As published in Technology, Instruction, Cognition and Learning vol 7 (3-4) 2010, pp 225-243. URL http://www.oldcitypublishing.com/TICL/TICLcontents/TICLv7n3-4contents.html

METADATA FOR LEARNING MATERIALS: AN OVERVIEW OF EXISTING STANDARDS AND CURRENT DEVELOPMENTS

Phil Barker, JISC CETIS, ICBL, Heriot-Watt University, Mountbatten Building, Edinburgh EH14 4AS, UK (corresponding author: email: philb@icbl.hw.ac.uk; phone: +44 131 451 3278; fax: +44 131 451 3327).
Lorna M. Campbell, JISC CETIS, CAPLE, University of Strathclyde, 50 George Street, Glasgow G1 1QE, UK.

Abstract
This paper provides an overview of specifications and standards for metadata relating to learning materials. It is structured to present first the currently established metadata schemas in use today (specifically the IEEE LOM and Dublin Core metadata), then to examine current developments and activities before looking at what might be the future challenges. The examination of current developments and activities highlights the increasingly recognized importance of metadata schema that describe what have in the past been thought of as secondary aspects of learning materials (for example who uses them and what for), and the importance of alternative approaches to structured metadata for resource description.

Keywords: metadata, resource description, learning resources, IEEE LOM, Dublin Core, Learning Object Metadata.

Introduction
What do we mean by "metadata" and "learning materials"? A useful definition of metadata is that used by NISO (2004) "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource". This definition has two important parts. Firstly, it distinguishes metadata from unstructured textual descriptions of a resource. The structuring of metadata normally takes the form of elements with defined semantics to describe specified characteristics of a resource and a syntactical binding for these elements, the aim of which is to allow machine processing of the information without requiring computational semantic analysis techniques such as text mining. Secondly, the NISO definition stresses that metadata exists to facilitate a range of activities. Resource discovery is the most visible activity facilitated by metadata, and is the one that seems most closely associated with metadata by most people; however, appropriate management and use of resources are no less important.

Defining what we mean by learning materials is more difficult. However, we think that "anything used for teaching and learning" captures the essence of what we are interested in. This approach makes the defining characteristic of learning materials their function and context, as opposed to characteristics that are inherent to the resource; this contrasts them with many other resource such as images, simulations, audio, etc which are more readily defined by resource specific characteristics. This has significant implications for the definition and development of learning resource metadata standards. We shall not try to distinguish here between educational materials, learning objects, educational resources, etc.

The following sections will focus on the two established metadata standards most relevant to learning materials, the IEEE LOM and Dublin Core, and will briefly describe and reflect on their characteristics and applications. This paper will also outline the current work being undertaken on these two schemas and on the development of a third related standard, ISO MLR. Finally we look at some of the future challenges facing the field of metadata for learning materials regardless of which specific standard one favours.

Established Metadata Schemas

IEEE Learning Object Metadata (LOM)

The IEEE LOM is (currently) an open and internationally recognized two-part standard for the description of "learning objects" and is composed of a conceptual data schema (IEEE, 2002) and an XML binding of that schema (IEEE, 2005). The definition of "learning object" used in the standard is "any entity digital or non-digital that may be used for learning, education, or training", which is comparable to the working definition used above. The LOM data schema specifies which characteristics of a learning object should be described and what vocabularies may be used for these descriptions; it also defines how this data model can be amended by additions or constraints.

The LOM conceptual data schema consists of a hierarchy of elements as shown in figure 1. The first level is composed of nine categories, each of which contains sub-elements; these sub-elements may be simple elements that contain data, or they may themselves be aggregate elements that contain further sub-elements. The data model specifies that some elements may be repeated either individually or as a group. For example, the elements 9.3 (*description*) and 9.1 (*purpose*) can only occur once within each instance of the *classification* category element, however the *classification* element may be repeated, thus allowing many descriptions for different purposes.

The semantics of LOM elements are determined by their context: they are affected by the parent or container element in the hierarchy and sometimes by other elements in the same container. For example the various *description* elements (1.4, 5.10, 6.3, 7.2.2, 8.3 and 9.3) each derive their meaning from their parent element: e.g. 5.10, *education.description* describes educational characteristics of the resource; 6.3 *rights.description* relates to the terms and conditions of use of the resource, and so on. In addition, *description* element 9.3 also derives some of its meaning from the value of element 9.1 *purpose* in the same instance of the *classification* category element.

