

## **D-Lib Magazine**

January/February 2010 Volume 16, Number 1/2 Table of Contents

## The Use of Metadata for Educational Resources in Digital Repositories: Practices and Perspectives

Dimitrios A. Koutsomitropoulos, Andreas D. Alexopoulos, Georgia D. Solomou, Theodore S. Papatheodorou High Performance Information Systems Laboratory, School of Engineering, Dept. of Computer Engineering and Informatics, University of Patras

{kotsomit, alexopoulo, solomou, tsp}@hpclab.ceid.upatras.gr

## Abstract

The wide availability of educational resources is a common objective for universities, libraries, archives and other knowledgeintensive institutions. Although generic metadata specifications (such as Dublin Core) seem to fulfill the need for documenting web-distributed objects, educational resources demand a more specialized treatment and characterization. In this article we focus on the use of learning-object specific metadata in digital repositories, as they are primarily incarnated in the LOM (learning object metadata) standard. We review relevant standards and practices, especially noting the importance of application profiling paradigms. A widespread institutional repository platform is offered by DSpace. We discuss our implementation of LOM metadata in this system as well as our interoperability extensions. To this end, we propose a potential LOM to DC mapping that we have put into use in DSpace. Finally, we introduce our implementation of an LOM ontology, as a basis for delivering Semantic Web services over educational resources.

## 1. Introduction

An implied requirement of today's World Wide Web is to provide mechanisms that store, manage and discover resources in an efficient way. This becomes a necessity for all kinds of electronically available resources, the number of which is increasing at an extremely rapid rate. Digital repositories are mechanisms that fulfill this requirement, provided that they utilize appropriate metadata schemata for the characterization of their content.

The work described in this article is mainly focused on the field of education, and we investigate metadata standards that are utilized by digital repositories, primarily intended for the preservation and management of educational and research material. Our main objective is to improve such mechanisms by enriching them with learning object metadata. A well-known and simple to use metadata element set that we particularly focus on is the IEEE LOM standard [6].

A popular system that operates as a digital repository for educational purposes is DSpace [<u>12</u>]. DSpace uses the qualified Dublin Core (DC) element set as its base metadata schema. However, because the DC schema is sometimes proven to be inadequate for the efficient characterization of educational material, we attempt the development of an application profile, extended with the LOM metadata standard, and we tailor it to the needs of an educational repository. At the same time, we propose an ontology using these LOM elements that helps to better capture the semantic notions of the underlying concepts and forms the basis for the deployment of digital repositories with advanced services.

# 2. Characterizing Learning Objects

Digital repositories, used by universities, libraries, archives and other education specific institutions, have the responsibility for dlib.org/dlib/january10/.../01kout.html

efficiently handling learning objects. In the *IEEE Draft Standard for Learning Object Metadata* [<u>6</u>], a learning object is defined as "any entity – digital or non-digital – that may be used for learning, education or training". The use of educational metadata and relevant application profiles is necessary for the description of learning objects.

# 2.1 Educational Metadata Standards

Several educational metadata schemata have been proposed over time in order to better characterize learning objects. A widely adopted metadata element set for this purpose is IEEE LOM, a standard which has been designed especially for the description of educational resources. According to Al-Khalifa and Davis [1], an important feature of LOM is that it is simple to use and has an inherent extension capability. This extensibility allows for the easy incorporation of new elements and enables LOM to meet the specific needs of applications.

IEEE LOM defines a hierarchy of elements that are grouped into nine categories: *General, Lifecycle, Meta-metadata, Technical, Educational, Rights, Relation, Annotation,* and *Classification*. Each category is comprised of sub-elements that have some basic characteristics in common and appear either as a single element or as an aggregation of other elements. The complete LOM element hierarchy is presented in Figure 1.



Figure 1: The element hierarchy of the IEEE LOM standard.

Apart from LOM, other well known metadata schemata designed to serve similar needs in the field of education are IMS [13], ARIADNE [14] and Dublin Core [3]. Their basic characteristics are summarized in Table 1.

