Multimedia Learning Object Metadata Management and Mapping Tool

Xiaofei Liu, Abdulmotaleb El Saddik
Multimedia Communications Research Lab
University of Ottawa, Ottawa, Canada, K1N 6N5
(1-613) 562 5800-6277
{xiaofei, abed}@discover.uottawa.ca

ABSTRACT

The Resource Description Framework (RDF) is an attempt towards the semantic web vision. RDF integrates a variety of applications using XML for syntax and URIs for naming. There are a number of RDF bindings for metadata of different domains such as Dublin Core (DC), IEEE-LOM/IMS, ADL SCORM and MPEG7. Nowadays there are a good number of learning resources on the Internet. Their metadata descriptions are presented in RDF/XML as well as in pure XML. As a matter of fact, very few RDF-based metadata tools exist. To build flexible and open e-learning systems, it is required to combine IEEE-LOM with other new and existing metadata standards. As a result, metadata mapping between diverse metadata schemes is an important design goal. In this paper we present a novel Multimedia Learning Object Metadata Management and Mapping tool. $MMLOM^3$ is a distributed RDF-based metadata management tool, which supports learning object metadata authoring, editing, as well as data querying and storing based on RDF binding. In addition to these basic metadata management functions, $MMLOM^3$ encompasses two key features: metadata transformation from XML binding to RDF binding and vice versa; and semantic mapping between metadata schemes.

Categories and Subject Descriptors

H.3.4 [**Information Storage and Retrieval**]: Systems and Software – *Distributed systems, Information networks*.

General Terms

Design, Standardization

Keywords

Metadata management, Semantic web, Web Services, RDF.

1. INTRODUCTION

New generation e-learning systems should provide high quality instruction and decision assistance anytime, anywhere and should be tailored to a learner's needs. Using technology to integrate and deliver sharable content may be the best means to reach this goal. More and more, metadata is widely used to describe content in numerous ways. Depending on the type of content, many different metadata schemes have been developed all aiming at particular interests and goals. RDF is a framework for describing and interchanging metadata. It provides interoperability between applications that exchange machine understandable information on the Web. The problem of building RDF-based e-learning systems has received significant attention over the last few years.

1.1 Related Works

UNIVERSAL
An online world-class brokerage service linking educators and trainers for the exchange and distribution of learning resources [7, 10], is implemented as a non-distributed three-tier application with a centralized web application server that serves web client's needs and synchronously accesses a central data-store; it only supports IEEE-LOM metadata standards and gives some modification and in-depth definition based on it.
EDUTELLA
A multi-staged project to scope, specify, architect and implement an RDF-based metadata infrastructure based on the Java peer-to-peer architecture JXTA [8], whose initial services include: Mapping Service, Query Service, Replication Service, and Annotation Service. EDUTELLA's mapping service has the objective of mapping metadata between different schemes, but so far, no mapping method has been proposed. Although, EDUTELLA has its own query language (RDF-QEL-I) which contains several query levels, it is more difficult to implement this query language than to use a mature one.
MetaNet
A common ontology for semantic interoperability used as a general architecture to enhance metadata interoperability across the web [9].

1.2 Contributions

In this paper we present a novel multimedia learning object metadata management and mapping tool, called MMLOM[3]. It is a distributed RDF-based metadata management tool, which supports learning object metadata authoring, editing, and data querying and storing based on RDF binding. In addition to these basic metadata management functions, it encompasses two key features of MMLOM[3] which are metadata transformation from XML binding to RDF binding and vice versa, and semantic mapping between those metadata schemes in order to allow interoperability among the existing metadata standards, like DC, LOM, and MPEG7. To our best knowledge, this paper presents the first work using the technology of RDF to design an interoperable tool for a metadata management system in the context of e-learning applications. We also evaluate the current RDF querying languages and choose RQL for our metadata management system. Based on the comparison of several existing RDF querying and storing tools, we apply the Sesame server [1] as our metadata management platform, upon which advanced functions like transformation and mapping guarantee interoperability across multiple domains. The rest of the paper is organized as follows. Section 2 provides an overview of preliminaries of metadata management system and the RDF technology. Section 3 a functional model of MMLOM[3] is given. Section 4 addresses some key issues in the design and implementation of such a system. Section 5 discusses the architecture and each component of MMLOM[3] in details. Finally Section 6 summarizes the work and point out several future research issues.