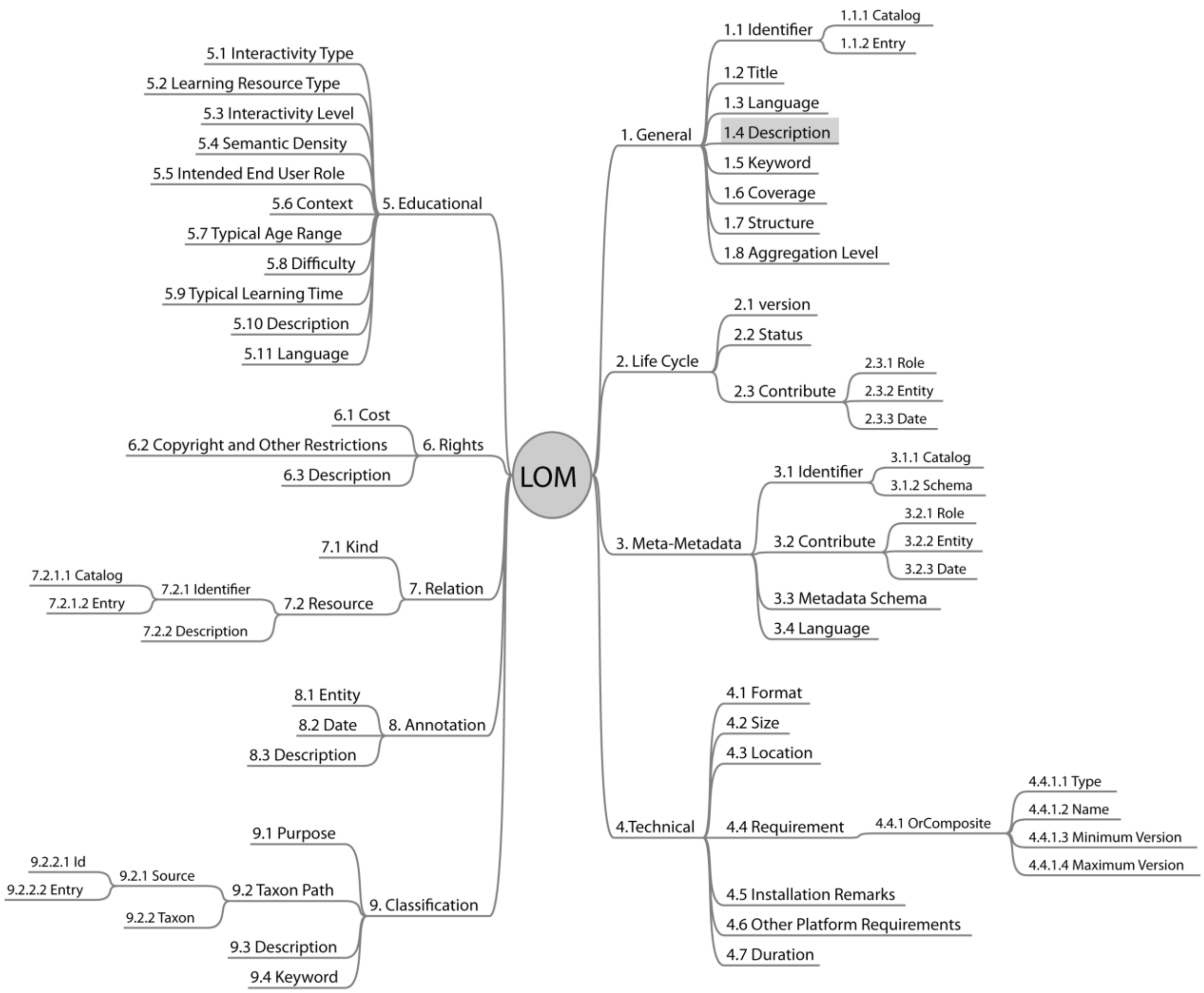

Figure 1: a schematic representation of the hierarchy of elements in the LOM data model.

The data schema also specifies the value space and datatype for each of the simple data elements. The value space defines the restrictions, if any, on the data that can be entered for that element. For some elements the value space allows any string of Unicode characters to be entered; for other elements entries must be drawn from a declared list (i.e. a controlled vocabulary) or must be in a specified format (e.g. date and language codes). Some element datatypes simply allow a single string of characters to be entered; others comprise two parts as described below:

- LangString datatype: where the data entered is likely to be text that would be read directly by a human the data is of a type defined by the LOM as a LangString. LangString items comprise two parts: one providing a language code and the second the Unicode text in the language specified by the code. The same information may be conveyed in multiple languages by repetition of data within an element as several LangStrings.

- Vocabulary datatype: where the LOM data schema requires an element to be described by a controlled vocabulary the element will be of the vocabulary datatype. Such elements are composed of Source-Value pairs; the source should contain the name of the list of terms being used and the value should contain the chosen term.

- DateTime and Duration datatypes: these datatypes allow a date or period of time to be given in a machine-readable format (the value space is based on the ISO 8601:2000 standard; an example of a correctly formatted date is 2003-11-22); a human-readable description may be provided instead of or in addition to the formatted date (e.g. "late 20th century").

Data or service providers implementing the LOM are not required to support all the elements in the conceptual data schema and the LOM data schema need not limit the information that may be provided. The creation of application profiles allows a community of users to specify which elements and vocabularies they will use. Elements from the LOM may be discarded and elements from other metadata schemas may be introduced; likewise, LOM vocabularies may be supplemented with values that are appropriate to the community that the implementers wish to support.

The LOM has been widely implemented by repositories and other learning resource providers, partly as a result of its status as an international standard, and partly through its association with other influential specifications, such as those produced by the IMS Global Learning Consortium (e.g. Content Packaging, Question and Test Interoperability) (IMS, no date, a, and b) and by ADL (SCORM) (ADL, no date). Examples of repositories and initiatives that have adopted the LOM are the JORUM (no date), a JISC funded repository of teaching and learning materials for UK Further and Higher Education; the European Ariadne foundation (Ariadne, no date); various European SchoolNet projects (European SchoolNet, no date); the Global Learning Objects Brokered Exchange (GLOBE, no date) federation; and many more.

Reflections on the LOM

The origins of the LOM can be traced back to developments initiated in the mid 1990s and it should be seen as an early attempt to deal with the difficulties of multiply-versioned complex objects. It is important to appreciate that requirements and expectations for the use of such resources was significantly different from those with which we are currently familiar. The LOM includes a multitude of pre-defined elements and complex structures, all of which were included for an envisaged need, but many of which do not seem to have been widely used. Evidence of this is recorded by Godby, 2004, and Friesen, 2004, however it is important to note that these papers date from very shortly after the LOM was standardized and it would be interesting to repeat these studies to ascertain whether practice has changed as understanding of the LOM has matured. Similarly, changes in technology-enhanced learning and technical infrastructure in the last fifteen years, most notably the web and the semantic web, have introduced new requirements and expectations that are not well reflected in the LOM.