NAME	CHARACTERISTICS
IMS	<ul> <li>A further work on IEEE LOM</li> <li>Elements can be mapped to DC</li> </ul>
ARIADNE	<ul> <li>A set of 47 elements, 27 of which can be directly mapped to LOM elements</li> <li>Organized in six categories: <i>General, Semantics, Pedagogical, Technical, Indexation,</i> <i>Annotation</i> [11]</li> <li>Fully compatible with IEEE LOM</li> </ul>
Dublin Core (DC)	• A set of 15 core elements that can be further refined using attributes (qualifications)

dlib.org/dlib/january10/.../01kout.html

- A general metadata standard, suitable for describing digital objects of any kind
- DC-Terms [2] constitute the most up-to-date and formal version of the metadata terms properties roughly correspond to the whole set of DC elements plus their qualifications

Table 1: Metadata schemata for learning objects.

# 2.2 Application Profiles

The adoption of a sole metadata schema is usually an inadequate way to efficiently characterize learning objects. As a solution to this problem, the use of *Application Profiles* [5] is proposed. According to Duval et al. [4], an application profile is an aggregation of metadata elements selected among one or more metadata schemata and combined into a new one. The goal is to adequately fulfill the application's special needs and to retain interoperability with its base schema. Therefore, applications that utilize application profiles can benefit by exploiting the features of an existing schema and by enriching them with desired characteristics. Some popular application profiles are CanCore [17], GEM [19], UK Learning Object Metadata Core [18], DC-Ed [16] and SCORM [15]. These profiles are briefly presented in Table 2.

NAME	CHARACTERISTICS		
Canadian Core (CanCore) Learning Resource Metadata Application Profile	<ul> <li>Designed to facilitate efficient description and discovery of educational resources (mainly in Canada)</li> <li>Based on and fully compatible with IEEE LOM and IMS</li> <li>Includes eight of the total of nine of the LOM schema's categories: <i>General, Lifecycle, Meta-Metadata, Technical, Educational, Rights, Relation</i> and <i>Classification</i></li> </ul>		
The Gateway to Educational Materials (GEM) Metadata Element Set	<ul> <li>Started as an effort to facilitate access to online learning objects</li> <li>Based on DC and enriched with education-specific elements</li> </ul>		
DC-ED	<ul> <li>Defined by the DCMI Education Community with attention to educational needs</li> <li>Incorporates the core DC elements, recommended qualifiers and some IEEE LOM elements</li> </ul>		
UK Learning Object Metadata Core	<ul> <li>Based on LOM</li> <li>Particularly optimized so as to meet the needs of the British educational system</li> </ul>		
Sharable Content Object Reference Model (SCORM)	<ul> <li>Integrates a set of related technical standards and specifications, including the IEEE LOM</li> <li>Defines a specific content model for ensuring interoperability of the designed content across various learning management systems</li> </ul>		

Table 2. Application profiles for learning objects.

# 3. The LOM Standard in DSpace

In this section we briefly introduce DSpace and the metadata that this system uses out-of-the-box. Then, we explain the importance of incorporating the LOM metadata in a DSpace-based digital repository that serves as an educational/institutional repository. Afterwards, we describe how such a modification in the inherent system's metadata schema affects several provided services, by introducing the case of the University of Patras live DSpace installation and by giving an overview of all implemented facilities.

# 3.1 DSpace and the University of Patras Installation

A widely known mechanism upon which many institutional repositories and digital libraries have been built is DSpace. This system offers a means to manage research materials and scholarly publications in a professionally maintained repository. The qualified version of the Dublin Core (DC) schema serves as the DSpace core metadata schema. This schema can be extended with additional qualifications and elements as well. But apart from DC, other metadata schemata can also be imported in DSpace thus enhancing its capabilities and expanding its applicability to a wider range of organizations and institutions. What is more, DSpace supports facilities that provide for greater visibility and accessibility of its content. The Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) [9] is one such facility and allows the exposure of the repository's metadata in XML format through the provided OAI-PMH interface.

The University of Patras institutional repository [20] has been deployed over the DSpace system, and hence it can exploit this particular system's characteristics. It operates as a digital repository responsible for the preservation and distribution of the University's research and educational material. It utilizes the DSpace inherent DC schema that we have enriched with educational metadata. More specifically, because of the repository's particular educational role, we have incorporated the LOM metadata schema into our institutional repository. The reason we opted for LOM is that this standard, as stated in Nair and Jeevan [10], includes "the minimal set of attributes needed to allow learning objects to be managed, located, and evaluated" and has proved to be a widely adopted and internationally recognized open standard for the description of learning objects. Moreover, as we have already mentioned in section 2.1, LOM is easily extensible, meaning that it is able to better satisfy the needs of an educational-specific repository.



