

Sampson D. and Karampiperis P. (2004). Reusable Learning Resources: Building a Metadata Management System Supporting Interoperable Learning Object Repositories. In Rory McGreal (Editor), *Online Education Using Learning Objects*, Taylor & Francis Books Ltd.

Reusable Learning Objects:
Designing Metadata Management Systems supporting Interoperable
Learning Object Repositories

Demetrios G. Sampson and Pythagoras Karampiperis

Department of Technology Education and Digital Systems, University of Piraeus
150, Androutsou Street, Piraeus, GR-18534 Greece

and

Informatics and Telematics Institute, Centre for Research and Technology – Hellas
42, Arkadias Street, Athens, GR-15234 Greece

Tel: +30-210-6839916/17, Fax: +30-210- 6896082

sampson,pythk@iti.gr

Abstract

Metadata can significantly assist in effectively accessing and managing learning objects. Currently, there are many systems which are intended to collect, share and reuse distributed learning objects, presenting the end-user with a uniform interface to search, access and evaluate these resources. On the other hand, although a generally accepted international standard for describing educational material exists, namely the IEEE Learning Object Metadata standard, many Learning object Metadata Management Systems are still using other metadata models for describing learning objects or previous versions of the IEEE LOM standard. This chapter outlines the requirements for a learning object metadata management system. These requirements go beyond metadata authoring and management, supporting the educational community in the integration of learning objects through the management of interoperable learning object repositories.

DRAFT VERSION

Introduction

Worldwide, a vast amount of educational content is constantly being produced, in order to support learning and teaching, in a wide range of contexts such as school, academic, training or life-long learning. The rapid increase in the quantity of available learning objects makes it difficult to search or re-use them by adapting or integrating them into new educational contexts. This problem can be handled efficiently by using learning object metadata. With consistent descriptions of the learning objects characteristics, *searching* becomes more specific and in-depth; *managing* becomes simpler and uniform; and *sharing* becomes more efficient and accurate [1].

The evolving need of introducing learning object metadata for the description of learning objects has grown to become a requirement when designing and developing metadata management systems. Such systems can be general identified with the term Learning Object Metadata Management Systems (LOMMSs) and can be described as environments that can access, maintain and support learning object repositories in such a way that they provide all the necessary services required for efficient indexing, storing, searching and reuse of the stored metadata information.

Currently, there are many LOMM systems that are designed to collect, share and reuse distributed learning objects, presenting the end-user with a uniform interface to search, access and evaluate the resources, such as the ARIADNE Knowledge Pool System [2], the CAREO (Campus Alberta Repository of Educational Objects) <<http://www.careo.org>>, the U.S.-based Science, Mathematics, Engineering and Technology Education Digital Library <<http://www.smete.org>>, the Educational Network Australia <<http://www.edna.edu.au>>, the Gateway to Educational Materials (GEM) digital library <<http://www.geminfo.org>>, the Scottish electronic Staff Development Library

(SeSDL) <www.sesdl.scotcit.ac.uk>, the LearnAlberta Portal <www.learnalberta.ca>, the COLIS <www.edna.edu.au/go/browse/0>, the SMETE <www.smete.org>, the Multimedia Educational Resource for Learning and Online Teaching (MERLOT) <www.merlot.org>, the Universal Brokerage Platform for Learning Resources <www.educanext.org>, the World Lecture Hall <www.utexas.edu/world/lecture/>, the Globewide Network Academy <www.gnacademy.org>, the McGraw-Hill Learning Network (MHLN) <www.mhln.com> and others.

The main goal when designing a LOMMS is to achieve interoperability between similar systems, so as to be able to reuse the stored and managed information, both at a lower representation level (*physical level*) and at the level of description and organization (*logical level*). The first goal can be achieved using standard interchange technologies such as XML (Extensible Mark up Language). The second goal can be achieved by adopting commonly agreed learning technology specifications. Although today a generally accepted international standard for describing educational material exists, namely the IEEE Learning Object Metadata standard [3], many LOMMSs are still using other metadata models for describing learning objects [4, 5, 6] or previous versions of the IEEE LOM standard, e.g. the Campus Alberta Repository of Educational Objects (CAREO) is using the IMS Metadata version 1.2.2.

Furthermore, the internationalization of each specification defined by the CEN/ISSS Learning Technologies Workshop [7] as the sum of processes whose purpose is to facilitate search, evaluation, reusability, and processing of learning objects within a multicultural and multilingual scenario, lead to the existence of multiple translations of each specification, providing evidence that two systems may not be able to interact, even when they use the same learning object metadata specification.

A possible solution to this problem is to define mapping specifications between different metadata schemas or different translations of a specific one, but this implies extra effort and extra cost. Another less costly approach is to make use of methodologies that are capable of transforming the metadata schemas used for the description of learning objects, minimizing human intervention. Several such approaches are currently being applied on other fields [8] with impressive results. However, even if simulations results over a wide range of learning objects prove the efficiency of such solutions [9], there are few LOM management systems incorporating such techniques.