2. BACKGROUND

2.1 Metadata Interoperability in E-learning Systems

The metadata standard for content in an educational domain is IEEE-LOM (Learning Objects and Metadata), which is inspired by the Dublin Core (DC). LOM does not support content structuring for textual content, nor for audio-visual content. However, it can borrow functionality from MPEG-7 in order to structure educational multimedia content into smaller objects, which in turn can be assigned keywords and/or semantic representations.

One of the key components in an e-learning system is an interoperable metadata management tool which provides easy access to different metadata schemes. In this paper the notion "interoperability" is defined as the possibility of unambiguous interchange between metadata, where components are able to share similar meanings for all systems and therefore implies that one particular system could (re)use different sorts of metadata at the same time, without confusion about their respective meanings or intentions of use. The technology of RDF provides a solution to this task.

2.2 RDF and RDF-based Learning Objects Metadata

The Semantic Web in which information is given well defined meaning, better enabling computers and people to work in cooperation, is an extension of the current web [13]. Figure 1 shows the Semantic Web architecture as layered structure in which XML forms the basis. The transport syntax, RDF, provides the information representation framework in the middle. On top of RDF, schemas and ontologies provide the logical apparatus necessary for the expression of vocabularies, enabling intelligent processing of information.

RDF is the language in which Semantic Web metadata statements are expressed and can be said to consist purely of so-called statements. An RDF statement consists of three elements: a subject, a predicate, and an object. The RDF Schema provides the basic vocabulary to express relationships between terms (resources being instances of terms, terms being subterms of other terms and so on) and also provides means to restrict the usage of predicates.

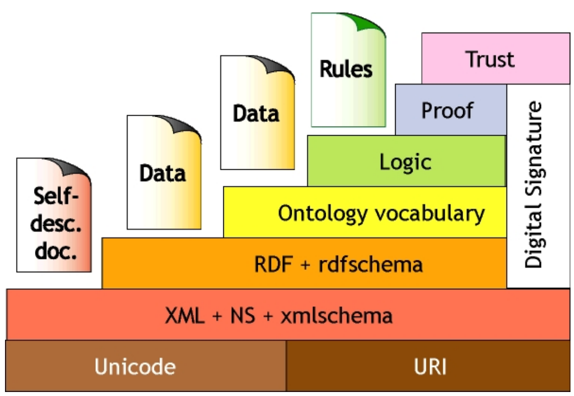

Figure 1. Layered structure of semantic web [17]

RDF which provides interoperability between applications that exchange machine understandable information on the Web [2], is the foundation for describing and interchanging metadata,. RDF is carefully designed to have the following characteristics: independence, interchange, and scalability. RDF is more flexible than XML in supporting metadata which by its very nature is subjective, distributed and expressed in diverse forms. There are many benefits of using RDF-based metadata descriptions in learning technologies [16]:

- The interoperability with other separate metadata standards is greatly increased as RDF allows a single storage model for very different types of data and schemas. Implementing a search mechanism that includes dependencies between metadata expressed in different schemas is therefore simplified.
- Reuse of existing metadata standards is greatly simplified; for example the VCard RDF binding [18] can be transparently included into LOM RDF binding [19] with no extra efforts.
- The relationship between metadata standards is represented in a machine-readable manner. For example, no conversion between Dublin Core metadata and LOM is required, and Dublin Core-aware tools can understand the Dublin Core parts of a given LOM metadata description.
- Several independent means of meaningful extensions (such as refinement, introduction and addition of new properties) are available in RDF, none of which cause interoperability problems On the other hand extending the XML binding is certainly possible using XML Schemas, the process would easily introduce interoperability problems.,.
- The graph representation and modularity of RDF effectively clean up the format and semantics of the specification for metadata that contain very complex interdependencies, such as IMS Content Packaging.
- A clean integration of different specifications is allowed in RDF, in a layered way. Currently, the work on IMS Content Packaging in RDF is built on top of the IMS Meta-data RDF binding, which is built on top of the VCard RDF binding and the Dublin Core RDF binding.

