Creating a LO metadata profile for Distance Learning: An ontological approach

George Nikolopoulos[1], Aikaterini Kalou[1], Christos Pierrakeas[1], Achilles Kameas[1]

{nikolopoulos, kalou, pierrakeas, kameas}@eap.gr
[1] Educational Content, Methodology
and Technology Laboratory (e-CoMeT Lab)
Hellenic Open University (HOU)
Patras, Greece

Abstract. The importance of Learning Objects (LOs) in the learning process - especially in the case of distance education - has been underlined significantly in the literature. The ability of administrating LOs in terms of accessibility, re-usability and interoperability seems to be ensured by adopting an appropriate metadata schema which fully and adequately describes them. Several metadata standards have been developed such as DC (Dublin Core) and IEEE-LOM. In the context of our work, we explore them and conclude that none of these standards does completely meet the requirements of distance learning material. Therefore, we propose a new metadata schema that is actually an application profile of the widely adopted IEEE LOM and has special orientation in distance education. We also enrich this subset with some additional attributes that represent concepts like learning outcomes. Then, we create an ontological representation of this new educational schema with a view to improving the potential of LOs' discovery and retrieval within an intelligent e-learning system.

1 Introduction

Metadata are structured information used to describe the features of a resource (digital or not), thus making easier its management and retrieval. According to the definition in [2], metadata are "machine-readable information about electronic resources or other things". A set of metadata elements combined so as to serve a specific purpose, constitute a metadata schema. The great importance of metadata lies in the fact that the "meta"-information they convey is machine readable, therefore interoperability among applications can be achieved.

In the case of educational recourses, the set of metadata used to describe their characteristics, needs to be able to capture their educational and pedagogical aspects. Therefore, apart from *author*, *title* or *type* – fields that are common in all metadata schemas - an educational metadata schema should also include information regarding the resource's particular learning type, its intended end users, the educational context and many more.

Learning objects (LOs) are a kind of educational resources and constitute a novel approach in organizing the educational material. They have been widely used for the creation of web educational content by many modern e-learning systems, such as Learning Management Systems or Learning Content Management Systems [13]. What we lack though, is a metadata schema for the proper characterization of learning objects, and especially for learning objects that are designed to serve the scope of a distance learning course. Existing metadata schemas, as described in literature, are not adequate enough to express all aspects of distance education, and the necessity for a new metadata schema arises.

Having realized this need, we propose a metadata schema with special orientation in education and particularly in the characterization of distance learning material. After reviewing existing approaches for describing educational resources, in section 2, our educational metadata schema is extensively presented in section 3. Its ontology transformation is given in section 4, where the need to adopt such a representation technique for a metadata schema is also outlined. Conclusions follow, in section 5.

2 Existing Metadata for LOs

Generally speaking, metadata standards are defined as schemas, developed by an organization or institution so as to cover their needs to the best possible extent. The use of a single metadata standard is not a recommended solution, since each application has its special features. Alternatively, the use of application profiles is suggested. According to [7], an application profile is *"an aggregation of metadata elements selected from one or more different schemas and combined into a new compound schema"*.

The IEEE Learning Technology Standards Committee (LTSC) has developed a standard for the description of learning material and learning resources, known as IEEE Learning Object Metadata (IEEE LOM) [8]. LOM is without doubt a widespread standard for educational metadata and focuses mainly on the description of LOs. It includes more than 60 elements classified into 9 categories, each one of them containing metadata for various aspects of a LO, including its technical characteristics and rights, as well as educational and instructional features.

On the other hand, the Dublin Core Metadata Initiative was developed by organizations so as to aid the sharing of web resources and has no particular focus in education. Its initial schema, the Dublin Core Metadata Element Set (DCMES) [3] known as Dublin Core, consists of 15 elements. A second version, the Qualified Dublin Core (QDC) [4], comes to extend the previous schema, by importing 7 new elements. At the same time, QDC includes a group of qualifiers specifying the semantics of the elements in such a way so that they may be useful in resource discovery.

The ARIADNE[1] profile intended to describe learning material used in secondary and post-secondary education in order to solve two major problems: a) the indexing of LOs (i.e., the creation of the metadata by persons) and b) the exploitation of metadata

[1] http://www.ariadne.ac.uk/

by users looking for relevant pedagogical material (which should be as easy and efficient as possible). The current version of ARIADNE is an IEEE LOM profile and is fully compatible with the LOM specification.

