

Karampiperis P. and Sampson D. (2003). Enhancing Educational Metadata Management Systems to support Interoperable Learning Object Repositories. *In Proc. of the 3rd IEEE International Conference on Advanced Learning Technologies ICALT 2003*, Athens, Greece, pp. 214-218.

Enhancing Educational Metadata Management Systems to support Interoperable Learning Object Repositories

Pythagoras Karampiperis, Demetrios Sampson

*Advanced e-Services for the Knowledge Society
Research Unit,
Informatics and Telematics Institute,
Centre for Research and Technology Hellas,
42, Arkadias Street, Athens, GR-15234, Greece
e-mail: {pythk, sampson}@iti.gr*

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

Abstract

Educational metadata can significantly assist the effective retrieval of educational resources. Currently, there are a number of tools for educational metadata authoring, facilitating the manipulation of educational metadata files. This paper outlines the requirements, and proposes an architecture for the development of a complete LRM management system which go beyond metadata authoring, supporting the educational community in the full management of LRM through the management of interoperable learning objects repositories.

web-based environments that can access, maintain and support the learning resources repositories in such a way that they will provide all the necessary services required for efficient indexing, storing, searching and reuse of the stored information.

The main goal when designing a learning resource metadata based system is to achieve interoperability between similar systems and reusability of the stored and managed information, both at a lower technical representation level and at the level of description and organization. The first goal is achieved using standard interchange technologies such as XML (eXtensible Markup Language); while the second goal is achieved using learning technology specifications. The problem rising at this point is the following: although a generally accepted standard for describing educational material (IEEE Learning Object Metadata [2]) exists, many LRMMSs are still using other metadata models for describing learning resources; or previous versions of the IEEE standard. Furthermore, processes whose purpose is to facilitate search, evaluation, reusability, and processing of learning objects within a multicultural and multilingual scenario (defined as internationalization problem by CEN/ISSS [3]), lead to the existence of multiple translations of each specification or guideline, so even when two systems use the same educational metadata standard it's highly possible that they cannot interact.

This paper presents an architectural approach to address such issues, starting from the identification and definition of the design considerations of a web based LRMMS and deepening in the encapsulation of mechanisms to address more specific problems as efficient mapping algorithms to address interoperability. The paper is structured as it follows: initially, we will address the issues related with

1. Introduction

A huge amount of learning content is constantly produced worldwide in order to support learning and teaching, in a wide range of contexts (school, academic, training, life-long learning etc.). The rapid increase of learning resources makes it difficult for searching, managing and reusing; using metadata for describing the content is an efficient way to solve this problem. With consistent descriptions of the characteristics of learning resources, searching becomes more specific and accurate, managing becomes more simple and uniform, and sharing becomes more efficient and in-depth [1].

The evolving need for introducing educational metadata for the description of learning resources has grown to become a requirement for the design and development of metadata management systems that can effectively and efficiently perform a series of operations on the educational metadata. Such systems can be general identified with the term Learning Resource Metadata Management Systems (LRMMSs) and are described as the

educational metadata management, identifying so the design and functional requirements for educational metadata management systems. Next section presents and the architectural design of a metadata management system, meeting the requirements set in the previous section. The final section presents a prototype implementation of the discussed architecture; and a simulation example of handling the interoperability between management systems. To present its feasibility, we demonstrate the integrated solution to the specific problem of metadata schema heterogeneity, by incorporation of advanced mapping features in the educational metadata management system.

2. Design & Functional Requirements of a LRM Management System

2.1 Ensuring the interoperability between Learning Object Repositories

The basis of many systems that integrate data from multiple sources is a set of correspondences between source schemata and a 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 there may be many sources and the source relations may have many attributes. Users can get lost and might miss or be unable to find some correspondences. Also, in many educational metadata schemata the attribute names reveal little or nothing about the semantics of the data values. Only the data values in the attribute columns can convey the semantic meaning of the attribute. Mapping mechanisms relieves users of the problems of too many attributes and meaningless attribute names, by automatically suggesting correspondences between source and target attributes [4]. These 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.
- *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 map 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.