This chapter discusses issues related to the management of learning object metadata, identifying the design and functional considerations of a Learning Object Metadata Management System, encapsulating modules such as efficient mapping algorithms.

Learning Object Metadata

Today, the web community has embraced the collection and use of metadata to characterize and index learning objects. This has the potential to a semantically more accurate retrieval of information. In general, metadata is information about data. The four aspects of metadata usage are shown in Figure 1. The aspect “*search, browse, retrieve*” is driven by the human user’s need to answer questions about the usefulness of the retrieved information or the effectiveness of the browsing process. The aspect “*ingest, assure quality, and reprocess*”, is driven by the need to acquire high quality information with a precisely defined data dictionary, ensuring the logical integrity of the stored metadata information. The aspect “*application to application transfer*” is driven by the need to transfer without human intervention, information from one repository to another on different platforms using different metadata schema. The aspect “*store and archive*” is

driven by the need for efficient implementation of search and retrieval within an overall goal of total cost minimization, including both infrastructure and human resources.

With the recent approval of the Learning Object Metadata (LOM) specification as a standard by the IEEE [3], learning object metadata models have achieved a stable common reference that provides implementors and developers with a solid foundation for creating metadata infrastructures to meet the needs of educators and learners. Given the necessarily abstract nature of this standard, the task of adapting it to meet the specific and concrete needs of these stakeholders, requires interpretation, elaboration, extension, and in some cases, the specialization of both the syntax and semantics. Such processes lead to multiple elaborations and/or representations of the same standard, depending on the application (*application profiling*). This fact can affect interoperability between learning object repositories, and reusability of the stored learning objects. Hence, it identifies the need for learning object metadata (LOM) management infrastructures and environment that can support the twin goals of interoperability and reusability with the minimum human interference. The CanCore metadata application profile is one attempt to address this problem (See Chapter XXX).

Both from the higher information abstract level, that is knowledge sharing, and from the lower level, that is data sharing, the need identified is to manage and facilitate the ability to share learning objects from one repository to another, with different management engines and without human intervention.

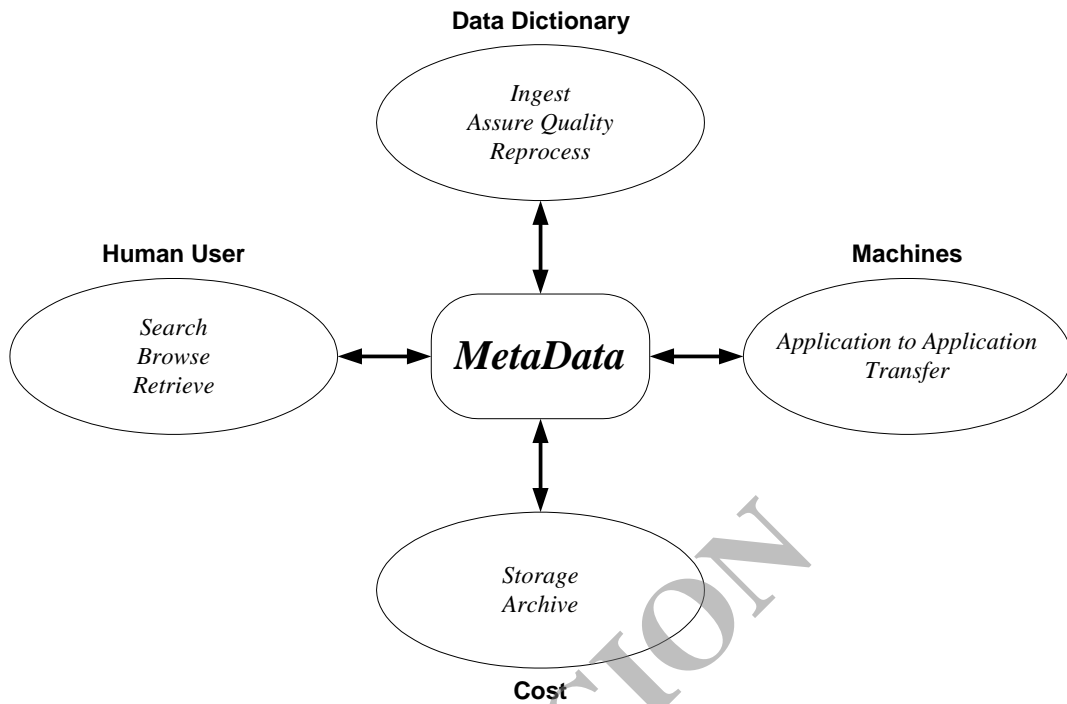


Figure 1: Aspects of Metadata Usage

Metadata Management Systems Needs and Requirements

Metadata Authoring and Management Tools

The need for creating and managing metadata had lead to the development of numerous software tools addressing the needs of metadata authoring and management. These tools can be roughly classified into two major categories:

- **Generic XML Tools:** This category includes tools for the creation of XML files. XML is the most commonly used format for metadata of all types and is not restricted to educational uses. These tools usually support a number of functionalities, including the mapping of XML documents to other metadata formats, updating, structural validating, searching, and manipulating XML documents, etc. These tools are not specifically developed for educational purposes, but they can be used for the creation of LOM files, if the user imports the corresponding learning object metadata XML schema. This, however, requires substantial expertise in metadata technologies.