Figure 2. The LOM schema in the DSpace system.

# 3.2 Implemented Facilities

The import of the LOM schema in the University of Patras institutional repository brought a number of changes to the underlying DSpace system. First, the system's registry was updated with the new schema. This task was accomplished manually, by importing each element separately, giving a name to each one and defining possible refinements. It is important to note, though, that we didn't incorporate all the LOM elements. We have chosen, instead, to adopt only a subset of the LOM elements that focus exclusively on the description of learning objects and, additionally, have no direct mapping to DC elements. These elements are *Version, Status, Interactivity Type, Learning Resource Type, Intended End User Role, Context, Difficulty, Typical Learning Time* and *Annotation*. The incorporation of the LOM metadata schema is reflected in the following DSpace facilities (Figure 2):

- *Item Submission Process* When a new item is imported into the DSpace system, a standard submission process is followed. This process is comprised of steps during which users have to characterize their material by filling its metadata into the supplied fields. Each field represents a specific DC element. In order for the LOM metadata elements to become utilizable, we added a new form (description step) to the submission process. The new form is only activated and appears to users when a learning object is submitted.
- Advanced Search Advanced search in DSpace allows for field-based search and Boolean combinations of search terms. The metadata fields among which the supplied keywords would be sought are provided in the form of a list. To this list we have added the "Educational Metadata" option, activating search among LOM elements as well.
- *Item View* The "Item View" page is where an item's metadata are shown, either in short or long format. In case we review a learning object, LOM metadata also appear.
- *Metadata Harvesting* DSpace supports OAI-PMH as a means for interoperable metadata harvesting. This harvesting facility is configurable as to what elements can be exported, supporting both simple and qualified DC. In order to ensure interoperability of our implementation to a maximum possible level, we initially chose to activate the OAI-PMH interface to export the institutional repository's metadata in qualified DC format. Because we implemented a mapping from the LOM schema (see section 4) to the qualified DC element set, it becomes possible to export LOM metadata through qualified DC.

In addition, we further extended the system's OAI facility, so that it can export LOM metadata directly as well. This feature is useful, for example, in LOM-conformant applications that are able to consume LOM directly.

# 4. Mapping LOM to DC: The creation of an LOM ontology

In order to incorporate the LOM schema in the University of Patras institutional repository, we first investigated which LOM elements were missing from the DSpace system. Afterwards, we performed a mapping to find exactly which LOM concepts correspond to existing DC elements. We then imported into the system's metadata registry only those LOM terms for which no direct correspondence to the underlying DC elements existed. By using this approach, we avoided repetitions of elements and thus the existence of multiple elements with similar semantic interpretations in the same schema.

For example, the LOM elements *Title*, *Language* and *Contribute* have a direct mapping to DC elements, and in particular they are mapped to dc.title, dc.language and dc.contributor respectively. The exact correspondence between these element sets can be seen in more detail in the *IEEE Draft Standard for Learning Object Metadata* [6]. The rest of LOM elements, for which no such direct correspondence to DC exists, were mapped to some DC-Terms properties that seem to better convey their meaning. The decision was made after considering both the IEEE LOM specification and the work suggested in the *IEEE Draft Recommended Practice for Expressing IEEE Learning Object Metadata Instances Using the Dublin Core Abstract Model* [7], which proposes a potential LOM to DC Abstract Model (DCAM) mapping.

As already mentioned in section 3.2, these LOM elements are *Version, Status, Interactivity Type, Learning Resource Type, Intended End User Role, Context, Difficulty, Typical Learning Time* and *Annotation* (nine in total). In the following subsections we explain in more detail how we implemented the mapping of these nine elements and how we constructed an LOM ontology on this basis. We then describe an example that shows how an educational resource, submitted to our repository system, finally obtains ontology-based metadata.

#### 4.1 Towards a LOM Ontology

Before implementing the mapping of the nine LOM elements, we created an ontology from these concepts. The reason we started with an ontology lies in the fact that this approach is a more efficient way to capture the semantic relations implied among them. What is more, it gave us the opportunity to further exploit LOM metadata is several semantic-aware extensions to the institutional repository, as described in Koutsomitropoulos et al. [8].

