublincore.org/documents/dc-elem-refine/

Klyne, G. & Carroll, J. J. (2004), Resource Description Framework (RDF): Concepts and Abstract Syntax, *W3C Recommendation 10 February 2004*. Retrieved July 1, 2005, from http://www.w3.org/TR/rdf-concepts/

Lagoze, C. (1996), The Warwick Framework – A Container Architecture for Diverse Sets of Metadata, D-Lib Magazine, July/August 1996. Retrieved April 30, 2006 from http://www.dlib.org/dlib/july96/lagoze/07lagoze.html

Manola, F. & Miller, E. (2004), RDF Primer, *W3C Recommendation 10 February 2004*. Retrieved July 1, 2005, from http://www.w3.org/TR/rdf-primer/

Miles, A. J. & Brickley, D. (2005), SKOS Core Guide, *W3C Working Draft 10 May 2005*. Retrieved July 1, 2005, from http://www.w3.org/TR/swbp-skos-core-guide

Memorandum of Understanding between the Dublin Core Metadata Initiative and the IEEE Learning Technology Standards Committee (2000). Retrieved July 1, 2005, from http://dublincore.org/documents/2000/12/06/dcmi-ieee-mou/

Nack, F., van Ossenbruggen, J. & Hardman, L. (2005), That obscure object of desire: multimedia metadata on the Web, part 2, *IEEE Multimedia 12* (1) 54-63. Retrieved July 1, 2005, from http://ieeexplore.ieee.org/iel5/93/30053/01377102.pdf?arnumber=1377102

Nilsson, M. (2001), The Semantic Web: How RDF will change learning technology standards, Feature article, Centre for Educational Technology Interoperability Standards (CETIS). Retrieved July 1, 2005, from http://www.cetis.ac.uk/content/20010927172953/viewArticle

Nilsson, M., Palmér, M. & Naeve, A. (2002), Semantic Web Metadata for e-Learning - Some Architectural Guidelines, *Proceedings of the 11th World Wide Web Conference (WWW2002)*, Hawaii, USA. Retrieved July 1, 2005, from http://kmr.nada.kth.se/papers/SemanticWeb/p744-nilsson.pdf

Nilsson, M., Palmér, M. & Brase, J. (2003), The LOM RDF Binding - Principles and Implementation, *Proceedings of the Third Annual ARIADNE conference*. Retrieved July 1, 2005, from http://kmr.nada.kth.se/papers/SemanticWeb/LOMRDFBinding-ARIADNE.pdf

Nilsson, M., Johnston, P., Naeve, A., Powell, A. (2006), The Future of Learning Object Metadata Interoperability, in Koohang A. (ed.) *Learning Objects: Standards, Metadata, Repositories, and LCMS*, in press.

Powell, A., Nilsson, M., Naeve, A., Johnston, P. (2005), DCMI Abstract Model, *DCMI Recommendation.* Retrieved July 1, 2005, from http://dublincore.org/documents/abstract-model/

Uschold, M. & Gruninger, M. (2002), Creating Semantically Integrated Communities on the World Wide Web, Invited Talk, Semantic Web Workshop, Co-located with

WWW 2002, Honolulu, HI, May 7 2002. Retrieved July 1, 2005, from http://semanticweb2002.aifb.uni-karlsruhe.de/USCHOLD-Hawaii-InvitedTalk2002.pdf

van Ossenbruggen, J., Nack, F. & Hardman, L. (2004), That obscure object of desire: multimedia metadata on the Web, part 1, *IEEE Multimedia 11* (4) 38-48. Retrieved July 1, 2005, from http://ieeexplore.ieee.org/iel5/93/29587/01343828.pdf?arnumber=1343828