While RDF-binding of most learning resource metadata specifications exists, (e.g., DC, IEEE LOM/IMS, and ADL SCORM) there is also RDF-binding of metadata from other domains, (e.g., CIDOC in Cultural Heritage/Libraries domain [12], MPEG7 [6] in the multimedia domain, and MetaNet in cross domain [5]) in today's learning object repositories, the learning objects exist mainly in two formats, XML or RDF. However, the use of a metadata management tool in e-learning systems to achieve those benefits mentioned above has not been widely explored. Therefore, it is very important to build an RDF-based metadata management tool that supports flexible data exchange between different systems with different metadata schemes, as well as adaptive learning content sequencing and presentation with different user profile and different user devices.

3. FUNCTION MODEL OF MMLOM[3]

In our earlier study, we designed a web-service-based architecture for an e-learning system known as Ubilearn [14]. Ubilearn allows content developers to create, store, reuse, manage, and deliver digital learning content in a distributed environment and also supports those learning-activity related functions such as registration, evaluation, collaboration, etc.

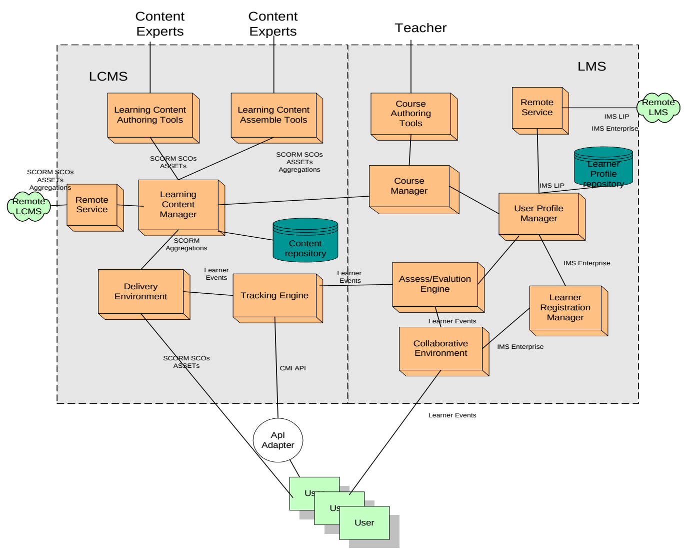

Figure 2. The function model of Ubilearn

In the presented work we will extend the results presented in [14] by means of refining the learning object metadata management system in Ubilearn and is achieved by our RDF-based metadata management tool MMLOM[3]. The architecture of MMLOM[3] is shown in Figure 2. The key role of the learning objects metadata management system in the whole e-learning system is to provide provides the most basic functions: indexing, storage, discovery (search), and retrieval of learning objects across multiple repositories, to support other tools in Ubilearn. For example, in order to assemble the learning object into a course, the assembly tool may need to search and browse the learning object metadata according to the user's profile, the delivery tool may need to search and get the metadata of the learning object in order to deliver the corresponding learning objects to the user.

A number of existing metadata schemes (such as DC, LOM, and MPEG7) may be coded in XML or RDF. In the spirit of allowing the e-learning environments to be interoperable, MMLOM^3 should be capable of exchanging learning objects with other repositories or systems. In addition to the basic metadata management functions as mentioned above, the other two key features that MMLOM^3 should possess are (1) metadata transformation from XML binding to RDF binding and vice versa, and (2) semantic mapping between the diverse metadata schemes.

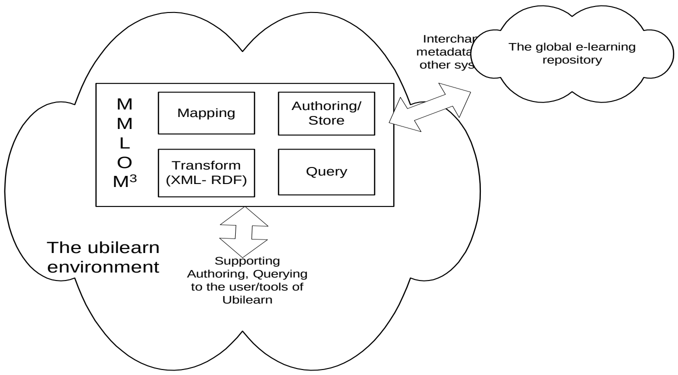

Figure 3. The functions of MMLOM3 in Ubilearn

MMLOM^3 is based on RDF and is the foundation for building a higher layer semantic for allowing learning objects to interoperate with user profile, device profile to support good quality learning process.