While the uptake and influence of the LOM has been considerable, and it has formed the basis for resource description in many repositories and federations of repositories, problematic issues have been noted. One such issue is that the LOM conceptual data schema (the stated aim of which is to "ensure that bindings of learning object metadata (LOM) have a high degree of semantic interoperability" IEEE, 2002, section 1.2) is not based on an abstract model shared with other metadata schema, and does not align with base standards for semantic interoperability, such as RDF. This makes semantic interoperability with other metadata standards problematic (Nilsson, 2008). Essentially it is impossible to import elements from other metadata schema, such as Dublin Core (see below) or schema developed to support specific resource types such as images or specific features such as rights management or preservation. This is especially problematic for a domain-specific standard since it inhibits what may be regarded as the modularization and specialization of interest because it is necessary for the LOM to accommodate general and non-educational characteristics (e.g. technical, rights, accessibility, etc) within the standard data schema rather than importing solutions from other domains. Pragmatic approaches to importing elements from other schema work only between those LOM systems that either understand the imported elements through some prior knowledge on the part of their implementers, or those that can work without the information the imported elements convey.

Other issues have arisen relating to sharing extensions: where extensions have been defined to meet the needs of a specific community or federation of repositories there has been little evidence of other communities with similar needs adopting these same extension. Frequently the same need seems to met be through slightly different extensions by different communities, thus further restricting the scope of interoperability.

Dublin Core Metadata

The Dublin Core Metadata Initiative (DCMI) develops metadata standards for the description of a broad range of resource types for a diverse variety of purposes. They are best known for the fifteen element "simple" Dublin Core Element Set (DCMI, 2008), which has been standardized as ISO Standard 15836-2003 (ISO, 2003). The core Element Set is intended to be "broad and generic, usable for describing a wide range of resources" (DCMI, 2008). The range of resource types to which Dublin Core metadata is applicable is emphasized in the formal definition of a resource, used elsewhere by DCMI, as "anything which might be identified" (Powell *et al*, 2007). These fifteen elements are:

> contributor, coverage, creator, date, description, format, identifier, language, publisher, relation, rights, source, subject, title and type.

All these elements are optional and may be repeated if required.

Since the inception of the DCMI it has been recognized that it would often be desirable to define more specific semantics for these elements. For example, for many resources it may be necessary to distinguish between different dates (e.g. date of submission and date of publication) associated with the lifecycle of a resource. Also, while the coverage of the Element Set is broad, it is not exhaustive: there are many characteristics of resources that are not covered, some of which are important in specialized domains. For example there is no way to describe the intended audience of a resource. For these reasons the "simple" Dublin Core Element Set has been supplemented with refinements and extensions resulting in what has historically been termed "qualified" Dublin Core metadata.

The Dublin Core Element Set dates from 1998. Since then understanding of metadata and the semantic web has evolved to require more formal definitions of metadata elements, their relationship to resources, and the values they may be assigned. The DCMI has responded with a set of specifications that comprises:

- the DCMI Abstract Model

- DCMI Metadata Terms
- the Singapore Framework for Dublin Core Application Profiles
- Guidelines for encoding Dublin Core metadata in RDF, XML and HTML/XHTML meta and link elements.

Taken together, these documents have attempted to update Dublin Core metadata to facilitate extensibility and harmonize it with the principles of the semantic web while at the same time ensuring backward compatibility with the original core Element Set.

The **DCMI Abstract Model** (Powell *et al*, 2007) "defines the nature of the components used [in Dublin Core metadata] and describes how those components are combined to create information structures". It provides three models. The *resource model* defines the relationship between the resource being described and the resources used in the description. A resource is defined as "anything which might be identified", and so includes real and imaginary things, and abstract intellectual constructs such as metadata elements themselves. According to the model each DC metadata description describes one and only one resource--this is known as the one-to-one principle. Real world descriptions, for example a catalogue record of a book, involve the description of several related resources, for example the book itself, the author, the publisher, etc. The *description set model* defines how statements about individual resources can be related to each other in order to provide such real world descriptions; Dublin Core metadata description sets may be instantiated as records. The *vocabulary model* defines the structure of vocabularies used in DC metadata descriptions, where the vocabulary is a set of defined terms with specific meaning in the abstract model.

DCMI Metadata Terms (DCMI, 2008b) defines all the metadata terms maintained by the DCMI. The terms are divided into properties, vocabulary encoding schemes, syntax encoding schemes and classes. *Classes* are formal categories of resources that share important characteristics, e.g. "bibliographic resources" (books, journal articles) or "file formats". *Properties* can be used to describe specific aspects, characteristics, attributes or relations of a resource, and include revisions of the 15 members of the Dublin Core Element Set. Dublin Core metadata properties may refine other properties, and may have specific domains and ranges, i.e. may be used to describe resources from a specific class or may have values that are drawn from a specific class. The *syntax* and *vocabulary encoding schemes* allow the identification of syntactical methods and vocabularies used to provide information in Dublin Core metadata descriptions.