- **Learning Object Metadata Tools:** This category includes tools that have been specifically developed for educational purposes. They usually facilitate (through user-friendly interfaces) the creation of learning object metadata files that are based on a specific learning object metadata model. However, these tools usually do not support the management of metadata files or the validation of both the structure and the semantics. Even if such tools are designed for the management of LOM files, they support only one LOM model which makes it almost impossible to integrate data from other sources.

Limitations

Metadata tools can provide functionalities for meeting specific requirements; however some limitations are encountered when applying them in the education field. They are not always oriented to educational needs, or they require the editor to be an expert having prior knowledge of at least XML technologies or educational metadata standards. The educational community has not yet exploited the full potential of learning object metadata, since many learning objects are available on the WWW without metadata description. Furthermore, for those learning objects that are described through LOM files, their description can be based on different metadata models (e.g. IEEE LOM, IMS Metadata, Dublin Core). Even when LOM files are based on the same metadata model, they can still differ if they are developed through different metadata tools, due to the fact that different tools use different XML bindings.

Standardization and Conversion

The basic function that underlies systems intercommunication is the exchange of information. The major barrier that prevents system intercommunication, limiting the interoperability between metadata management systems, is the use of different

specifications that define the structure of the exchanged information (*standardization diversity*). However, assuming that two systems use the same standardization format, interoperability cannot be ensured if this common format is described in different natural languages (*internationalization problem*). In both cases, there are two possible ways to achieve interoperability between LOM management systems: either the use of a neutral, standardized format or a conversion between varying formats (either different specifications or different translations of the same specification) [10, 11].

Figure 2 shows the overall standardization and conversion costs depending on the standardization level. Standardization costs contain all the costs that are necessary to implement a standard [12], e.g., software costs, hardware costs and personnel costs. Obviously standardization costs are proportional to the level of standardization. The graph of overall conversion costs is reversed, since with high standardization hardly any conversion is necessary, whereas precise conversions between multitudes of specifications and guidelines cause comparatively high costs.

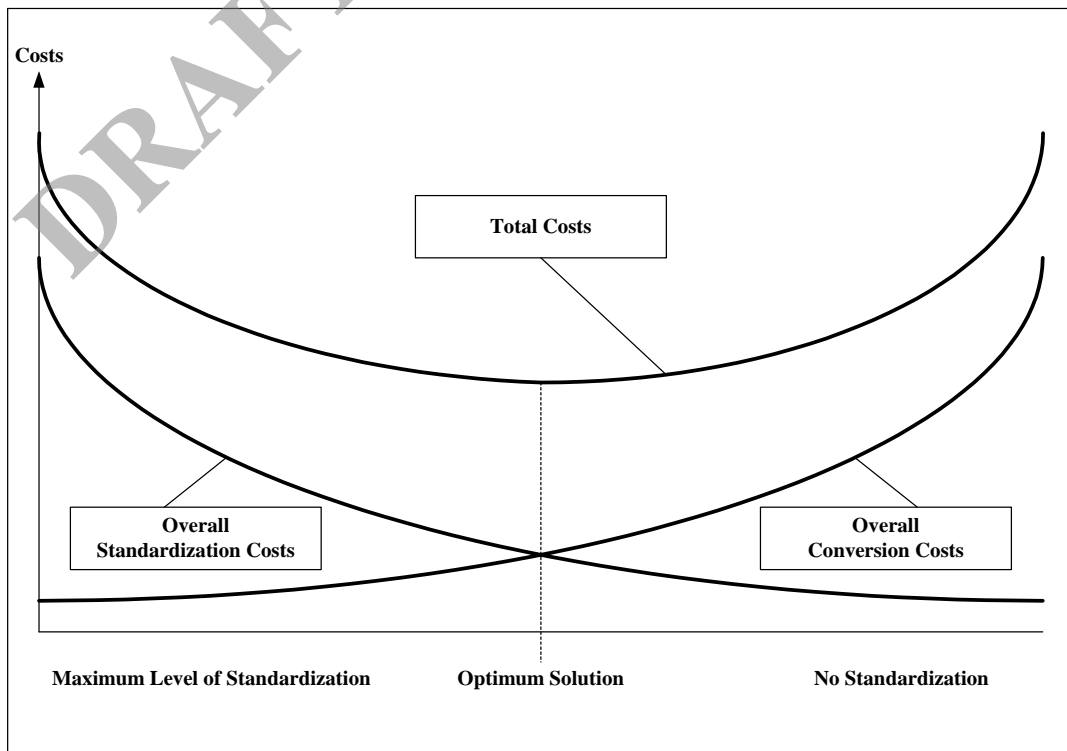


Figure 2: Trade-off between standardization costs and conversion costs

The overall conversion costs, as schematized in figure 2, are the sum of:

- **Costs for generating the converter:** These accrue through developing the necessary software and through acquiring a thorough knowledge of the data that has to be converted. The process of acquiring this kind of knowledge is mentioned by CEN/ISSS that addresses the problem of internationalization in the case of the IEEE LOM standard.
- **Costs resulting from an insufficient conversion result:** These costs can occur if the conversion instrument is error-prone or information loss could not be avoided. The probability of information loss obviously increases with the heterogeneity of the used metadata schemas. Costs resulting from insufficient results include expenses for manual post editing of the conversion result.