As a subsystem of the Ubilearn, MMLOM^3 should have a distributed and scalable architecture which can be integrated seamlessly with the Ubilearn system. Its main functions are:

- To provide learning objects metadata authoring for different formats, such as LOM, MPEG7, and DC.
- To store metadata in SQL database and/or export/import them to/from XML/RDF binding files
- To allow flexible searching via an RDF query language cross multiple repositories
- To support the mapping function between metadata schemes, including DC, LOM, and MPEG7 as well as the format transformation between XML binding and RDF binding which will improve the interoperability with other systems.

4. DESIGN ISSUES

4.1 The Implementation Infrastructure - Web Services and J2EE

Web Services provides a standard means of communication among different software applications running on a variety of platforms and/or frameworks. The excitement over Web Services is largely based on the potential for a combination of XML, Web, SOAP and WSDL specifications, and yet-to-be-defined protocol stacks so as to address many of the problems that these technologies have encountered. Web Services is designed as a standard reference architecture which would promote interoperability and extensibility among web applications as well as allow them to be combined in order to perform more complex operations.

We chose Java 2 Enterprise Edition (J2EE) as our technical infrastructure based on the following considerations: 1) J2EE is an industry standard; we have a much wider choice of vendor for our pre-built software (application servers mostly) including numerous open source projects, 2) developers can leverage existing J2EE technologies to build a complete and fully interoperable Web Services that complies with XML standards and many benefits of J2EE are preserved for Web services such as Portability, Scalability, and Reliability, and 3) our other systems are all java based; no architectural change is needed when we implement web services on J2EE in our new tool.

4.2 RDF Query Language

RDF-enabled search technologies have the potential to provide significant improvements over the current keyword-based or theme navigation search engines. RDF, with the help of RDF Schema Language, provides a rich infrastructure for representing meaningful information in the form of ontologies. There are a number of query languages for RDF available some have been in use for several years.

Magkanaraki et al. [11] give a comprehensive comparison of RDF query languages based on five criterions: Modeling Constructs, Ontology Querying, Data Querying, Data/Ontology Querying and Additional Features. Compared to other query languages, RQL has best supports for Ontology Querying such as descendant traversal of class, property hierarchies, filtering conditions on class and on Data Querying such as Boolean Filters, Set-based operations, etc. It also has better supports for Data Ontology Querying and Additional Feature than other querying languages. It is the only one to support RDFS querying in existing querying languages. Based on the advanced features of RQL, we decided to use it as the query language of our MMLOM^3.

4.3 Querying and Storing Platform

The main components of RDF-based metadata management in MMLOM[3] are RDF querying and storing. Since there are already some RDF storing and querying tools and since some of these tools provide API for building applications upon them, it is not necessary to build these functions from the very beginning. Well established tools include RDFSuite [11], Sesame [1], Extensible Open RDF (EOR), Redland [11] and Jena [11]. Considering the objective of building an interoperable multimedia learning objects metadata management and mapping tool, our RDF querying and storing platform was chosen from those candidates based on the following criterions:

- A distributed architecture implemented using java-based technology,
- Easy integration into our existing Ubilearn system; a set of APIs to support RDF storing, querying, and management is provided
- Web Services are supported
- RQL as query language is used and database independency is supported

Sesame is the best selection as our RDF storing and querying server based on the above analysis. As an RDF Schema-based Repository and querying facility, it is being developed by Administrator Nederland Bv as one of the key deliverables in the European IST project On-To-Knowledge (EU-IST-1999- 10132). Sesame consists of a repository, a query engine and an administration module for adding and deleting RDF data and Schema information. It is also a Java system for RDF and RDF schema that provides data and schema query via SeRQL, RQL and RDQL, a backend database (PostgreSQL, MySQL, Oracle 9i), a RDF Semantics inferencer, a RDF explorer web application, and HTTP/SOAP/JMI access. These characteristics of Sesame meet all our requirements described above for the storing and querying platform, upon which our multimedia learning object metadata engine can be easily build.

