Conceptual Interpretation of LOM Semantics and its Mapping to Upper Level Ontologies

M. Elena Rodríguez, Jordi Conesa
(Universitat Oberta de Catalunya, Barcelona, Spain
mrodriguezgo@uoc.edu, jconesac@uoc.edu)

Elena García-Barriocanal, Miguel Ángel Sicilia
(Universidad de Alcalá de Henares, Madrid, Spain
elena.garciab@uah.es, msicilia@uah.es)

Abstract: In this paper we discuss about LOM semantics as result of using ontologies that help to find possible improvements to LOM (extensions and resolution of semantic ambiguities). The paper also presents a mapping between the LOM standard and an upper level ontology that promotes semantic interoperability among heterogeneous learning systems.

Keywords: LOM standard, learning object, upper-level ontology, semantic interoperability
Categories: L.1.2, L1.3, D.2.12.

1 Introduction

The use of Information and Communication Technologies (ICT) in education has widened the feasibility of e-learning. Among others, e-learning needs to encompass both pedagogic and technological issues. Technological aspects include hardware and software architectures based on standards and specifications that make possible the interoperability, i.e., the search, access, reusing, sharing and interchange of learning resources (Learning Objects (LO)) and services between learning systems. Standards and specifications provide a common language. The more rigorous and formal the language is the better semantic interoperability levels we obtain.

In this paper we are interested in metadata standards that allow the description of learning resources. More specifically, our goal is to increase their semantic expressiveness by means of ontologies in order to improve searching capabilities and to facilitate the annotation of learning resources. Achieving this objective requires answering research questions as: "Can we use ontologies to validate and improve standard metadata definitions?" and if so, "Can we establish mappings between ontologies and a given standard definition to make the standard even more generalized?" Along the paper the reader will find the answers and results associated to the previous questions.

The paper is structured as follows: section 2 describes briefly the LOM standard, revises the LO concept and justifies the benefits to cope with LO categorizations. This section also presents an overview of different ontology kinds and the advantages and requirements that has to have an ontology in LO context, justifying why we use Opencyc as the support ontology. Section 3 describes our results of comparison

between LOM and Opencyc, presenting a LO categorization and its mapping in OpenCyc. Last section concludes the paper and presents the further work to be done.

2 Related Work

2.1 The IEEE LOM Standard, the Learning Object Concept and Types of Learning Objects

The LOM standard ([IEEE LOM 2002]) defines, grouped in categories, the metadata elements set to be used for describing LO. All metadata elements are optional and they are structured hierarchically. We can find generic metadata elements (general category), metadata that describe the history and current state of a LO (life cycle category), data elements over specified metadata elements (meta-metadata category), metadata that describe LO technical requirements (technical category), metadata elements stating pedagogical characteristics of the LO (educational category), metadata that dealt with legal issues (rights category), metadata that allow specifying relationships between LO (relation category), metadata for adding comments about LO use experiences (annotation category) and metadata that define the LO according to a classification system (classification category).

A LO, in terms of LOM standard, is defined as any entity, digital or non-digital, that may be used for learning, education or training. This broad definition is complemented and/or restricted by some authors in order to clarify the LO meaning and boundary. For example Wiley ([Wiley 2002]), restricts LO to be digital entities and emphasizes its potential to be reused in different learning experiences. On the other hand, Polsani ([Polsani 2003]), to the reusing requisite, adds the need to provide instructional context to LO.

Despite the efforts, a main problem when working with LOs, is they are fairly heterogeneous with regards to several characteristics. Moreover, current specifications and implementations do not include the notion of LO type, according to different criteria, as guiding method for metadata structuring ([Sicilia et al. 2004]). For example, all LOM standard metadata are equally applicable to any type of LO, irrespective of its type. In fact, attending to metadata elements, we can specialize LOs with regards to several characteristics as, for example, their internal structure (1.7 *structure* element), granularity (1.8 *aggregation level*), interactivity form (5.1 *interactivity type*) or kind of learning resource (5.2 *learning resource type*). Other interesting distinction between LOs can be made from a conceptual point of view. When we design and develop LOs we can distinguish, at least, between the LO understood as creative work, and the different representations we offer from this creative work, i.e. the final digital or physical content elements (i.e. the existing LOs).

Exploiting the LO type notion inside ontologies is key for different processes that can be (semi-)automated, given that LO types determine the reasoning processes that are applicable to each kind of LO. Some examples of these processes are: 1) tagging: the distinction between kinds of LOs simplifies their tagging process, given that some metadata are only applicable to a specific LO type. In addition, some other metadata can be automatically derived depending on the LO type. 2) Location: the kind of a LO can act a discriminator element in LO searches in large repositories. 3) Composition and sequencing: the LO kind can be used for LOs combination and for specifying

their use restrictions in a learning experience. And 3) personalization: types of LO constitute a way to model user preferences about kind of LO, interaction styles, etc.