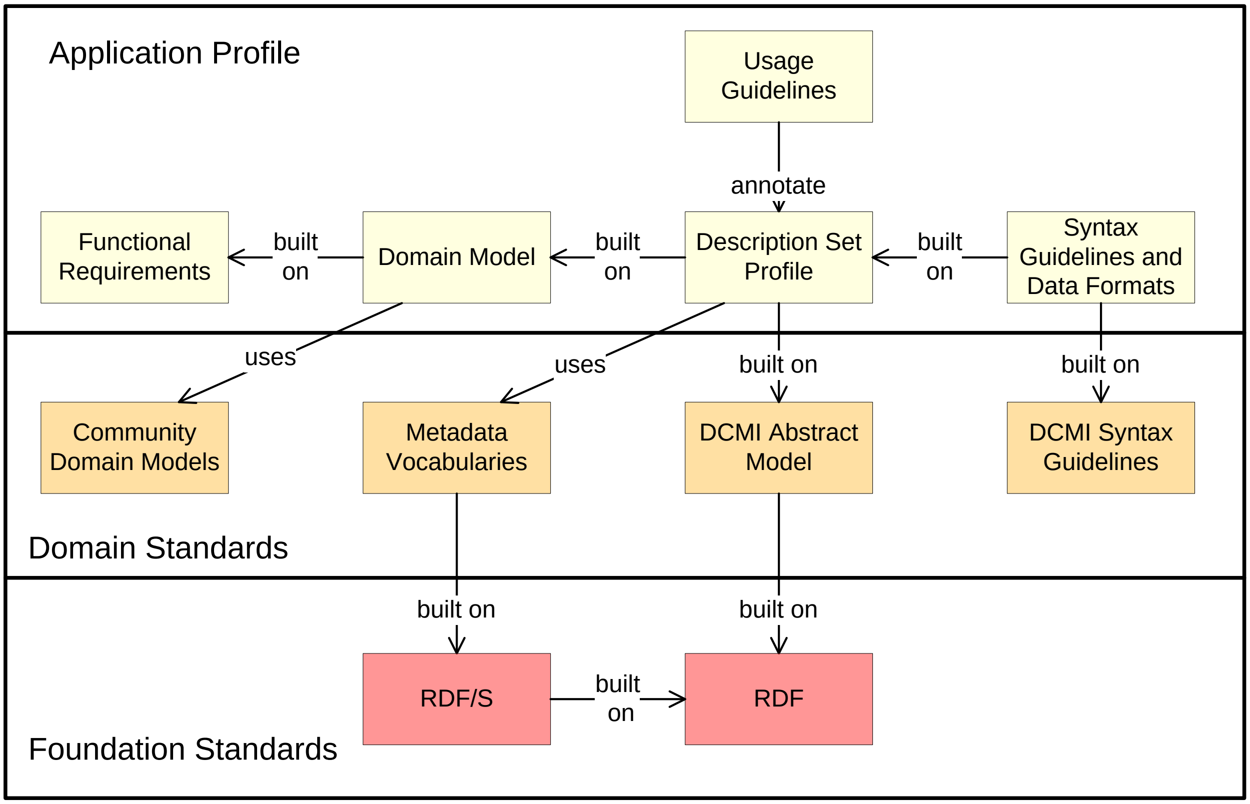

Figure 2: a graphical representation of the Singapore Framework (from Nilsson *et al* 2008).

The Singapore Framework for Dublin Core Application Profiles (Nilsson *et al* 2008) describes an approach to creating and documenting application profiles based on the Dublin Core abstract model and metadata vocabularies, such as the DCMI Metadata Terms, that are compatible with the model. The Framework is represented graphically in figure 2. The central component of the application profile is the Description Set Profile, which "defines a set of metadata records that are valid instances of an application profile". This is built on a domain model, which describes the scope of the application profile by defining "the basic entities being described by the application profile and their fundamental relationships". The domain model itself is built on functional requirements. Optional usage and syntax encoding guidelines describe how to apply the application profile and define any syntactical structures that are specific to the profile. The whole application profile is based on "domain standards" such as the DCMI Abstract Model, DCMI syntax guidelines and metadata vocabularies.

Reflections on Dublin Core Metadata

One key difference between the approach taken by Dublin Core and that of IEEE LOM is that the LOM sets out to define what is called an "instance" (IEEE, 2002, section 1.1), that is a block of metadata which if expressed in XML starts with `<lom>` and ends with `</lom>`, whereas the Dublin Core approach is to define individual terms and the rules for their application. To conform to the LOM one may not use elements from other metadata schema if doing so replaces an existing LOM data element (IEEE, 2002, section 5), thus it is not possible to use most Dublin Core elements and conform with the LOM standard; no equivalent restriction exists for the use of Dublin Core metadata. Rather, DCMI has the concept of levels of interoperability (Nilsson *et* *al*, 2009), which allows for the use of individual Dublin Core terms for semantic interoperability with or without reference to any specific Dublin Core concept of a record for syntactic interoperability.

The uptake of "simple" Dublin Core, i.e. the Element Set, has been considerable, notably in specifications such as: OAI-PMH (Open Archives Initiative, 2002) where it is an integral part of the specification defining the minimum metadata requirement; SRU (Library of Congress, 2007); and RDF (e.g. Manola and Miller, 2004) where DC Elements were used in examples contained in the documentation and have widely been used in implementations of the specification. As a result many repositories and information systems support the Dublin Core Element Set, if not natively then as a translation of the native metadata used for export and interoperability purposes. However, the more recent approaches based on the DC Abstract Model, DC Terms and the Singapore Framework, while guided by sound theoretical principals, are as yet unproven by mass implementation. This is by no means a trivial task, as considerable expertise is required in a range of areas and one has to ask whether the necessary extra effort will yield significantly greater benefits.

Current Developments and Activity

IEEE LOM in query and harvest

One of the reasons for using a standardized metadata schema such as the LOM is to achieve efficiencies through sharing metadata or sharing services based on the metadata. Although these potential benefits are well recognized and standard metadata schema have been used in the learning domain by individual projects and services, there remains a lack of shared practice for the application of these standards to the domain. This inhibits the provision of joined-up services linking repositories of teaching and learning materials and reduces the ability of teachers and learners to find content appropriate to their educational needs. The IMS Global Learning Consortium's Learning Object Discovery and Exchange (IMS, no date, c) project is attempting to address this deficiency by providing profiles and guidelines for the use of existing standards and specifications for sharing metadata and related services in the learning, education and training domain. A related initiative is the ASPECT project, coordinated by European Schoolnet, one of the projects involved in developing IMS LODE. ASPECT is a European best practice network aiming to improve the adoption of learning technology standards and specifications in Europe (ASPECT, no date).

The IMS LODE project intends to build on other existing standards and specifications, including two generic standards: Search/Retrieval via URL (SRU, see Library of Congress, 2007) for sharing search services, and the Open Archives Initiative's Protocol for Metadata Harvesting (OAI-PMH, see Open Archives Initiative, 2002) for exchanging metadata records.