Moreover, the IMS LRM[2] is a set of specifications for learning resources, developed by the global organization Instructional Management System (IMS). It includes elements useful for the description of learning resources, while the specifications address issues like content packaging, question and test interoperability, learning design and simple sequencing. IMS LRM adopts all the categories and elements of the LOM standard.

Two more LOM application profiles, that were created so as to describe resources locally, are CanCore[3] (Canadian Core) and UK LOM Core[4]. CanCore, used mainly in Canada, simplifies LOM maximizing at the same time interoperability between different projects. UK LOM Core, designed for United Kingdom educational system, intends to provide guidelines to those who desire to create, use and apply metadata.

GEM[5] (Gateway to Educational Materials) is a RDF metadata vocabulary, designed for the description of educational resources. It includes all DCMI elements with new properties focused on education. GEM has also created controlled vocabularies including catalogs for the level of end users, evaluation methods and tools and the types of resources. GEM Consortium has access to GemCat, a tool which generates metadata in a format consistent with the GEM standard.

Finally, the Sharable Content Object Reference Model[6] (SCORM) is a reference model which controls how the learning content is organized, described and linked with Learning Management Systems. The CAM (Content Aggregation Model) is the one of the three specifications that handles and adopts the IEEE LOM schema. However, SCORM allows the extension of LOM, thus enabling organizations to add new elements and enhance the existing controlled vocabularies.

3 Educational Metadata Schema for the Hellenic Open University

There is not one metadata standard appropriate to fulfill the requirements and needs of every application. Some standards focus on technical metadata, other on educational metadata while some other on more specialized elements. When an institution or organization needs a standard in order to characterize, retrieve or archive its resources, it uses an existing schema enriched with other elements or creates a new one.

To build our schema, we took into account several criteria and the needs of an institution, such as the Hellenic Open University which is specialized in lifelong and distance education. An eligible, flexible and functional schema is required so as to characterize a large amount of learning material. Moreover, the metadata elements to be integrated into the proposed schema, should meet the requirements of other struc-

[2] http://www.imsglobal.org/metadata/
[3] http://cancore.athabascau.ca/en/
[4] http://zope.cetis.ac.uk/profiles/uklomcore
[5] http://www.learningcommons.org/educators/library/gem.php
[6] http://www.adlnet.gov/capabilities/scorm/scorm-2004-4th

tures that manage LOs, like institutional libraries. So, since institutional libraries are usually based on a cataloging standard, it is necessary for our proposed schema to describe, at least, some common characteristics like title, description, format and creation date of a LO.

Our proposed schema is actually an application profile of IEEE LOM but with a particular orientation in distance learning material. The IEEE LOM was chosen as a basis for creating the profile due to its wide acceptance in the academic environment and its extensive usage by institutional repositories. The profile adopts the majority of IEEE LOM's elements, augmenting them with some additional attributes in order to represent concepts commonly used in distance education, like learning outcomes. The proposed schema is rich enough, so that it can effectively describe all aspects of a LO (educational, technical, etc.), but not exceedingly analytic as to become difficult to use.