This ontology comprises of a set of classes corresponding to each LOM element that accepts a list of controlled values. The purpose of these classes is multiple:

- They serve as a means to group LOM vocabulary values into ontology classes. For example the values *{Student, Author, Teacher}* are used as fillers for the LOM element *Intended End User Role*. We then create the class lom:intendedenduserrole and assert that the instances *Student, Author* and *Teacher* belong to this class.
- 2. They precisely characterize the nature and type of these values so that, every time any such value is used, we know where it comes from and exactly to which LOM element it corresponds.
- 3. The mapping of these LOM elements to DC-Terms is unavoidably incomplete. For example, dcterms:extent represents, in principle, a more general-purpose relation than lom.typicallearningtime (see section 4.2). However, every time a corresponding DC term is used in this manner, we ensure that it accepts values originating only from the specific LOM class, thus disambiguating its intended purpose.
- 4. By grouping LOM values into classes, it is easy to relate them to DC-Terms classes by specifying exactly this generalization relation (i.e., using SubClassOf). For example, the class lom:intendedenduserrole is asserted as a specialized sub-type (subclass) of the dcterms:AgentClass, thus giving a more precise semantic interpretation between LOM values and DC-Terms notions.

The hierarchy of the produced LOM ontology can be seen in Figure 3.



Figure 3. LOM ontology hierarchy.

## 4.2 The Mapping Procedure

When we finished with the creation of the LOM ontology, we proceeded with the actual work of implementing a mapping from these nine LOM elements to the DC-Terms properties. This mapping, as mentioned earlier, was performed carefully, taking into account related work in this field and the corresponding LOM specification. For example, we put into the DSpace metadata registry the element lom.intendedenduserrole in order to represent the LOM element *Intended End User Role*. The closest DC term to this element is dcterms:audience. At the same time, we also augmented this mapping relation by declaring that the values of this element are of type lom:intendedenduserrole. In this way we explicitly define from which class the lom.intendedenduserrole may take values and insert an implicit reference to the corresponding LOM ontology class. Similarly, let's consider the LOM element *Typical Learning Type*. This element is represented in the metadata registry by lom.typicallearningtime, and we mapped it to a very close-in-meaning DC term, dcterms:extent. In order to clarify its actual value space, we also need to state that the values of this element are of type lom:typicallearningtime. The rest of the LOM elements were treated in the same way (Table 3).

LOM Element		DC-Terms Property
2.1 Version	lom.version	dcterms:hasversion
2.2 Status	lom.status	dcterms:type type="lom:status"
5.1 Interactivity Type	lom.interactivitytype	dcterms:instructionalMethod type="lom:interactivitytype"
5.2 Learning Resource Type	lom.learningresourcetype	dcterms:type type="lom:learningresourcetype"
5.5 Intended End User Role	lom.intendedenduserrole	dcterms:audience type="lom:intendedenduserrole"
5.6 Context	lom.context	dcterms:educationLevel type="lom:context"
5.8 Difficulty	lom.difficulty	dcterms:type type="lom:difficulty"

lom.annotation



dcterms:description

The proposed mapping can be seen as a means to further enhance semantic interoperability between institutional repositories that are able to expose their content metadata using interoperability facilities like the OAI protocol. OAI allows for the harvesting of a repository's metadata in any format, including both simple DC and qualified DC formats. Among the qualified DC metadata that are being harvested, the "mapped" LOM elements are also included in a way that minimizes loss of semantics. So, apart from DSpace, it is possible for any other OAI-compliant application, extended with LOM, to easily make these metadata available to service providers.



Figure 4. The LOM metadata in DSpace. See <u>larger version</u> of Figure 4.

# 4.3 A Running Example

8 Annotation

Suppose we want to submit some kind of educational material in our — extended with LOM metadata — institutional repository. The consequent characterization of our resource as a learning object augments the submission process with an additional description step. During this step we are able to describe our object's educational characteristics by completing the necessary values in the corresponding text fields (left part of Figure 4). Each field represents a specific LOM element. When the submission process is successfully completed, a metadata record is created for this object in the DSpace system, as depicted in the right part of Figure 4. In this metadata record we can see which literal value corresponds to each LOM element. The corresponding schematic representation of (only a small part of) this metadata record, is given in Figure 5.