SRU facilitates remote searching of repositories, i.e. it allows a search service to be constructed for one or more target repositories that is independent of the repository(ies) being searched. In order to do this it is necessary to know something about the metadata at the target repository for two reasons: firstly, in order to search specific fields (e.g. to find everything where the “author” field equals “William Shakespeare” one needs to know how to address fields relevant to the author name), secondly in order to interpret the result set that is returned. For SRU, search terms are formatted according to the Contextual Query Language (CQL) part of the specification, which uses so called "context sets" to define the indexes being searched and the relationship required between the search term and the index term. Simple Dublin Core element names are commonly used in this context, for example dc.author = = "William Shakespeare" would form part of a request searching the index of the author fields for values exactly equal to the string "William Shakespeare". The matching records may be returned by the repository being searched in any XML-encoded metadata schema, though simple DC is defined as a *de facto* default. Clearly where educational characteristics of resources are required for resource discovery and selection it is important to be able search for these characteristics and to return records that include their description: i.e. there is a need for a context set that allows a search to be performed on LOM elements and for LOM records to be returned.

OAI-PMH facilitates the transfer of metadata records from a repository (a data provider) to another system either singly or, more usually, in bulk. This allows the receiving system to provide a service based on the metadata records for the content of one or more OAI-PMH data providers—the typical service being the search for content in several repositories. While OAI-PMH mandates that metadata must be provided in a form of simple Dublin Core, it provides the option of transferring other metadata formats as well. In practice, however, the use of richer metadata schema, such as those suitable for describing educational characteristics of resources is variable, is often confined to private collaboration projects and is sorely in need of openly available, widely endorsed best practice guidelines. Some guidelines do exist for generic application of OAI-PMH, for example those by the Digital Library Federation (DLF, no date) and those from the DRIVER project (Vanderfeesten *et al*, 2008). However, there remains a need to address some issues that are specific to the use of the LOM with OAI-PMH.

Mapping the IEEE LOM to the DCMI Abstract Model

It has long been acknowledged that it will often be necessary to use specialist metadata terms alongside those defined in the IEEE LOM conceptual data schema (Barker *et al*, 2006, section 4 and Duval *et al*, 2002). This is a consequence of the broad range of resource types that may be used in learning contexts and the broad range of activities that need to be supported in order to manage these materials. It would be unfeasible for the LOM conceptual data schema to provide metadata elements for the description of every single characteristic that may be important for every resource type and activity (see Barker, 2008, for a description of the wide range of characteristics that conceivably may need to be described). One approach to this problem is for application profiles to draw on elements from other schemas to extend the IEEE LOM conceptual data schema. However, this approach is somewhat problematic, because of fundamental differences in how the semantics of elements are expressed in metadata schemas based on different abstract models (Nilsson, 2008).

Furthermore, it is desirable that LOM-based systems should be capable of operating within networks based on other metadata schema, such as Dublin Core and the semantic web, which requires that LOM metadata elements can be expressed in those schema (DCMI, no date). To this effect it would be advantageous if LOM descriptions could be expressed in RDF, RDF providing a common model for making assertions about characteristics of a resource that is independent of the nature of those characteristics. This common model allows terms from different metadata vocabularies to be "mixed-and-matched", so that educational characteristics could be expressed using terms drawn from the LOM and complemented with descriptions of characteristics of other types (for example rights and licensing arrangements) using terms drawn from some more appropriate standard RDF vocabulary.

A joint DCMI/IEEE LTSC[1] taskforce is working to address these issues (DCMI, no date). The taskforce will produce two outputs the first of which will be an IEEE Standard for an RDF vocabulary for expressing the semantics of the data elements in the LOM conceptual data schema. The standard will draw on the RDF Schema description language and the Dublin Core Abstract Model. Where necessary it will define RDF terms for new properties, classes, vocabularies, syntax encoding schemes and vocabulary encoding schemes; where possible data elements will be expressed using terms that can be drawn from existing RDF vocabularies (e.g. Dublin Core). The second output will be "recommended practice for expressing IEEE Learning Object Metadata instances using the Dublin Core Abstract Model". This will specify how to use the metadata terms defined by the RDF vocabulary for LOM data elements from the first output to express IEEE LOM conforming instances as Dublin Core description sets.

Dublin Core Education Application Profile

A related DCMI activity is the creation of an application profile for education based on the principals outlined in the Singapore Framework (DCMI, no date b). The approach being taken is described as a modular profile, covering only those properties and relationships of a resource that are relevant to education. The intention is that this can be "plugged-in" to metadata descriptions (at least those that share the same underlying model as Dublin Core) that cover other generic or specific characteristics of the resource. For example one can imagine that an application profile used to describe educational videos might comprise some elements of generic metadata (e.g. title, date of creation etc.), metadata specific to video (e.g. aspect ratio, frame rate etc.) and elements from the DC-Ed profile to describe educationally significant characteristics of the videos.

The requirements of the Singapore Framework[2] for creating application profiles are to document functional requirements, produce a domain model, and to select terms from suitable metadata vocabularies on the basis of these.

Some DC Terms already exist that are relevant to education, for example for stating the educational level of the intended audience of a resource, or the relationship between a resource and an educational attainment standard; the work of the DCMI/IEEE LTSC Taskforce provide further relevant DC Terms for expressing elements from the LOM base schema. The DC-Ed working group has already collected 49 use cases for educational metadata from 23 organizations in 6 countries; these have been related to functional requirements and properties of resources (Currier, 2008). It is worth noting that many of the properties to which the requirements relate are not properties of the primary educational resource being described (i.e. the learning material) but rather are properties of related resources. For example, a use case along the lines of "a teacher wants to find resources that have been used successfully in classes similar to her own

1 LTSC: the IEEE Learning Technology Standards Committee, which is the committee within the IEEE that is responsible for LOM development and maintenance.

2 There are clearly other actions required to implement such a profile in an information system, these requirements relate to approval of the work by the DCMI.

(e.g. 1st year undergraduate physics)" will lead to a requirement that a description is provided of the educational level of learners in classes where a resource has been used is provided. These inter-relationships between resources that need to be described in a application profile are outlined in the domain model of the Singapore Framework. A draft domain model has been proposed for the DC-Ed profile, and is reproduced as figure 3 below, however it should be emphasized that this represents a starting point for discussion rather than the final finished model.