4.4 Metadata Scheme Mapping

XSLT (XSL Transformations) [3] is a language for transforming XML documents into other XML documents. A transformation expressed in XSLT describes rules for transforming a source tree into a result tree. This transformation is achieved by associating patterns with templates: A pattern is matched against elements in the source tree; a template is instantiated to create part of the result tree, the result tree is separated from the source tree. The structure of the result tree can be completely different from the structure of the source tree. In constructing the result tree, elements from the source tree can be filtered and reordered, and arbitrary structures can be added.

In [15] a method was used to extract mappings from a pre-defined multiple-domain mapping matrix. This approach involves linking a mapping matrix to the XSLT processor. The mapping matrix explicitly defines the semantic mappings between a pair of metadata vocabularies belonging to different domains. This approach has, however, certain debilitating limitations. A matrix is only capable of specifying simple one-to-one mappings. A 2-dimensional matrix only works if the mappings are symmetrical in both directions across all of the domains. If the mappings are asymmetrical, the matrix becomes highly complex and multi-dimensional. We observe that the primary limitation of this approach is its lack of scalability; as the number of domains grows, the mappings become asymmetrical and the matrix becomes excessively complex and unwieldy.

In order to conquer this weakness and enable the coexistence of metadata interoperability, extensibility and diversity, a more generic approach is integrated in MMLOM[3]. Web Metadata Architecture [6, 9], proposed by Jane Hunter extracts mappings dynamically from a thesaurus of metadata terms, which is generated by formally defining relationships between a fixed set of standardized vocabularies chosen from different domains, such as LOM, DC and MPEG7. The semantic relationships are called MetaNet expressed using DAML+OIL [4]. By linking the semantic knowledge provided by MetaNet and XSLT together, semantic, structural and syntactic mappings between XML-encoded metadata descriptions from different domains can be performed.

As illustrated in Figure 4, the key components in the metadata mapping architecture of MMLOM[3], are described as follows:

- Domain-specific namespaces express each domain's metadata model and vocabularies using RDF Schema and coded by XML and XMLS.
- A metadata ontology (MetaNet) expressed using DAML+OIL [4] and based on a common underlying extensible vocabulary which can be generated by merging the domain-specific ontologies (RDF Schemas) from each namespace.
- XSLT; the language of transformation between XML-encoded metadata descriptions, combined with the semantic knowledge of the ontology, is capable of flexible and dynamic mappings between application profile instantiations.
- Application Profiles, which are XML Schema definitions that combine, restrict, extend and redefine elements from multiple existing namespaces, could also embed RDF Schema definitions of new classes or properties derived from classes and properties defined in the domain – specific RDF Schemas.

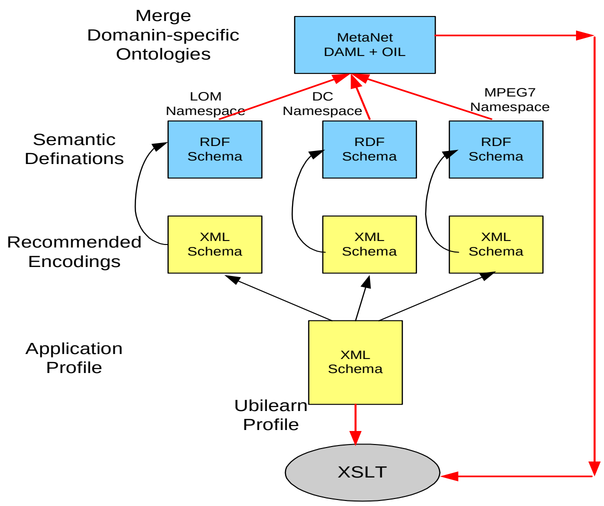

Figure 4. The metadata mapping architecture of MMLOM³

4.5 Reuse the Components of Current Metadata Mapping Tool

The MMLOM³ presents a subsystem in Ubilearn. It is an update of Ubilearn's current metadata mapping tool. There is a detail introduction of the tool used in Ubilearn [15]. The existing system is a metadata generator and search tool, which is able to store metadata in a relational database as well as to generate XML files and allows the mapping between the following metadata standards DC, LOM and ARIADNE in a transparent way. This time we will add RDF binding of DC, LOM and MPEG7 to support more semantic description of the metadata and add RDF-based query to those metadata, change the mapping methods with a more dynamic mapping ones that are different from the matrix mapping used in the current tool. At the same time we will add RDF to XML metadata format transformation to support the interoperability with other RDF-based tools and XML-based metadata tools. The new tool has the same architecture as the old one, reuses some of its components and can be integrated to Ubilearn seamlessly.