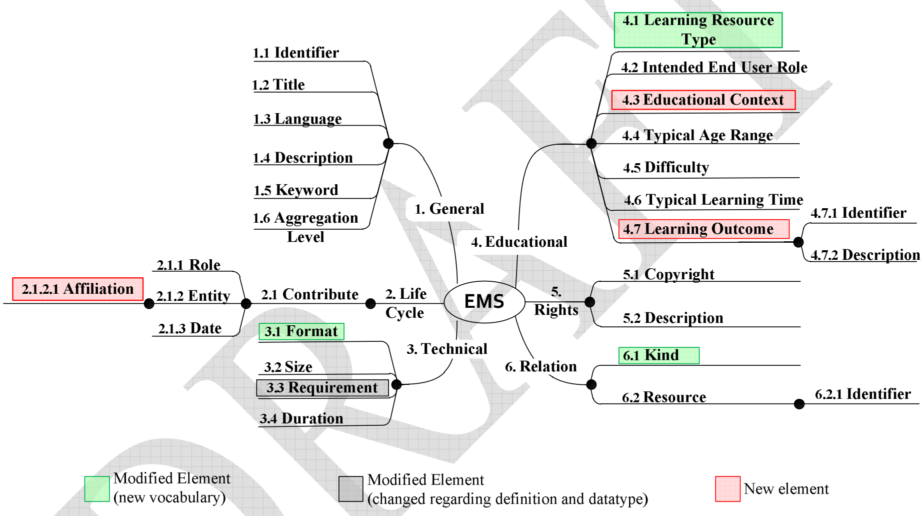

Fig. 1. The proposed Educational Metadata Schema (EMS).

Fig. 1 summarizes the elements of which the proposed metadata schema consists. Apart from those that have been directly taken from IEEE LOM (those not marked with a specific color in Fig. 1) and are considered to come with equivalent semantics, three additional types have been included. The first type concerns elements which have differences in the value space of their controlled vocabulary compared to their counterparts in the IEEE LOM schema. More specifically, for those elements we have modified the predefined set of values that they can accept in an attempt to meet a LO's specific characteristics. The second type includes elements that come with modifications in their definition and data type, while the third one regards new entries in the proposed schema.

The elements of our proposed metadata schema that come with modifications in the value space of their controlled vocabulary are three in total: 1) *Format* of the

Technical category, 2) the *Learning Resource Type* of the *Educational* category and 3) *Kind* of the *Relation* category.

As far as the *Format* element is concerned, we have defined a new set of allowable values which appears in Table 1. The given set is based on IANA MIME types[7], given in IEEE LOM, but is oriented for the characterization of distance learning material and particularly for LOs that are going to be utilized by an adaptive learning system. The set in Table 1 emerged after studying the characteristics of the educational material that is already used by the Hellenic Open University (HOU) and seems to be quite broad in order to cover a wide range of technical data types.

Table 1. Possible values of the *Format* element (i.e. technical data types of the resource).

Proposed Value Space			
Text	document hypertext	Streaming Media	audio recording animation self-running presentation webcast video
Image	Photo Map Graph Image presentation	Application	interactive software hypermedia application wiki presentation

Learning Resource Type is another element with substantial modifications regarding its controlled vocabulary compared to the one in the IEEE LOM schema. The main problem with the corresponding IEEE LOM element is that apart from the values which express educational information (e.g. *Exercise, Problem Statement, Simulation*) it also contains values that refer to technical information (e.g. *Diagram, Figure, Graph*). Moreover, some important types of LOs such as *Example*, *Serious Game*, *Case Study* or *Project* are absent.

To this end, we define a completely new list of acceptable values that reflects the most common types of educational material used within distance education courses and incorporates only information about the instructional perspective of a resource. This list of values is, to a certain degree, based on the content object types provided by the ALOCOM generic content model[8]. The complete list of the learning types we propose is presented in Table 2.

Another important modification concerns the *Kind* element that belongs to the *Relation* category. In this element, the controlled vocabulary, used as its value space, expresses the various kinds of relationships among LOs. So, apart from the existing

[7]http://www.iana.org/assignments/media-types/index.html

[8]http://kuleuven.academia.edu/ErikDuval/Papers/1227319/ALOCOM_a_generic_content_model_for_learning_objects

relationship *has part* and its inverse one *is part of* we make provision for two additional types of relationships:

1) *supports* and its inverse *is supported by* attempts to correlate a "supportive" LO that contains complementary or prerequisite knowledge, with one that has a key role in the learning process (a "core" LO).
2) *is alternative type* relationship correlates two or more LOs that are exactly the same in their educational content and differ only in their technical format. This is a highly significant relationship especially if the objective is to provide personalized learning depending on the preferences of each learner.

Table 2. Possible values of the *Learning Resource Type* element.