# Supplying LOM metadata



Figure 5. Schematic representation of a learning object's LOM metadata in DSpace.

What is more, we can use the OAI facility to expose the repository's content metadata. In particular, we choose to export them in their qualified DC format. In Table 4, we can see a part of the metadata record that was created for our object, as shown through the OAI-PMH interface. It is actually an XML representation of the object's metadata where the implemented LOM to DC mapping becomes evident. From this XML document we can easily produce an ontology document, provided that we apply the appropriate transformations. In combination with the LOM ontology, this document can be further utilized by reasoners and other Semantic Web compatible applications.

```
<record>
  < . . . >
      <dspace-ont:author xml:lang="en">Solomou, Georgia
     </dspace-ont:author>
      <dcterms:title>Introduction to Algorithms
      </dcterms:title>
      <dcterms:type type="dspace-ont:dspacetype" xml:lang="en">Learning Object
     </dcterms:type>
     <dcterms:educationLevel type="lom:context" xml:lang="en">Undergraduate
      </dcterms:educationLevel>
     <dcterms:type type="lom:difficulty" xml:lang="en">Medium
     </dcterms:type>
     <dcterms:audience type="lom:intendedenduserrole" xml:lang="en">Student
      </dcterms:audience>
     <dcterms:instructionalMethod type="lom:interactivitytype" xml:lang="en">Active
     </dcterms:instructionalMethod>
      <dcterms:type type="lom:status" xml:lang="en">Draft
     </dcterms:type>
      <dcterms:extent type="lom:typicallearningtime" xml:lang="en">One semester
     </dcterms:extent>
 <...>
</record>
```

Table 4. A part of a learning object's metadata record in DSpace, as exposed through the OAI-PMH interface.

For example, in Figure 6 we can see how our learning object's LOM metadata can be represented in an ontological context. The correspondence of these elements to DC-Terms properties is also apparent. In contrast to Figure 5, which depicts correlations between elements and mere literal values, we can now see how this particular object is related to LOM classes and subclasses, using the LOM ontology.



Figure 6. Schematic representation of a learning object's ontology-based metadata in DSpace.

## 5. Conclusions

After a brief overview of existing metadata standards and application profiles, it is apparent that IEEE LOM constitutes a solution for the characterization of educational resources. By incorporating this schema inside a repository specific to educational purposes, like our DSpace-based institutional repository, we see how it can be utilized in practice, and how it contributes to the efficient description and retrieval of learning objects.

The incorporation of the LOM schema required a careful process through which we investigated exactly which LOM concepts were missing from the system's metadata registry. Those concepts were imported into the repository's inherent metadata schema and were mapped to the appropriate DC-Terms properties. In addition, we extended the DSpace OAI harvesting facility to make it possible to expose these newly adopted LOM metadata. This idea can be similarly applied in any other OAI-compliant repository, resulting in the interoperable export and reuse of its educational metadata.

As a means to better capture the semantics of the relation between LOM and DC, we have also proposed the basis for an LOM ontology. This ontology explicitly represents LOM filler-values and vocabularies as proper entities, while allowing their classification and association with DC-Terms notions under well-defined relations. Therefore, it can be further utilized to enable semantic-aware services and suggests a means for providing semantic interoperability.

#### References

[1] Al-Khalifa, H., & Davis, H. (2006). The evolution of metadata from standards to semantics in E-learning applications. In *Proc.* of the 17th ACM Conference on Hypertext and Hypermedia.

[2] DCMI Usage Board (2008), DCMI Metadata Terms. DCMI Recommendation. http://dublincore.org/documents/dcmi-terms/.

[3] DCMI (2008). *Dublin Core Metadata Element Set, Version 1.1*. DCMI Recommendation. http://dublincore.org/documents/dces/.

[4] Duval, E., Hodgins, W., Sutton, S., & Weibel, S. (2002). Metadata Principles and Practicalities. *D-Lib Magazine*, 8(4). doi:10.1045/april2002-weibel.

[5] Heery, R., & Patel, M. (2000). Application Profiles: Mixing and Matching Metadata Schemas. *Ariadne, Issue 25.* <u>http://www.ariadne.ac.uk/issue25/app-profiles/</u>.