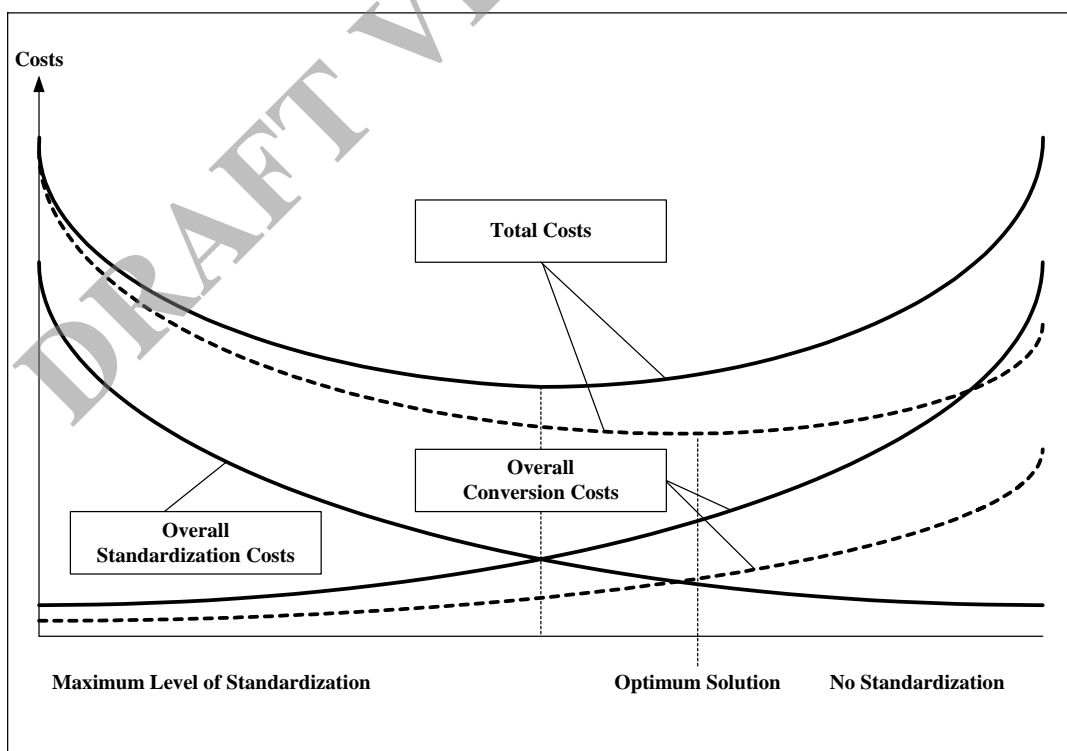


Figure 3: Trade-off between standardization costs and conversion costs when a schema-mapping algorithm is applied.

Apparently there is a trade-off between the overall costs of standardization and overall conversion costs. The optimum solution lies at the point where the sum of standardization and conversion costs is minimal. We make the assumption that the use of mapping algorithms implicates a right-shift of the overall conversion cost, as illustrated in figure 3. This right-shifting is due to the fact that the conversion costs have been reduced, since there is less need required for the user to interfere in the conversion process, thus reducing the cost of acquiring knowledge of the data that needs to be converted. On the other hand, this right-shifting also implies a further reduction of the needed standardization costs. This does not mean that no standardization is needed, but that fewer efforts are required to support the exchange of information in the context of internationalization.

Improving the interoperability between LOMMSs

The basis of many systems that integrate data from multiple sources is a set of correspondences between source and target schema. Correspondences express a relationship between sets of source attributes, possibly from multiple sources, and a set of target attributes. In real life scenarios it is very difficult to identify the correspondences since the metadata schemas are very complex and in most cases the attributes are related with each other not with a “one-to-one” relationship. Mapping mechanisms relieves users of that problem, by suggesting correspondences between source and target attributes almost automatically [9].

These mapping mechanisms can be roughly classified into two major categories: *Attribute-Driven*, when the mapping process is based on the names of the attributes and not on the values that they hold; and *Data-Driven* when the mapping process is based on the similarity of the data values that the attributes hold.

The Data-Driven mechanisms have better performance since the corresponding transformation maps can be the result of comparing more than one example. This property

does not exist in Attribute-Driven mechanisms, which produce the mapping only by comparing the name of the attributes between the two given schemas. The two categories of methods have comparable performance when only one input example is used by a Data-Driven mechanism.