Proposed Value Space		
1. Guidelines	**9. Activity**	**11. Self-Assessment**
2. Presentation	− Case Study	− Multiple Choice Questions
3. Demonstration	− Problem Solving	− Open Type Question
4. Lecture	− Text Composition	− Problem Statement
5. Definition-Principle-Law	− Question	**12. Experiment**
	10. Simulation	**13. Serious Game**
6. Narrative Text	− Interactive	**14. Exercise**
7. Analogy	− Non Interactive	− Multiple Choice Questions
8. Example		− Open Type Question
		− Problem Statement
		15. Project

The only element that has been adopted with different definition and data type is *Requirement*. In the context of our metadata schema, the *Requirement* element is used to describe any software or hardware requirements which are necessary in order to use a LO. Its data type has been altered to *LangString*. Thus, its completion simply requires writing statements like *"This LO requires the use of Adobe Acrobat Reader. Version newer than 6.xx*". On the other hand, the original IEEE LOM element needs much more, so this modification simplifies the process of describing the requirements of a LO to a great extent.

Finally, we have augmented our schema with some new elements that were necessary in order to represent concepts commonly used in distance education. We added the sub-element *Affiliation*, as shown in Fig. 1, which determines the status of the entity that has contributed to the creation and development of the LO. The *Learning Outcome* element, placed under the *Educational* category, expresses the correlation of a LO with one or more learning outcomes. In particular, for each learning outcome that a LO satisfies, one needs to give its natural language statement, via the *Description* element, as well as to assign to it an identifier (*Identifier:Entry* and *Identifier:Catalog*), according to a specific identification system. The *Educational Context* element now implies the actual context where the learning process takes place, and can accept values like "distance education", "face to face learning" and "blended learning". This is a key element in terms that it gives information about the mode of learning for which the particular object is appropriate.

4 An ontological approach

In what follows, we first outline the reason for using an ontology for representing our proposed metadata schema. We then explain in details the structure of the deployed ontology.

4.1 Why we need an ontology

The elements of any educational metadata schema should be managed in any available format e.g. SQL tables, text files, HTML meta-tags and so on [11]. Such a technical realization of the abstract model in a specific format is called a *binding*. As it is known, for IEEE LOM, XML and RDF bindings are defined. The usage of XML for the LO metadata expression facilitates the indexing process and the retrieval of annotated learning resources. However, this format seems to be not sufficient enough to address the limitation of text-based searching, since XML does not provide the meaning of the described structures. The Resource Description Framework[9] (RDF) attempts to overcome the problem by adding semantics to each metadata element of any schema. The description of LOM elements via RDF facilitates their integration into e-learning systems which nowadays are dominated by Semantic Web technologies and especially by the notion of ontology.

Even though RDF is intended for representing knowledge, it lacks reasoning abilities; RDF does not support making inferences or deductions. Therefore, a much more expressive framework is required, so that metadata can be meaningfully encoded. Ontologies, expressed in OWL (the most widely-used ontology language) are the pillar of Semantic Web and provide the ability to represent any domain of interest in a more structured way. The ontological expression of LO metadata converts them into machine-understandable information. Moreover, metadata that are represented by ontological models in OWL, are enhanced with richer properties.

In general, a survey of the literature on the usage of ontologies in web-based education returns a great number of systems that embed ontological models in their implementation [1]. These ontological models can reflect many various aspects of an e-learning system, such as student profiles and knowledge domains. The integration of LOs components to such systems requires a more semantically enriched representation [5]. Therefore, many research groups have attempted to annotate semantically the LO metadata. Some representative examples are ALOCOM, SCORM, OntoLo, EduBank, ACM CCS, etc.

[9] http://www.w3.org/TR/rdf-primer/

4.2 The LO ontology

In this subsection, we describe the deployed ontology that represents our proposed application profile. In order to build this ontology, we followed a widely-adopted methodology, proposed in [12]. For its formal representation, we used OWL 2 - the most recent version of the Web Ontology Language - whereas for creating and managing our ontology, we used the Protégé editor.

The notion of LO is reflected in the *LearningObject* class. The elements of the *General* category are represented by the corresponding datatype properties, such as *title*, *language*, *description*, and *aggregation level*. In particular, the *aggregation level* is an integer datatype property that can take values from 1 to 4. The *language* property can be filled with any of the known language identifiers, like *"en"* for English or *"el"* for Greek.