[6] IEEE LTSC. (2002). Draft Standard for Learning Object Metadata (IEEE 1484.12.1-2002).

http://ltsc.ieee.org/wg12/files/LOM\_1484\_12\_1\_v1\_Final\_Draft.pdf.

[7] IEEE LTSC. (2008). Draft Recommended Practice for Expressing IEEE Learning Object Metadata Instances Using the Dublin Core Abstract Model (IEEE P1484.12.4/D1). <u>http://dublincore.org/educationwiki/DCMIIEEELTSCTaskforce?</u> action=AttachFile&do=get&target=LOM-DCAM-newdraft.pdf.

[8] Koutsomitropoulos, D., Solomou, G., Alexopoulos, A. & Papatheodorou, T. (to appear). Knowledge Management and Acquisition in Digital Repositories: A Semantic Web Perspective. In *Proc. of the 2009 International Conference on Knowledge Management and Information Sharing (KMIS 2009)*.

[9] Lagoze, C., Van de Sompel, H., Nelson, M., & Warner, S. (2002). *The Open Archive Initiative Protocol for Metadata Harvesting*. <u>http://www.openarchives.org/OAI/openarchivesprotocol.html</u>.

[10] Nair, S. & Jeevan, V. (2004). A Brief Overview of Metadata Formats. *DESIDOC Bulletin of Information Technology*, 24(4), pp. 3-11.

[11] Najjar, J., Duval, E., Ternier S., & Neven, F. (2003). Towards Interoperable Learning Object Repositories: the ARIADNE Experience. In *IADIS WWW/Internet 2003 Conference*, pp. 219-226.

[12] DSpace, http://www.dspace.org/.

- [13] IMS, http://www.imsproject.org/.
- [14] ARIADNE, http://www.ariadne-eu.org/.
- [15] SCORM, http://www.adlnet.org/.
- [16] DC-Ed, http://projects.ischool.washington.edu/sasutton/dcmi/DC-EdAP-7-18-06.html.
- [17] CanCore, http://www.cancore.ca/.
- [18] UK LOM Core, http://zope.cetis.ac.uk/profiles/uklomcore/.
- [19] GEM, http://www.thegateway.org/.
- [20] University of Patras institutional repository, http://repository.upatras.gr/dspace.

## About the Authors



**Dimitrios Koutsomitropoulos** is an Adjunct Assistant Professor at the Department of Computer Engineering and Informatics and a researcher at the High Performance Information Systems Laboratory (HPCLab), University of Patras. He has received a Ph.D. and a M.Sc. from the same department, specializing in knowledge management and discovery on the Web. He has participated in many R&D projects with European and national funding. His research interests include knowledge discovery, automated reasoning, ontological engineering, metadata integration, semantic interoperability and the semantic web, where he has published an important number of research articles and papers.



**Georgia Solomou** is currently pursuing a doctoral degree in Semantic Web technologies at the University of Patras in Greece. She holds a master's degree in computer science from the same University. She is a member of the High Performance Information Systems Laboratory (HPCLab) where she has taken part in many European research projects. Georgia's research interests are related to the Semantic Web standards and technologies, mainly focusing in the field of Digital Libraries.

#### 2010/01/20

The Use of Metadata for Educational R...



Andreas Alexopoulos has received his BSc in Computer Science from the Computer Science Department of University of Crete in Greece. He is currently a postgraduate student in the Department of Computer Engineering & Informatics of the University of Patras. His master thesis and his research interests are about educational metadata and interoperability in digital repositories.



Theodore Papatheodorou is a Professor at the Department of Computer Engineering and Informatics, University of Patras, since 1984. He has also served as the Department's Chair between 2005 and 2009. He is the head of the High Performance Information Systems Laboratory (HPCLab). He has received a Ph.D. in Computer Science in 1973 and a M.Sc. in Mathematics in 1971 from Purdue University as well as a B.Sc. in Mathematics from University of Athens in 1968. He has authored hundreds of scientific publications in several areas of Computer Engineering and Computer Science.

Copyright © 2010 Dimitrios A. Koutsomitropoulos, Andreas D. Alexopoulos, Georgia D. Solomou and Theodore S. Papatheodorou

doi:10.1045/january2010-koutsomitropoulos