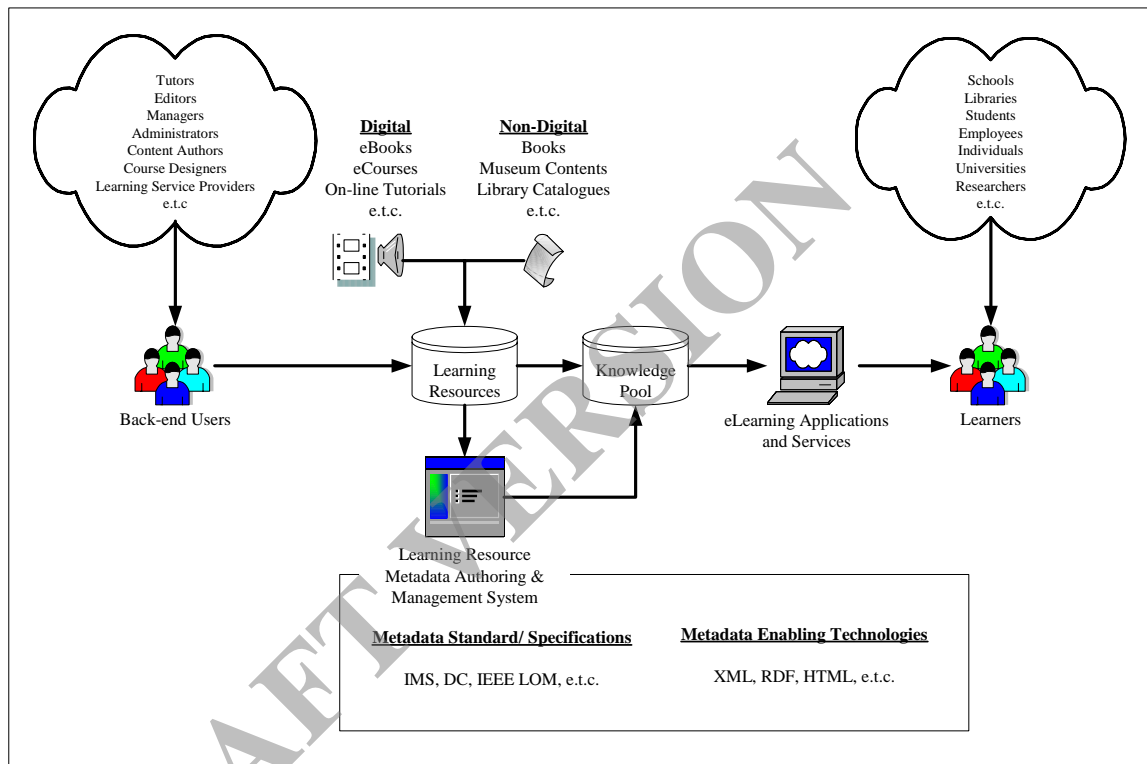


Figure 4: Dimensions of a Learning object Metadata Environment

Requirements for Learning object Metadata (LOM) Management

The main requirements for learning object metadata management can be derived from Figure 4, where the different actors, technologies, tools, etc., involved in metadata management are shown [13]. As it is shown in this figure, metadata management towards a collective and harmonized LOM repository requires the support of most common LOM standards / specifications; the creation of new and the modification of existing LOM files; the validation of metadata information; the support of most common metadata

technologies, etc. The design considerations of a LOM management system supporting the above requirements are briefly elaborated in Figure 5.

- **Support of most common learning object metadata models.** LOM management should support the creation of LOM files based on most common LOM standards/specifications. Moreover, a LOM management system should support the definition of new LOM models in order to support many application profiles over a specific metadata model.

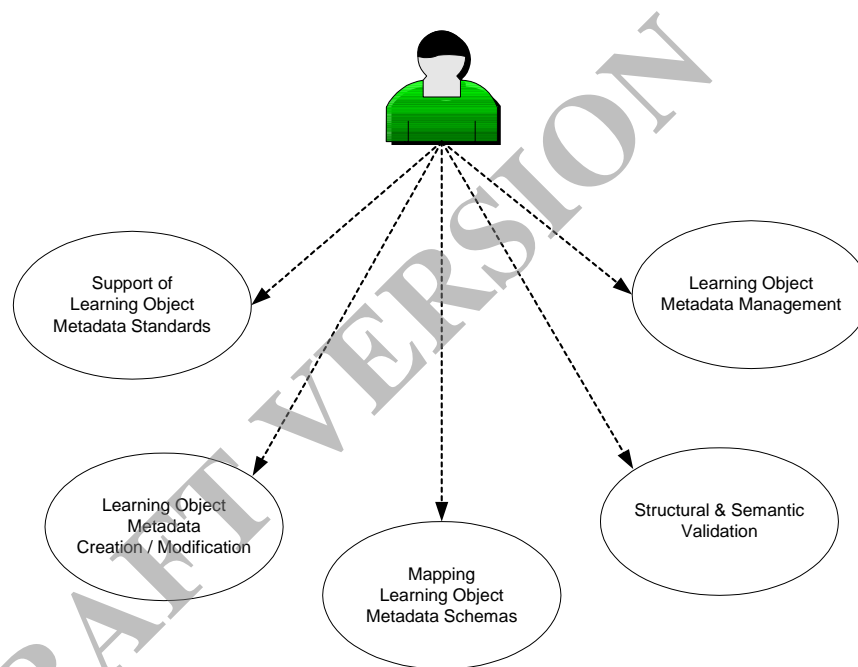


Figure 5: Design considerations of an LOM management tool

- **Creation/modification of learning object metadata files.** This is the most basic function in LOM management. The user should have the option to define a new or modify an existing LOM file according to any of the supported LOM standards and/or specifications (e.g., create an IEEE LOM). Moreover, since this function is mainly targeted to educational content authors, who are not necessarily experts in metadata technologies, it should be supported through a user-friendly interface (e.g., through wizards), providing help concerning the information that needs to be inserted into each LOM field.