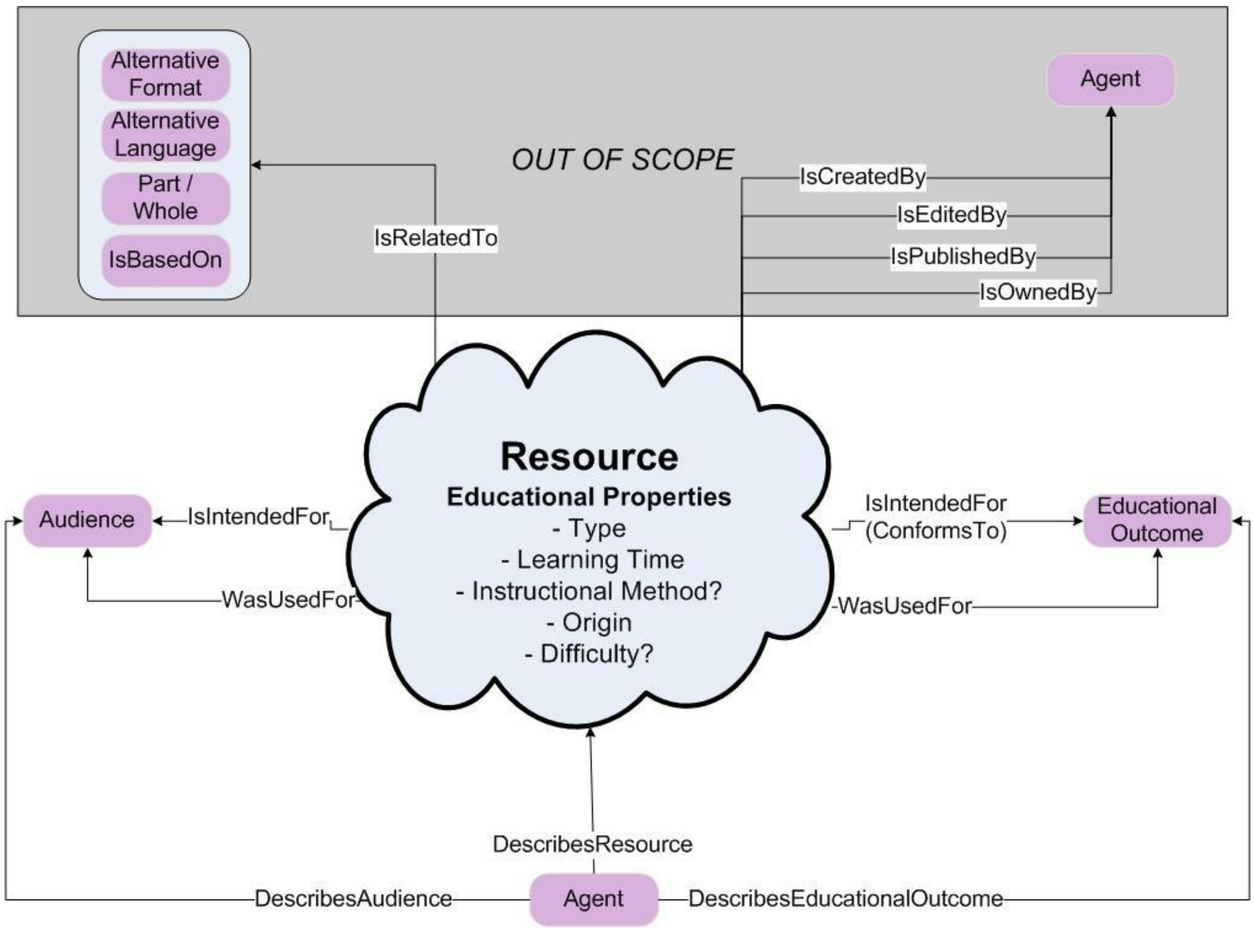

Figure 3. A domain model proposed for the DC-Education application profile. The central cloud represents any resource-type-specific domain model for the primary material.

<u>Other Educationally Relevant Metadata Specifications in Development</u>

The inclusion of entities such as Audience and Educational Outcome into the domain model for educational metadata highlights the importance of some types of metadata that have, perhaps, been regarded as secondary metadata in the past, but which increasingly appear to be of primary importance to education. Indeed, it seems to follow from the working definition of learning materials as "anything used for teaching and learning" that the defining educational characteristics pertain not to the material itself but to the use of that material. The relevant metadata schemas are those describing audience interests (so called attention metadata), courses, and competencies, and there is interesting work in progress in all these areas.

The Attention Profiling Mark-up Language (APML, no date) enables the description of topics and sources that a person is interested in to be shared in the form of an XML file. The attention profile may be generated explicitly by the person concerned or may be derived from attention data, i.e. information about what a person has been looking at derived from a record of their activities. This specification is being developed by a community with no direct affiliation to any formal specification or standards body. Notwithstanding the specification's draft status several prototype services have implemented it including digg.com and the BBC's radiopop.co.uk.

The European Committee for Standardization (CEN) has endorsed a workshop agreement and a commitment to develop a European Norm for Metadata for Learning Opportunities (see Wilson, 2008). This work has its origins in course description metadata initiatives from several European countries, and describes a common model for learning opportunities so that they may be aggregated by other services. The initial focus is on course advertising; however there is scope for wider application to course description for other purposes.

There is a long history of work on standardizing competency definitions and the like, including the IEEE Reusable Competency Definition (IEEE, 2007) and HR-XML (HR-XML consortium, no date). Currently working group 3 of ISO subcommittee 36 (ISO/IEC JTC1 SC36 WG3) is developing a conceptual reference model for competences and related objects. Again, there is potential scope to apply this model to the educational outcomes object in the DC-Education domain model.