The *Contribute* element of the *Life Cycle* category is captured by the *Contributor* class and the corresponding elements are represented by datatype properties, such as *contributeDate*, *affiliation* and *contributorRole*. *ContributorRole* is an enumerated dataype property that can be filled with the values *publisher*, *creator*, *reviewer* or any other similar concepts, expressing allowable roles in the life cycle of a LO. Instances of the class *Contributor* are linked to any instance of the class *LearningObject* via the object property *contributor*.

The *Format* category consists of the following four elements: *Text, Image*, *Streaming Media* and *Application*. Each of them is captured as a subclass of the *Format* class. This hierarchical structure is shown in Fig. 2. The controlled vocabulary, expressing allowable values for each element in the *Format* category, are represented as instances of the corresponding classes (see Table 1 – the second column involves the instances). The remaining elements in this category, expressing the physical size, software or hardware requirements or time duration for streaming media, are described by the datatype properties *size*, *requirement* and *duration*, respectively.

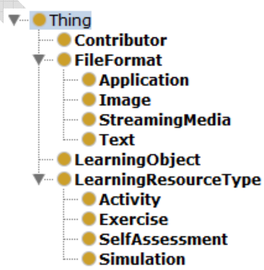

Fig. 2. The class hierarchy of the ontology.

The *Learning Resource Type* element that specifies the different educational types of LOs is captured by the *LearningResourceType* class. This element is associated with a predefined list of terms, represented as instances in the ontology. However, these terms can be further refined and hence they have been placed as sub-classes of the main *LearningResourceType* class. The ontological structure of the *Learning Resource Type* element is indicated in Fig. 3.

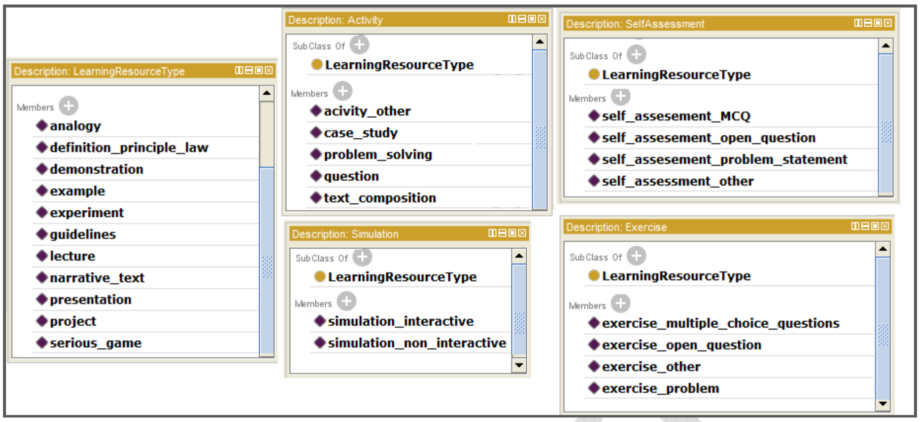

Fig. 3. The *Learning Resource Type* category in the ontology.

The remaining elements of the proposed metadata schema, belonging to the *Education* category and expressing concepts like the intended end users, the instructional context, the age range of end users, as well as difficulty and average learning time, are captured by the corresponding datatype properties.

The datatype property *rights* has been defined in order to capture the copyright data that apply for a LO. Finally, the potential relationships among LOs, appearing as members of the *Relation* category, are represented as object properties in our ontology linking two instances of the *LearningObject* class. These object properties have been placed as sub-properties of the main property *isRelatedtoMA*, as shown in Fig. 4.

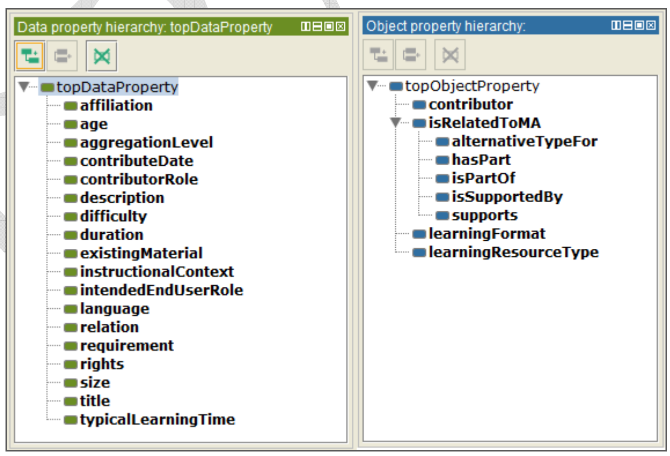

Fig. 4. Object and Datatype Properties of the developed ontology.