- **Mapping of learning object metadata models.** The LOM files can be created according to a number of LOM models or different bindings. Therefore, a LOM management system should be able to map LOM files that are based on a specific LOM model or binding to another, in order to create a homogeneous and harmonized metadata repository.
- **Validation of learning object metadata.** One of the main problems with LOM files is that they may contain inaccurate information. Therefore, a LOM management system should facilitate the validation of the information in LOM files, when this is possible. The user should be informed if the entries in the fields are unacceptable (e.g., when text is inserted in fields where a number is expected). In addition, LOM management should facilitate the validation of the structure of LOM files, concerning their conformance with the selected learning object metadata model.
- **Metadata document management.** A LOMMS should provide the functional framework in order to support repository managers in finding, updating, deleting, sorting and grouping any set of LOM files through multiple document selections, multiple editing in LOM files, through a graphical interface that supports user friendly features like ‘drag & drop’.

In terms of non-functional requirements the system should meet the following principles [14]:

- **Modularity:** The system should consist of several independent modules.
- **Portability:** The system should be able to run in any platform.
- **Extensibility:** The system should be extendable (e.g., metadata specifications should be kept in a metadata schema repository and not be hard coded to allow import of new metadata specifications and allow translation of the interface language to other languages, etc.)

LOMMS Architectural Design

Figure 6 presents the architectural diagram of a LOM management system showing the structural components of the system and their interconnection paths. Interconnection between components is modeled by associations (directed arrows). The direction of each association shows which component initiates communication. These associations can represent direct connections or they can also be used to abstract away details of more complex connection and communication patterns (e.g. indirect communication based on events). Interfaces are shown by the round interface symbol and by adding dependency arrows between the interfaces and the components using them. The components of this architecture can be grouped into two different layers.

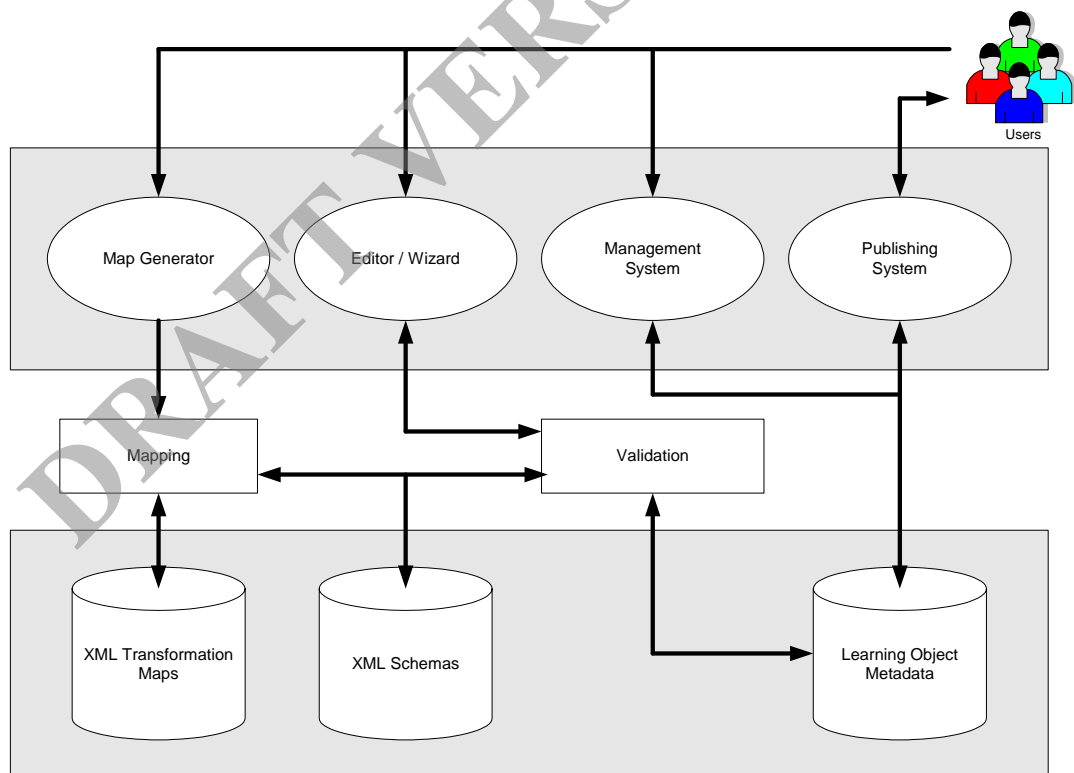


Figure 6: Architecture Diagram of a LOMMS

- Interface layer: A layer visible by the users of the LOMMS. It contains all the components of the user-interface. These are the XML editor/wizard, the management interface, the publishing interface and the map generator.