5. THE ARCHITECTURE AND COMPONENTS OF MMLOM³

The architecture of our MMLOM³ tool, as shown in Fig. 5, describes the organization of components in a J2EE framework, which includes client tier, web tier, business tier and information tier. Each component in MMLOM³ corresponds to a certain function in our function model described in Section 3, and pertains to a certain tier in our MMLOM³ architecture. The client tier displays information and collects input from the end user. The corresponding components in this tier include metadata authoring (DC, LOM and MPEG7) and mapping interface between these schemes. The web tier generates presentation logic and accepts user input from clients, and generates appropriate responses for the user. The technology used to implement the components in this tier is Java Servlet, which can also connect to business partners using web services technologies (SOAP, UDDI, WSDL, ebXML) through the Java APIs for XML (the JAX APIs). The business tier implements both presentation and business logic based on Enterprise JavaBeans technology. In our MMLOM³, the business tier it plays as metadata engine including components of RDF/XML based metadata generation, Mapping, and RDF and XML transformation. As the lowest tier, Sesame server provides the basic RDF storing and querying functions. The metadata engine accesses and stores metadata (in RDF format) from/into the Sesame Server through SOAP or HTTP protocol. It can access metadata (in XML format) and learning object directly from the database through JDBC.

5.1 Metadata Engine

- **Metadata authoring and mapping:** It provides an abstract interface for metadata editor to author the different metadata schemes like LOM, DC and MPEG7. It also provides the mapping interface between the schemes.
- **XML generator and mapping:** It first extracts the instance of metadata through the interface, and then generates its corresponding XML binding. It also maps XML binding of one of scheme into another scheme, for example, LOM to DC, DC to MPEG7, if required.
- **RDF generator and mapping:** Its function is very similar to that of XML generator and mapping, except that the generation and mapping is based on RDF-binding instead of XML-binding.
- **RDF and XML transformation:** it provides the function of transforming between XML binding and RDF binding for any given metadata scheme.
- **Metadata Search and browsing interface:** It generates query for every metadata scheme or the general query to more than one different schemes and then returns the results to the user.

- **Search engine:** It evaluates the query from users or from external web services-based systems and generates the corresponding RQL query to the sesame server. When the results return from the sesame server, the search engine represents them in a user friendly format at the user interface.

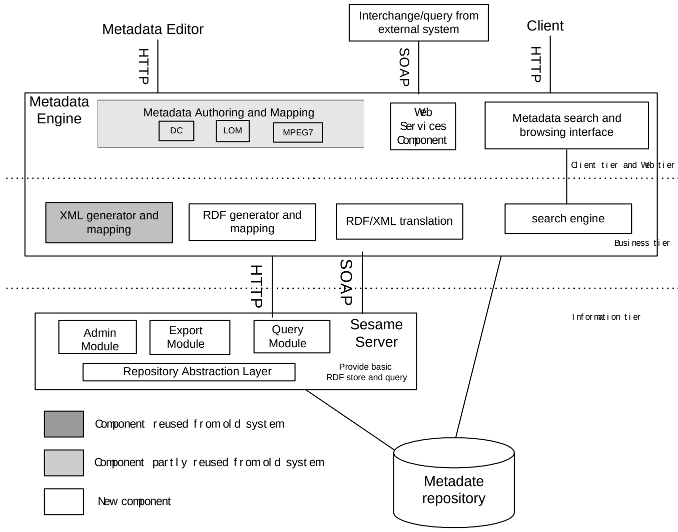

Figure 5. The architecture of MMLOM3

5.2 Sesame Server

- **Admin module** which allows incremental uploading of RDF data and schema information, as well as information deleting.
- **Query module** which evaluates RQL queries posed by the user.
- **Export module** which allows the extraction of the complete schema and/or data from a model in RDF format.
- **Repository Abstraction Level (RAL)** is an interface that offers RDF-specific methods to its clients and translates these methods in order to call its specific DBMS. This separation makes it possible to implement Sesame over a wide variety of repositories without changing any of Sesame's components. [1]

5.3 Metadata Repository

For each metadata instance two binding formats, XML and RDF are stored in the repository in our system. Any DBMS could be chosen as our metadata repository since the RAL component in Sesame server guarantees that high-level applications are independent from the database. Instead of MySQL, PostgreSQL was used as our database system based on the consideration that it now supports internazionalization. New tables are created and added to the database whenever a new class or property is required to be added into the repository. If a class is a sub-class of another class, the table created for it will also be a sub-table of the table corresponding to the super-class. Likewise for properties being sub-properties of other properties. Instances of classes and properties are inserted as values into the appropriate tables [1].