ISO Metadata for Learning Resources

A third initiative related to educational metadata that is currently underway is Metadata for Learning Resources (MLR), which is being undertaken by working group 4 of ISO subcommittee 36 (ISO/IEC JTC1 SC36 WG4). The origins of this work can be traced back to proposals in 2002/3 to adopt IEEE LOM as an ISO standard. This proposal was rejected for a variety of reasons including that of support for internationalization in the LOM. A working group was established to investigate these issues further and propose solutions. The outcomes of the working group were published in 2006 by Norm Friesen in a CanCore article "Building a better LOM". Since then attempts to address the issues raised have proceeded within SC36. Although a range of concerns were raised about the initial direction of this work, recent developments have focused on the proposed adoption of a semantic model that will hopefully maximize ISO MLR's compatibility with current efforts in Dublin Core and the IEEE LTSC.

As currently proposed ISO MLR will be a multipart standard composed of six parts: 1, the framework; 2, data elements; 3, the core application profile; 4, technical elements; 5, education elements; 6, availability and rights management. However, other parts may be defined in the future.

Future Challenges

Requirements for Educational Metadata and a Domain Model

During the course of a recent study on the range of metadata that may be necessary to describe educational materials (Barker, 2008) it became evident that metadata requirements for educational resource types and purposes are not well understood and are less well articulated.

Pinning down the details of which educationally significant characteristics pertain to which entities and which relationships are important is a crucial step in understanding what information is needed to create resource descriptions that meet educational requirements, and how to go about gathering that information. For example, information about how a resource is used, such as what course is it used for and the subject and educational level of that course, may be gathered by course management systems such as VLEs or MLEs, but this information is rarely, if ever, passed back to the system that manages the resource descriptions, i.e. the repository or catalogue.

Distributed Metadata

The above example is one of many that illustrates how metadata may be distributed across many systems. A more conventional example would be a resource discovery service that searches across several independent repositories. The rationale for the latter example is that the quality of the search can be enhanced by aggregating the contents of several repositories; similarly in the former example the quality of the information gathered, and hence the service offered, can be enhanced by aggregating information about usage from several systems. As well as being distributed across many systems it is highly likely that the metadata will be heterogeneous: different systems will record different metadata and make it available in different formats. The concepts of the semantic web may be useful in dealing with such distributed heterogeneous metadata but this has yet to have much impact in practice, particularly in the educational domain.

Limits of Metadata

Another observation made during the recent study on metadata requirements for educational materials is that when precise metadata requirements are not well articulated for a particular domain it is often common practice to provide descriptions in the form of free text. The original rationale for creating structured metadata was to record resource descriptions that were machine readable without some form of computational semantic analysis of free text. Key to this requirement is the assumption that a computer will be taking action on the basis of information conveyed in a resource description (for example selecting an appropriate resource for a given scenario) rather than a human taking this action. However, it is quite possible that in many cases it may be sufficient find a description of the right thing (in terms of an entity or relationship in an agreed domain model) and to present this in human readable form to the user who can then take action. This reduces the role of metadata to the well-understood role of supporting resource discovery, i.e. allowing the user to find the human readable description.

Also highly relevant to situations when precise metadata requirements are not widely agreed are approaches such as (social) tagging and folksonomies. These allow users, or groups of users, to apply descriptive keywords to resources without worrying about the details of the precise relationship between the concept expressed by the keyword and the resource. The users also do not necessarily have to agree with others about what term should be used to express the concept, though many of the systems that implement tagging approaches also include mechanisms for identifying commonly used tags for each resource, which can be useful in identifying any emerging consensus about which terms are appropriate.

Closing Observations

Over the last fifteen years or so there has been a general shift away from individual, monolithic systems and solutions, and towards systems that comprise distributed components, as exemplified by service oriented architectures, and, arguably, the web as a distributed information system (particularly when one considers the use of RESTful APIs). In many ways the IEEE LOM standard now appears to be a typical product of the age of monolithic systems: resource description based on a coherent record describing all aspects of a "learning object" and its use, complying with a single standard. Description of any characteristics not already included in the LOM conceptual data schema was envisaged as being achieved by extending that schema. Better understanding of the semantic complexities associated with "mixing and matching" metadata schema has lead to a move away from this single schema approach and towards one where metadata from different schema can be mixed if they are based on a unifying abstract model.

At the same time there has been growing recognition that the educational resources being described are not discrete "learning objects" but amalgams of multiple resources each with their own characteristics and bound together by complex relationships. This recognition has resulted in considerably more complex requirements for resource descriptions to be generated by a range of agents, and yet to a large extent we still rely on manual cataloguing: there has been no main-streaming of automatically generated educational metadata. This does not seem sustainable, especially when compared to the Google approach for resource discovery. An alternative approach has to be found which enables independent actors to create descriptions of multiple related resources in heterogeneous formats, both human and machine-readable, and which facilitates the aggregation of these descriptions (and their exposure to search tools such as Google) based on an agreed abstract model of resource relationships.

References

ADL (no date). *Sharable Content Object Reference Model (SCORM).* Retrieved 28 February 2009, from http://www.adlnet.gov/scorm/

Ariadne Foundation (no date). *ARIADNE*. Retrieved 28 February 2009, from http://www.ariadne-eu.org/

APML (no date) *The Attention Profile Mark-up Language* Retrieved 28 February 2009, from http://www.apml.org/

ASPECT (no date). *Adopting Standards and Specifications for Educational Content.* Retrieved 28 February 2009, from http://aspect.eun.org/

Barker, P. Campbell, L. M., Roberts, A. & Smythe, C. (2006) *IMS Meta-data Best Practice Guide for IEEE 1484.12.1-2002 Standard for Learning Object Metadata* Retrieved 23 September 2009 from http://www.imsproject.org/metadata/mdv1p3/imsmd_bestv1p3.html

Barker, P. (2008). *Learning Material Application Profile Scoping Study – final report.* Tech Rep. Retrieved 28 February 2009, from http://www.icbl.hw.ac.uk/lmap/lmapreport.d3.pdf

Currier, S. (2008). *Presentation on the Dublin Core Education application profile module* at DC-2008 International Conference on Dublin Core and metadata applications, Berlin 22-26 September. Presentation retrieved 28 February 2009, from http://www.slideshare.net/morageyrie/dublin-core-metadata-initiative-education-application-profile-task-group-meeting-berlin-2008