- A layer non-visible by the users of the LOMMS. It contains all the repositories involved and the operations, which are performed. The repositories involved are the Learning object Metadata repository, the XML Schema repository and the XML Translation Maps repository and the operations are validation and mapping.

Repositories

Every XML metadata file is accompanied by an XML Schema. The purpose of an XML Schema is to define the legal building blocks of an XML document. It defines the document structure with a list of legal elements and additional information such as the type of the elements. An XML Schema file can be declared in the XML document, or as an external reference. To store both the learning objects metadata and the corresponding schemas, a LOMMS has to use three different repositories:

- **XML Schemas Repository:** This repository is a system directory containing all the XML Schema files for the corresponding learning object metadata standards/specifications that the LOMMS supports.
- **Learning Object Metadata Repository:** This repository stores the metadata description of the learning objects. This repository is an XML-based database, whose information structure is inherited from the corresponding XML Schema. The best practice is to use internally only one database and transform the metadata structure if desired, through the mapping procedure, to every supported metadata schema, instead of maintaining one database per standard/specification, since then the number of required resources is dramatically increasing each time a new XML Schema needs to be imported.
- **XML Transformation Maps Repository:** As previously mentioned an LOMMS should allow mapping of XML files between learning object metadata models. Transformation maps should be automatically generated by a corresponding

mechanism, by associating a number of elements of one metadata schema to a number of elements of another schema.

Validation process

The validation mechanism of a LOM management system should provide two different types of validation: structural validation and semantic validation. Structural validation checks if the XML files conform to the element structure and hierarchy of the associated XML Schema files. On the other hand semantic validation checks if XML files conform to the associated XML Schema in terms of data types and/or vocabularies defined (Figure 7).

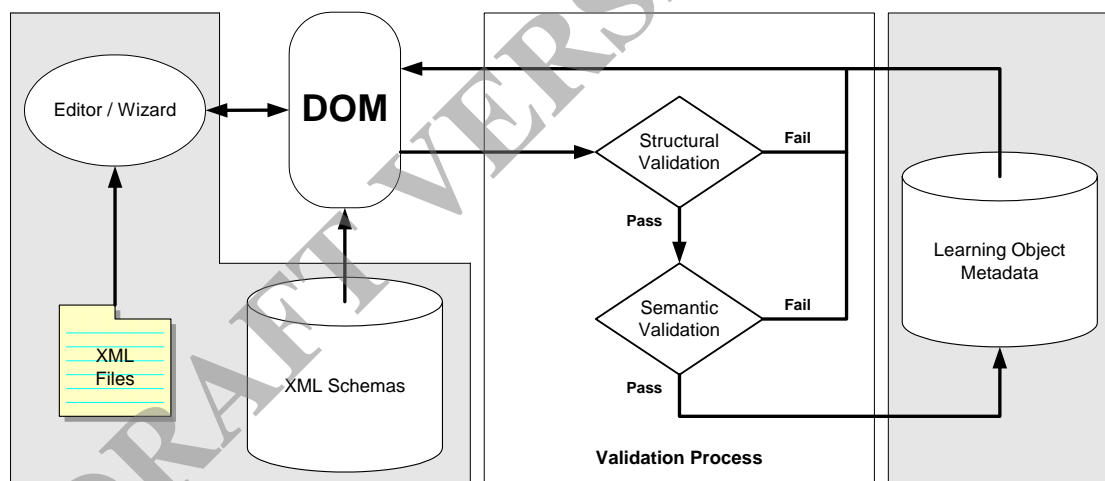


Figure 7: Validation process flow chart

Mapping process

Before a user can convert LOM from one metadata schema to another, the relationship between the nodes of the source metadata schema and the nodes of the corresponding destination metadata schema must be defined. A description of this relationship is stored in a special XML document called a transformation file or map. Each transformation file contains two things: the mapping between the nodes of the source schema and the nodes of the destination schema, and a skeletal XML document that

represents the structure for the results of the transformation. A transformation is a one-way mapping: from an XML schema or document to another XML schema or document. Once the user has created a transformation file that indicates how to transform an XML document based on a specific schema into another schema, he can create XML documents based on the desired schema for any XML document that conforms to the original schema used in the transformation. In Figure 8, the general diagram of the mapping procedure is shown.