2.2 Requirements for Learning Resource Metadata (LRM) Management

The main requirements for educational metadata management, towards a collective and harmonized LRM repository requires: support of most common LRM standards / specifications, creation of new and modification of existing LRM files, validation of meta-data information, support of most common meta-data technologies, etc. The design considerations of a LRM management system supporting the above requirements are briefly elaborated below [7]:

- *Support of most common educational metadata standards*: LRM management should support the creation of LRM files based on most common LRM standards/specifications. Moreover, LRM management should support the definition of new LRM schemas.
- *Creation/Modification of educational metadata files*: this is the most basic function in LRM management. The user should have the option to define a new or modify an existing LRM file according to any of the supported LRM standards and/or specifications (e.g. create an IEEE LOM). Moreover, since this function is mainly targeted to educational resources authors, who are not necessarily experts in meta-data 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 LRM field.
- *Mapping of LRM standards*: the LRM files can be created according to a number of LRM standards. Therefore, LRM management requires that the user is able to map LRM files that are based on a specific LRM standard to any other.
- *Validation of semantic educational metadata*: one of the main problems with LRM files is that they can include inaccurate information. Therefore, LRM management should facilitate the validation of the information included in LRM 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, LRM management requires that the validation of the structure of LRM files, concerning their conformance with the selected LRM standard / specification.
- *Meta-data document management*: LRM management would also take into account the needs of LRM repository managers to find, update, delete, sort and group any set of LRM files through multiple document selections, multiple editing in LRM files, and with the help of a graphical interface including drag & drop features.

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

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

3. Architectural Design of LRM Management Systems

Figure 1 presents the architectural diagram of a LRMMS 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.

- *Interface layer*: A layer visible by the users of the LRMMS. 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 LRMMS. It contains all the repositories involved and the operations, which are performed. The repositories involved are the Learning Resource Metadata repository, the XML Schema repository and the XML Translation Maps repository and the operations are validation and mapping.

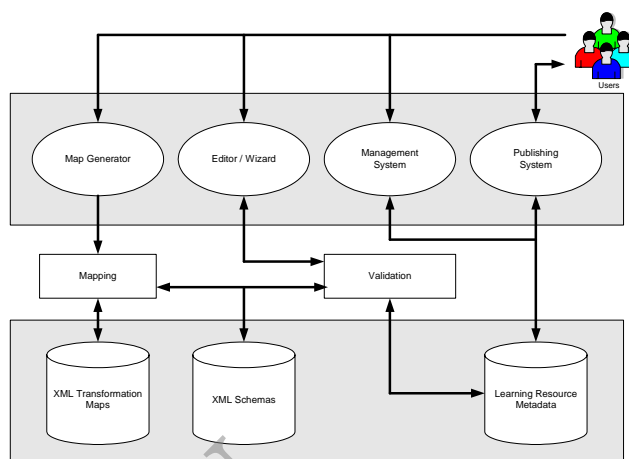


Figure 1: Architecture Diagram of a LRMMS.

Every metadata XML 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 resources metadata and the corresponding schemas, a LRMMS has to use three different repositories:

- *XML Schemas Repository*: This repository is a system directory containing all the XML Schema files for the corresponding educational metadata standards/specifications that the LRMMS supports.
- *Learning Resource 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 the number of required resources is dramatically increasing each time a new XML Schema is imported.
- *XML Transformation Maps Repository*: As previously mentioned an LRMMS should allow mapping of XML files between educational metadata specifications. 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.

4. A prototype implementation of LRMMS and simulation examples

4.1 EM2 - Educational Metadata Management Tool

In this section we will present an implementation of a LRM management system that complies with the introduced design considerations. The EM2 (Educational Metadata Management) system [7], provides several metadata services based on the IEEE LOM schema for Learning Resources Metadata authoring, indexing, storing and retrieval along with a number of advanced features and mechanisms for achieving interoperability between LRM repositories.

The case under study is the Metadata Schema-Mapping mechanism. The EM2 platform provides a complete, automatic environment that supports the creation of transformation files that indicate how to transform an XML document based on a specific schema into another schema and provides the mechanisms for this transformation (Figure 3 and 4).

4.2 Case Study: Transforming Metadata Repositories for Educational Objects

As it was introduced before, although a generally accepted standard for describing educational material (IEEE Learning Object Metadata) exists, many educational metadata management systems are using other metadata models or previous versions of the IEEE standard; or even different translations of the IEEE LOM. Therefore, in order to achieve the interoperability between educational metadata management systems, efficient tools and procedures that will enable the transformation between different metadata schemes should be integrated. In traditional data integration tools the users of the system are required to understand both their source's data and the target representation, in order to manually create a mapping between them. Other tools have been presented, trying to completely automate the mapping process but these tools are expecting from the user to describe with completely the same way the content of the different schemas [6]. The EM2 mapping mechanism except from being fully automatic, handles in a more precise way real case mapping problems such as the mapping between different educational metadata description schemes that uses not only different description schemas but also different standards for representing the content information. In this section we are going to present the simulation results of the EM2 mapping mechanism when is used to map one educational metadata schema to another.