2.2 Ontologies and their Role in E-learning

Ontologies may be seen as the representation of one view‘s world. As much as more consensus exists in ontologies, more uses they have. Ontologies may be classified in several ways ([Guarino 98]). Depending on the kind of knowledge that they represent, ontologies may be classified in application ontologies, which describe one or more domains in a very particular context. Obviously, this kind of ontologies cannot be reused out of the context they belong. Another more general kind of ontologies is domain and task ontologies. These ontologies deal with a given generic domain or task, respectively. They are more reusable than the previous ones, but the problem is that their reusability is limited according to the domains of interest, i.e., if we are interested in dealing with more than one domain, they may be incomplete. The third kind of ontologies is the upper-level ontologies, which represent general information about several domains and, sometimes, take into account different cultural perspectives. These ontologies are usually larger than the others and more reusable in general. However, to reuse them we need to overcome the usability problems of dealing with large ontologies. In our work we plan to use ontologies with two main purposes:

1. Checking and validating LOM standard, as well as finding possible improvements (extensions and resolution of semantic ambiguities).
2. Establishing a correspondence between the LOM metadata and the ontology. This will improve semantic interoperability of learning contents and services among different learning systems ([Aroyo et. al 2006], [Sicilia 2006]).

For the first purpose we need an ontology that describes the domain (or domains) that are dealt in the LOM standard. For the second objective we need an upper-level ontology because they provide a more reliable background (and top level) concepts, its knowledge is shared for more people and therefore it is more generalized. Since LOM deals not only with the educative data, but also with the different ways of tracking changes, composing objects, authorship etc. we cannot choose a single domain ontology to accomplish the first purpose. Hence, we use one of the most well

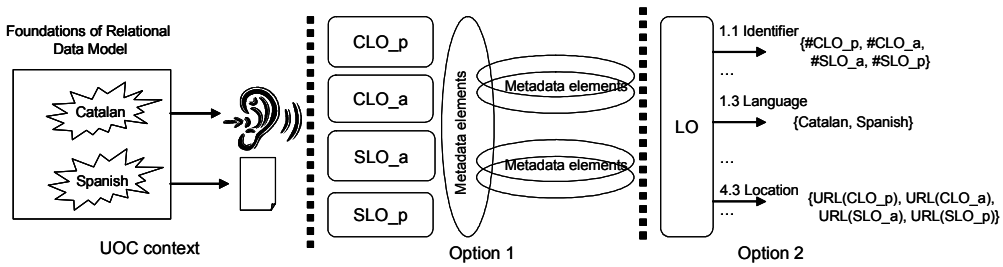

Figure 1: UOC motivating example

known and large ontologies available: the Cyc (encyclopedia) ontology ([Lenat 95]). Cyc is an ontology that contains with more than 2.2 million assertions (facts and rules) describing more than 250,000 terms, including 15,000 predicates.

Since Cyc is a commercial product, we decided to use OpenCyc (http://www.cyc.com/cyc/opencyc) for that work. Opencyc is a public version of Cyc and contains a subset of the Cyc knowledge. Even being a subtype of Cyc, we believe that OpenCyc contains enough knowledge to deal with our validation and mapping processes.

3 Common Sense Ontologies

3.1 Motivating Example: the UOC Case

As we have stated, LOM standard does not explicitly consider the possibility of specialize LOs. In addition, we find very specific metadata (implementation-dependent metadata) merged with very general metadata that are independent of a concrete final implementation of the LO

For example, the UOC is a virtual university that, among others, offers several degrees related with TIC knowledge field. Moreover, the UOC offers learning in two linguistic environments (Catalan and Spanish). Let consider a subject as "Introduction to Databases" where, among others topics, we explain the foundations of the relational data model. Imagine we have a LO that is a lesson in narrative text that explains the previous topic. This LO is available both in Catalan and Spanish languages. We also assume that, for each language, it is available in textual (pdf) and audio format (figure 1 shows the described situation). Assuming the UOC has selected metadata in an application profile ([Duval et al. 2002]), when we proceed to tag the LO at instance level, we have many different options. Without loss of generality, let examine two possible options (depicted also in figure 1).