There are also some additional elements coming from various categories of the metadata profile that could be represented in the ontology, only after importing some other supporting ontologies. Such ontologies are the *LearningOutcome ontology* presented in [9], the *Java ontology* described in [10] and the well-known FOAF ontology[10]. The Java ontology could be replaced by any other ontology that captures a knowledge domain which we want to teach.

The element *keyword* of the *General* category can be expressed via the object property *subject* that relates instances of the *LearningObject* class to instances of classes included in an ontology that models a particular knowledge domain. As far as the *Life Cycle* category is concerned, we have defined the *LearningObject* class as equivalent to the *Person* class of the FOAF ontology and then we make use the built-in object properties *foaf:name* and *foaf:surname* so as to formulate the element *Contribute:Entity*.

The main objective of the combination of *LearningObject* and *LearningOutcome* ontologies is to assign learning outcomes to LOs. To this end, we determined the object properties *satisfies* and *satisfiesInd*, both of which associate instances of *LearningObject* class with *LearningOutcome* instances, directly and indirectly, respectively. The latter is going to be used in implicit inferences. The indirect correlation of LOs can be only elaborated by restriction rules and by exploiting existing reasoning mechanisms. As a consequence, we have defined the following rule, expressed in the Semantic Web Rule Language (SWRL): "*IF (x is-a LearningObject) AND (y is-a LearningObject) AND (p is-a LearningOutcome) AND (y satisfies p) AND (x complements y) THEN (x satisfiesInd p)*". With this rule, it can be inferred that a LO satisfies a learning outcome indirectly, if this LO is supportive for another one, created in order to serve the aforesaid learning outcome.

4.3 Evaluating some example queries

The Java ontology, presented in [10], models the knowledge domain of the Java programming language that is covered by the HOU's course module of 'Software Engineering'. In [9], except for the ontological representation of the structure and taxonomy of learning outcomes, we attempted to apply this model to a selected piece of educational material provided by the HOU that was relevant to the Java programming language. Therefore, in the current work, we continue with the design and the implementation of LOs that are going to relate with learning outcomes that are already constructed and concern concepts of the Java.

Moreover, we ran some representative queries and evaluated them against the populated ontology aiming at examine our model's capability to infer knowledge. These queries are expressed in the Manchester OWL Syntax and tested through the DL query tab of Protégé. We demonstrated that apart from running simple lookup queries, we can request more complex things, based on this ontology.

Consider, for example, that we want to obtain all LOs that are difficult and satisfy learning outcomes that concern Java operators and fall into Cognitive Domain. We can express this by the query#1 in Table 3. Moreover, with the query#2 we can retrieve those LOs that their learning resource type is narrative text. Finally, with query#3, we can retrieve all LOs that satisfy in anyway those learning outcomes that are relevant to Java operators.

[10] http://xmlns.com/foaf/spec/

Table 3. Some Example Queries in Manchester OWL Syntax.

#	*Query*
1	**satisfies** some (subject some **Operator** and domain value **CognitiveDomain)** and **difficulty** value **'difficult'**
2	**learningResourceType** value **narrative_text**
3	**satisfies** some (subject some **Operator**) and **satisfiesInd** some (subject some **Operator**)

Of course similar requests can be made for different domain subjects, different levels of knowledge and any kind of relationship modeled in the ontology as a property. All these semantic queries are actually examples of competency questions for the proposed ontology. Competency questions are a commonly used technique for evaluating such formalisms [6].

5 Conclusions

Having realized the need to effectively characterize the educational material that serves the needs of a course delivered from distance, we propose a metadata schema with educational perspective and particular orientation in distance learning. The aim of the proposed schema is to provide a complete element set for the characterization of LOs or of any other kind of educational material, intended to be exploited by e-learning systems and applications.