In our simulation experiment we try to map different representations of the same real-world entity using the Dublin Core, Ariadne, Gem, and IEEE LOM educational metadata specification schemes. In order to examine the efficiency of the proposed algorithm, we have designed several datasets for each one of the five previously mentioned educational metadata schemes. For the designing of the testing datasets we used the ISO 639 and

ISO3166-1 standards as the language format scheme and the ISO 8601 standard as the date format scheme, according to all metadata schemes of our example.

In order to evaluate the total efficiency of the mapping mechanism we have designed three different evaluation criteria, which are defined by:

- Confidence = $\frac{\sum \text{Similarity of Content}}{\forall \text{mapping}} \div \text{Total number of mappings produced}$
- Success = $\frac{\text{Number of correct mappings}}{\text{Total number of mappings produced}}$
- Mistakes = $\frac{\text{Number of wrong mappings}}{\text{Total number of mappings produced}}$

All of them are used to evaluate a total criterion that is the mean value of confidence, success and mistakes for all of the different educational metadata description schemes. Since the EM2 mapping mechanism is data-driven, it is obvious that the efficiency of the mapping procedure depends on the similarity between the entity values of the different schemas, on which the algorithm is applied.

So, we have split the testing datasets in three categories according to the measure of the similarity between entity values, which is defined by:

$$\text{DataSetSimilarity} = \frac{\sum \text{Similarity}}{\forall \text{entity of the DataSet}} \div \text{Total number of entities in DataSet}$$

Thus, we have datasets with low (less than 40%), medium; and high (more than 70%) similarity (Figure 2).

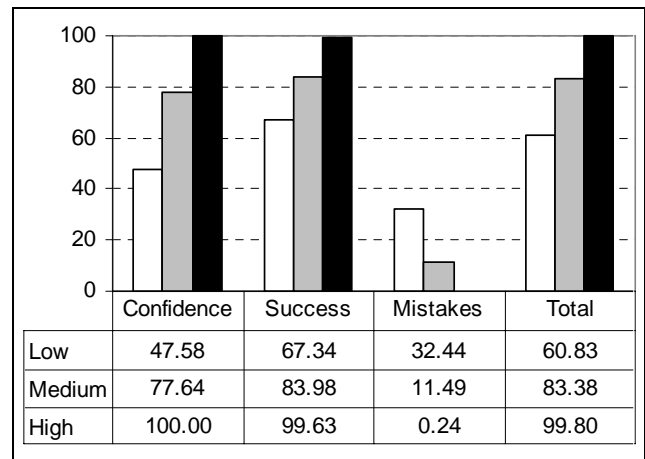


Figure 2: Confidence, Successful mappings, False mappings & Total Efficiency of the mapping procedure when it is applied to datasets with low, medium and high similarity between entity values respectively.

From the simulation results it is clear that the EM2 mapping mechanism succeeds in producing accurate

mappings, and thus it has a very good total evaluation indicator.

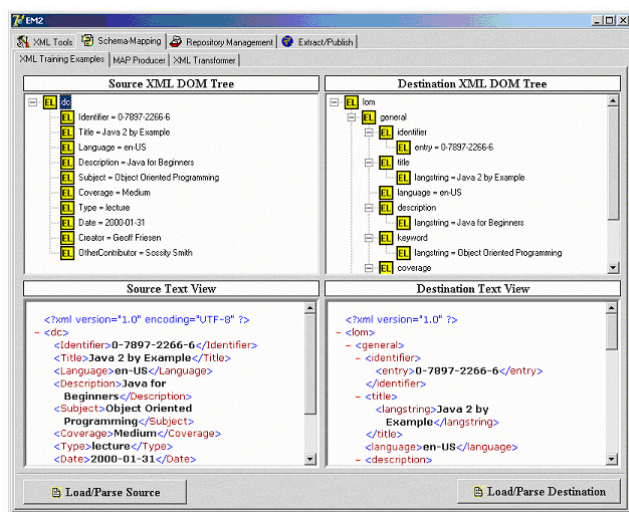


Figure 3: Mapping of IEEE LOM to Unqualified Dublin Core (XML DOM Tree).