DCMI (2008). *Dublin Core Metadata Element Set, version 1.1.* DCMI Recommendation. Retrieved 28 February 2009, from http://www.dublincore.org/documents/dces/

DCMI (2008b). *DCMI Metadata Terms.* DCMI Recommendation. Retrieved 28 February 2009, from http://dublincore.org/documents/dcmi-terms/
DCMI (no date). *Joint DCMI/IEEE LTSC Taskforce.* Retrieved 28 February 2009, from http://dublincore.org/educationwiki/DCMIIEEELTSCTaskforce
DCMI (no date b). *DC-Education application profile*. Retrieved 28 February 2009, from http://dublincore.org/educationwiki/DC_2dEducation_20Application_20Profile
Digital Library Federation (no date). *Best practices for OAI data provider implementations and shareable metadata.* Tech Rep. Retrieved 28 February 2009, from http://webservices.itcs.umich.edu/mediawiki/oaibp/
Duval, E., Hodgins, W., Sutton, S., & Weibel, S. (2002) *Metadata Principles and Practicalities*, D-Lib Magazine, Vol 8 (4) Retrieved 23 September 2009, from http://www.dlib.org/dlib/april02/weibel/04weibel.html
European SchoolNet (no date) *European Schoolnet*. Retrieved 28 February 2009, from http://www.eun.org/
Friesen, N. (2004). *Final Report on the "International LOM Survey".* Tech. rep. Retrieved 28 February 2009, from http://staff.uow.edu.au/eteaching/str/metadata/intlLOmsurvey.pdf
Friesen, N. (2006). *Building a better LOM.* CanCore tech. rep. Retrieved 28 February 2009, from http://www.cancore.ca/betterlom.html
GLOBE (no date). *Global Learning Objects Brokered Exchange (GLOBE).* Retrieved 28 February 2009, from http://www.globe-info.org/
Godby, C. J. (2004). What do application profiles reveal about the learning object metadata standard? *Ariadne*, (41).Retrieved 28 February 2009, from http://www.ariadne.ac.uk/issue41/godby/
HR-XML Consortium (no date). HR-XML Retrieved 28 February 2009, from http://www.hr-xml.org/
IEEE (2002). 1484.12.1—2002, *Standard for Learning Object Metadata*. The Institute of Electrical and Electronics Engineers, Inc.
IEEE (2005). 1484.12.3—2005, *Standard for eXtensible Markup Language (XML) Binding for Learning Object Metadata data model*. The Institute of Electrical and Electronics Engineers, Inc.
IEEE (2007). 1484.20.1—2007, *Standard for learning technology - data model for reusable competency definitions.* The Institute of Electrical and Electronics Engineers, Inc.
IMS (no date, a.) *Content Packaging Specification.* Retrieved 23 September 2009, from http://www.imsglobal.org/content/packaging/
IMS (no date, b.) *Question & Test Interoperability Specification.* Retrieved 23 September 2009, from http://www.imsglobal.org/question/
IMS (no date, c.) *Learning Object Discover and Exchange Project Group*. Retrieved 28 February 2009, from http://www.imsproject.org/lode.html
ISO (2003). ISO Standard 15836—2003, *Information and documentation - The Dublin Core metadata element set.* The International Organization for Standardization.
JORUM (no date). *JORUM*. Retrieved 28 February 2009, from http://www.jorum.ac.uk/
Library of Congress (2007). *SRU version 1.2 specifications.* Retrieved 28 February 2009, from http://www.loc.gov/standards/sru/specs/
Manola, F. & Miller, E. (2004). *RDF Primer.* Retrieved 28 February 2009, from http://www.w3.org/TR/rdf-primer/
Nilsson, M. (2008). *Harmonization of metadata standards.* Tech. rep., PROLEARN Consortium. Retrieved 28 February 2009, from http://ariadne.cs.kuleuven.be/lomi/images/5/52/D4.7-prolearn.pdf

Nilsson, M., Baker, T., & Johnston, P. (2008). *The Singapore framework for Dublin Core application profiles*. DCMI Recommended Resource. Retrieved 28 February 2009, from http://dublincore.org/documents/singapore-framework/

Nilsson, M., Baker, T., & Johnston, P. (2009). *Interoperability levels for Dublin Core metadata* DCMI Recommended Resource. Retrieved 28 September 2009, from http://dublincore.org/documents/interoperability-levels/

NISO (2004). *Understanding Metadata.* Retrieved 28 February 2009, from http://www.niso.org/publications/press/UnderstandingMetadata.pdf

Open Archives Initiative (2002). *The Open Archives Initiative protocol for metadata harvesting version 2.0.* Retrieved 28 February 2009, from http://www.openarchives.org/OAI/openarchivesprotocol.html

Powell, A., Nilsson, M., Naeve, A., Johnston, P., & Baker, T. (2007). *DCMI abstract model.* DCMI Recommendation. Retrieved 28 February 2009, from http://dublincore.org/documents/abstract-model/

Vanderfeesten, M., Summann, F., & Slabbertje, M. (eds) (2008) *DRIVER Guidelines 2.0 : Guidelines for content providers - Exposing textual resources with OAI-PMH.* Tech. Rep. DRIVER project. Retrieved 28 February 2009, from http://www.driver-support.eu/documents/DRIVER_Guidelines_v2_Final_2008-11-13.pdf

Wilson, S. (2008) *CEN endorses European metadata for learning opportunities* Retrieved 28 February 2009, from http://zope.cetis.ac.uk/members/scott/blogview?entry=20081021140752

Acknowledgements

Figure 2 is a copy of a diagram from DCMI, Copyright © 2008 Dublin Core Metadata Initiative. All Rights Reserved. http://www.dublincore.org/about/copyright/

Figure 3 was created by Sarah Currier, John Robertson, Phil Barker and Lorna Campbell and is licensed under Creative Commons Attribution, Share Alike (CC-By-SA) http://creativecommons.org/licenses/by-sa/2.0/.

Licence