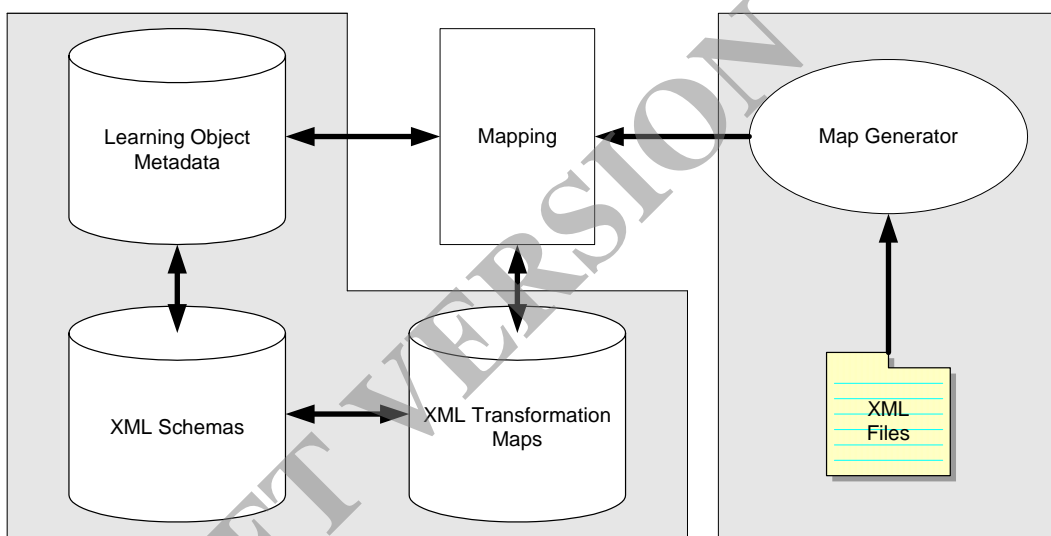


Figure 8: Mapping process of educational metadata schemas (A data-driven approach).

Conclusion

It is widely accepted that the use of metadata can improve the efficiency and effectiveness of information retrieval. In addition, it can provide the means for customized retrieval, based on user knowledge and preferences. This chapter outlined a number of reasons that make the use of learning object metadata essential in e-learning environments. It described the main design considerations that should be satisfied to provide an effective LOM management system that offers features such as the creation and modification of LOM files, structural and semantic validation, support of emerging XML technologies and

support of any learning object metadata model through the mapping between different metadata schemas.

References

- [1]. V. Torres da Silva and C. Jose Pereira de Lucena, "ContentNet: a framework for the interoperability of educational content using standard IMS", Computers and Education, Vol 37, Issue 3-4, 2001, p 273-295.
- [2]. Erik Duval, and Eddy Forte, "The Ariadne Knowledge Pool System", Communications of the ACM, May 2001, Volume 44, Issue 5.
- [3]. IEEE Draft Standard for Learning Object Metadata, IEEE P1484.12.1/d6.4, 2002.
- [4]. Dublin Core Metadata for Resource Discovery. Internet RFC 2413. URL address: <http://www.ietf.org/rfc/rfc2413.txt>, accessed January 2003.
- [5]. ARIADNE project. URL address: <http://ariadne.unil.ch>, accessed January 2003.
- [6]. S. Sutton, "Conceptual Design and Deployment of a metadata framework for educational resources on the Internet". Journal of the American Society for Information Science 50(13), pages 1182-1192, 1999.
- [7]. CEN/ISSS Information Society Standardization System, Learning Technologies Workshop, URL address: <http://www.cwnorm.be/iss/Workshop/lt/>, accessed January 2003.
- [8]. L. Popa, M.A. Hernandez, Y. Velegrakis, R.J. Miller, F. Naumann & H. Howard, "Mapping XML and Relational Schemas with Clio", Proceedings of the 18th IEEE International Conference on Data Engineering (ICDE) 2002.
- [9]. P. Karampiperis, K. Kastradas, and D. Sampson, "A Schema-Mapping Algorithm for Educational Metadata Interoperability", Proc. of the 15th World Conference on Educational Multimedia, Hypermedia and Telecommunications ED-MEDIA 2003, Honolulu, Hawaii, USA, June 23-28, 2003.
- [10]. E. Wustner, T. Hotzel & P. Buxmann, "Converting Business Documents: A Classification of Problems and Solutions using XML/XSLT", Proceedings of the 4th IEEE International

Workshop on Advanced Issues of E-Commerce and Web-Based Information Systems (WECWIS) 2002.

- [11]. Critchlow T., Ganesh M., Musick R., “Meta-Data Based Mediator Generation” In Proceedings of the 3rd IFCIS Conference on Cooperative Information Systems, August 1998, URL address: <http://citeseer.nj.nec.com/critchlow98metadata.html>, accessed January 2003.
- [12]. T. Buxmann, F. Weitzel, F. Westarp, & W. Konig, “The Standardization Problem in Networks – A General Framework”. In Jakobs K., (eds.): Standards and Standardization: A Global Perspective, Idea Publishing Group, 1999.
- [13]. Sampson D., Papaioannou V. & Karadimitriou P., “EM²: An Environment for Editing and Management of Educational Metadata”, Educational Technology & Society 5 (4) 2002, ISSN 1436-4522.
- [14]. H. Singh, “Achieving Interoperability in e-Learning”, Learning Circuits, March 2002, URL address: www.learningcircuits.org/mar2000/singh.html , accessed January 2003.

DRAFT VERSION