In order to make a more extensive examination of the efficiency of the EM2 mapping mechanism, we created full datasets of LOM and Dublin Core [7] metadata schemes representing the same learning objects (Figures 3 and 4). The result of the mapping algorithm (Table 1) was according to Annex B (Mapping to Unqualified Dublin Core) of IEEE 1484.12.1-2002 (Draft Standard for Learning Object Metadata) standard, in which the mapping between LOM and Dublin Core metadata schemes is defined. This fact proves that the automatic, without the user interference, mapping between two different schemas is not impossible and also that this process can be as effective as the manual mapping of a very experienced user.

Table 1: Mapping between Unqualified Dublin Core and IEEE LOM.

Dublin Core		IEEE LOM
Data Element	Identifier	Data Element
DC.Identifier	1.1.2	General.Identifier.Entry
DC.Title	1.2	General.Title
DC.Language	1.3	General.Language
DC.Description	1.4	General.Description
DC.Subject	1.5	General.Keyword
DC.Coverage	1.6	General.Coverage
DC.Type	5.2	Educational.LearningResourceType
DC.Date	2.3.3	LifeCycle.Contribute.Date
DC.Creator	2.3.2	LifeCycle.Contribute.Entity
DC.OtherContributor	2.3.2	LifeCycle.Contribute.Entity
DC.Publisher	2.3.2	LifeCycle.Contribute.Entity
DC.Format	4.1	Technical.Format
DC.Rights	6.3	Rights.Description
DC.Relation	7.2.2	Relation.Resource.Description
DC.Source	7.2	Relation.Resource

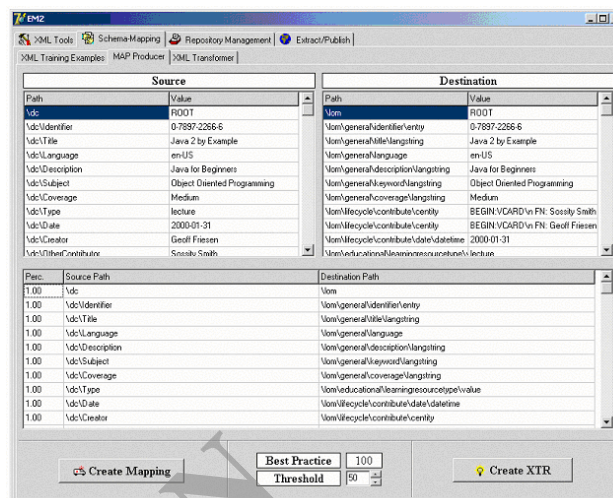


Figure 4: Mapping of IEEE LOM to Unqualified Dublin Core (Mappings).

5. Conclusions

It is widely accepted that the use of metadata can improve the efficiency and effectiveness of information retrieval from the web. In addition, it can provide the means for customized retrieval, based on user knowledge and preferences. This paper outlined the main design considerations that should be satisfied to provide an effective LRM management system. The discussed LRM management system architecture promotes the use of educational metadata specifications. It offers features such as the creation and modification of educational metadata, structural and semantic validation, support of emerging XML technologies, support of any learning resources standard or specification and mapping between different metadata schemas. Finally, we presented a prototype implementation of the discussed architecture and demonstrated the use of it on an educational schema-mapping problem. The results showed that the mapping mechanism except from being fully automatic, handles in a precise way real case mapping problems such as the mapping between different educational metadata description schemes that use not only different description schemas but also different standards for representing the content information, or the mapping between metadata descriptions that represent the same real-world entity in different languages.

6. 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]. IEEE Draft Standard for Learning Object Metadata, IEEE P1484.12.1/d6.4, 2002.

- [3]. CEN/ISSS Information Society Standardization System, Learning Technologies Workshop, URL address: <http://www.cwnorm.be/iss/Workshop/lt/>, accessed January 2003.
- [4]. 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.
- [5]. Papaioannou, V., Karadimitriou, P., Papageorgiou, A., Karagiannidis C., & Sampson, D. (2001). From Educational Metadata Authoring to Educational Metadata Management. In Proceedings IEEE Computer Society International Conference on Advanced Learning Technologies. Madison, Wisconsin, USA, August 6-8, 2001 (best paper award).
- [6]. 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.
- [7]. Dublin Core Metadata for Resource Discovery, Internet RFC 2413. URL address: <http://www.ietf.org/rfc/rfc2413.txt>, accessed January 2003.

DRAFT VERSION