6. SUMMARY AND FUTURE WORK

The main design goal of this paper was to build an RDF-based metadata management system to support interoperability in the e-learning domain using up-to-date RDF technology. We presented the functional model of $MMLOM^3$, a new multimedia learning objects metadata management and mapping tool in the system. Based on the analysis of the key design issues of $MMLOM^3$, we proposed a detailed description of the architecture and address each component of this tool.

A major challenge, and a topic of continued research is to extend our model to build a semantic multimedia learning system, which may be achieved by building a semantic layer upon learning objects metadata to integrate domains, user profiles, and device profiles. It allows catalogue management of learning resources (IMS packaging, SCORM), learning object assembling, sequencing, and dynamic delivery and personalized learning.

7. ACKNOWLEDGMENTS

The authors acknowledge the financial assistance of the Natural Sciences and Engineering Research Council Canada (NSERC).

8. REFERENCES

[1] Jeen Broekstra, Arjohn Kampman, Sesame: A generic Architecture for Storing and Querying RDF and RDF Schema. ISWC02, 2002.

[2] Resource Description Framework (RDF) Model and Syntax Specification, W3C Recommendation Feb.1999

[3] XSL Transformations (XSLT) Version 1.0 Specification W3C Recommendation Nov. 1999

[4] DAML+OIL (March 2001) Reference Description W3C Note Dec. 2001

[5] Alimilia Magkanaraki, Sofia Alexaki, Vassilis Christophides, Benchmarking RDF Schemas for the Semantic Web. ICS-FORTH June 2002

[6] Jane Hunter, Adding Multimedia to the Semantic Web – Building an MPEG – 7 Ontology. SWWS, Stanford, 2001.

[7] The EducaNext Portal for learning resource technical white paper (http://www.educanext.org/ubp)

[8] Wolfgang Nejdl, Boris Wolf, Changtao Qu. The Edutella, A P2P Networking Infrastructure Based on RDF. WWW2002.

[9] Jane Hunter, Meta net - A Metadata term thesaurus to enable semantic interoperability between metadata domain, [9] Journal of Digital information, Volume 1 issue 8, 2001 –02-08

[10] Bernd Simon, Learning Object Brokerage: How to make it happen, Ed-media 2003

[11] Aimilia Magkanaraki, Grigoris Karvounarakis, Ontology Storage and querying, technical report of FORTH Institute of Computer Science Information Systems Laboratory No 308 April 2002

[12] The International Committee for Documentation of the International Council of Museums (ICOM-CIDOC)

[13] Tim Berners-Lee, James Hendler and Ora Lassila. The Semantic Web. Scientific American. 2001 May.

[14] Xiaofei Liu, Abdulmotaleb El Saddik, and Nicolas Georganas. An implementable architecture of e-learning system. CCECE 2003. Montreal, Quebec, Canada, May 4-7, 2003.

[15] Zheng Yin and Abdulmotaleb El Saddik. Study of Metadata for Advanced Multimedia Learning Objects. CCECE 2003. Montreal, Quebec, Canada, May 4-7, 2003.

[16] Mikael Nilsson. The semantic web: How RDF will change learning technology standards. www.cetis.ac.uk/content/211010927172953/viewArticle

[17] Marja-Riitta Koivunen, W3C Semantic Web Activity, presented at Semantic Web Kick-off in Finland meeting, Nov 2, 2001, Helsinki, Finland

[18] Representing vCard Objects in RDF/XML http://www.w3.org/TR/vcard-rdf

[19] IEEE Learning Object Metadata RDF binding. http://kmr.nada.kth.se/el/ims/md-lomrdf.html

[20] Schwartz, M., and Task Force on Bias-Free Language. Guidelines for Bias-Free Writing. Indiana University Press, Bloomington, IN, 1995