Our educational metadata schema is actually a profile of the widely known IEEE LOM. It adopts the majority of LOM's elements, it modifies some other or even augments it with additional elements representing concepts that are common in distance education. Learning outcomes are, for example, such concepts and any LO is constructed with the view to serve a specific learning outcome. Our schema is rich enough, so that it can effectively describe both educational and technical aspects of a LO, but not exceedingly analytic so as to become difficult in use.

As a next step, we tried to model our schema as an ontology. Ontologies come with many applications in the field of education and their usage for representing the structure of a LO could end up to the development of advanced, intelligent applications. The ontological representation of our educational metadata schema was accomplished by appropriately "translating" its structural elements to classes, properties and instances in the ontology.

The proposed work comes to address the drawbacks that come up by the usage of a cataloging standard and the deficiency of the current metadata schema in the Hellenic Open University. Furthermore, the expression of the presented schema in an ontology language makes more feasible the integration of the HOU's infrastructures into semantic-aware e-learning systems.

Our future work is focused on the development of an ontology-based e-learning system for the HOU that is going to offer personalized learning from distance. The resulted learning object metadata ontology can be easily combined with other ontologies, modeling other notions in distance education. The importance of such an onto-

logical representation lies in its possible exploitation by systems that can handle semantics and thus understand the role of each single piece of educational material. Furthermore, we are going to design tools and methods able to create reusable LOs in accordance with the proposed metadata schema. Our major goal is to provide HOU's students with advanced, personalized services for efficiently handling and disseminating educational material.
Acknowledgments. This research described in this paper has been co-financed by the European Union (European Social Fund – ESF) and Greek National Funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) (Funding Program: "HOU")

References

1. Al-Khalifa, H.S., Hugh, D.: The Evolution of Metadata from Standards to Semantics in E-Learning Applications. *Proceedings of Seventh Conf. Hypertext and Hypermedia (Hypertext '06),* (2006).
2. Berners-Lee, T. Web architecture : Metadata. Retrieved from World Wide Web Consortium (W3C): http://www.w3.org, (1997)
3. DCMI: Dublin Core Metadata Element Set, version 1.1. DCMI Recommendation (2008) URL: http://www.dublincore.org/documents/dces/.
4. DCMI: DCMI Metadata Terms. DCMI Recommendation (2008) URL: http://dublincore.org/documents/dcmi-terms/.
5. Gaević, D., Jovanović, J., Devedžić, V.: Ontologies for Reusing Learning Object Content. *In Proceedings of the 5th IEEE International Conference on Advanced Learning Technologies (3th International Workshop on Applications of Semantic Web Technologies for e-Learning)*, Kaohsiung, Taiwan , pp. 650-654, (2005).
6. Grüninger, M. and Fox, M. S.: Methodology for the design and evaluation of ontologies. In IJCAI95 Workshop on Basic Ontological Issues in Knowledge Sharing, (1995)
7. Heery, R., & Patel, M.: Application profiles : mixing and matching metadata schemas. *Ariadne* (25) (2000)
8. Hodgins, W., Duval, E.: Draft Standard for Learning Object Metadata. Institute of Electrical and Electronics Engineers, Inc. (2002) URL: http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf
9. Kalou, A., Solomou, G., Pierrakeas, C., Kameas, A.: An Ontology Model for Building, Classifying and Using Learning Outcomes. In International Conference on Advanced Learning Technologies, Rome, (2012)
10. Kouneli, A., Solomou, G., Pierrakeas, C., Kameas, A.: Modeling the Knowledge Domain of the Java Programming Language as an Ontology. In International Conference on Web-based Learning, Sinaia, Romania, (2012)
11. Nilsson, M., Palmér, M., Brase, J.: The LOM RDF binding - principles and implementation. 3rd Annual Ariadne Conference, Katholieke Universiteit Leuven, Belgium (2003)
12. Noy, N. and McGuiness, D.: "Ontology Development 101: A Guide to Creating Your First Ontology", Stanford Knowledge Systems Laboratory Technical Report KSL-01-05, (2001)
13. Schreurs, J. & Al-Zoubi, A.Y.: Converting Content to Reusable Learning Objects Adaptable to User Preferences and Infrastructure. In T. Bastiaens & S. Carliner (Eds.), *Proceedings of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education 2007* (pp. 6537-6544). Chesapeake, VA: AACE (2007)