In option 1, four LO metadata registers are created, one for each LO instance, according to available languages and formats. Each metadata register includes all

Option 1	Option 2
- Too many data redundancy; therefore hard data maintenance - All metadata are defined at the same level - There is not way to know that, for example, CLO_p and SLO_p are the same content with different languages even when relationship between them is defined ("version of" does not have the right semantic, "translation of" should be a more appropriate value - Changes in content imply new LO instances and then new metadata registries + Tagging fits LO definitions (each available instance is a LO, and each instance has its own metadata registry)	- Very unclear and confusing conceptual structure - All metadata are defined at the same level - Different semantic interpretations, for example: one LO that uses two languages or two LO, each one with different languages? - Adhoc heuristics are needed to interrelate metadata values, i.e., to know which is the location of the Spanish versions - It is not clear, for example, what to do when a change in the content involves only one of the languages - Tagging does not fit LO definitions (except LOM LO definition in a very broad sense interpretation "[…]as *any entity*, digital or *non-digital*, that may be used for learning […]." + No data redundancy

Figure 2: Disadvantages and advantages of each tagging option

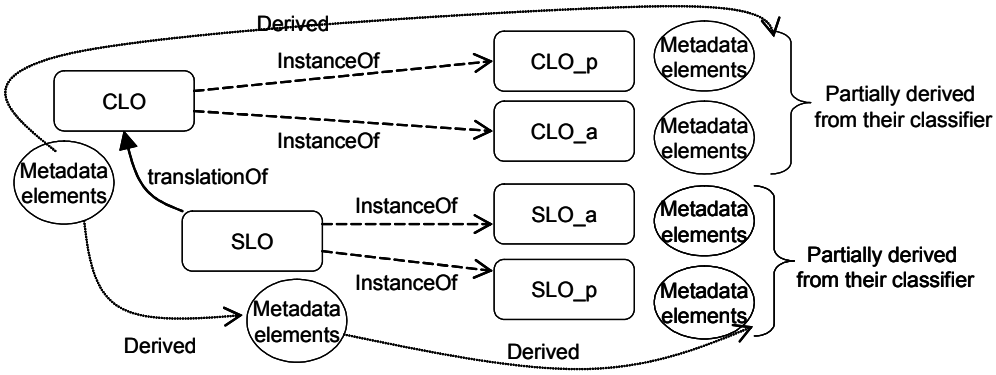

Figure 3: LO specialization and associations

metadata included in the UOC metadata profile. We want to point out the following issues: 1) some metadata values depend on the language of the LO instance under consideration (e.g. general metadata as 1.2 *title*, 1.3 *language*, 1.5 *keywords* etc.). 2) Some other metadata depend on the implementation format (as technical metadata: 4.1 *format*, 4.4 *requirements* etc.) of the LO instance. 3) Other metadata remain invariable, independently of previous aspects (1.6 *coverage*, 1.7 *structure*, 1.8 *aggregation level* most of educational metadata etc.). And 4) relationships among LO instances (not shown in the figure) can be declared though metadata contained in the relation category. LOM vocabulary values for the *kind* (metadata element 7.1) of the former relationships would be, respectively, "is version of" and "is format of".

In option 2, only one metadata register is provided for the different LO instances. This option takes advantage on the fact some metadata can be multivalued. For instance, metadata element 4.3 *location* (technical category) stores the URL where the LO instances are available, metadata element 1.3 *language* the human languages within the LO to communicate to the intended user etc.

Figure 2 summarizes the most relevant disadvantages (-) and advantages (+) of the described metadata tagging options.

3.2 An Interpretation of LOM semantics

In that subsection we present an interpretation of the LOM semantics which is the result of a deep comparison process we have done between LOM metadata and the part of Opencyc that deals with their semantic equivalent (or almost close) concepts.

Our strategy is driven by a LO specialization taking as criteria the nature, from a conceptual point of view, of LOs. At a first level, we can distinguish between the conceptual LOs (the LO consequence of a creative work) and existing LOs (the final accessible and available LO instances). In addition, we can identify, at least, a subset of conceptual LOs, i.e. the different conceptual LOs that we can derive as versions of the previous conceptual LO (i.e. the copy conceptual LOs). This specialization has been used to filter metadata according to their abstraction level, distinguishing which metadata are applicable to each LO type. In addition, applicable metadata are classified in basic and derived. Only basic metadata elements will potentially receive value(s). The value(s) of derived metadata could be inferred trough the specification

of the appropriate relationships. Non applicable metadata elements to a LO type do not appear either in the LO type or in its supertypes.

CLO represents the original textual LO conceived in Catalan (Conceptual LO). Only metadata elements related to the intellectual process of design and creation need to be defined only for CLO. Examples of these elements include the ones of the general category (e.g. 1.6 *coverage*, 1.7 *structure*, 1.8 *aggregation level*) as well as (among others), metadata belonging to educational and classification categories. Given that our LO is a textual LO, it is also required to specify language dependant metadata elements (like 1.2 *title*, 1.3 *language*, 1.5 *keywords*, etc.) which take value according to Catalan language.

The translation of CLO to Spanish gives rise to SLO (Copy Conceptual LO, subtype of Conceptual LO). Some of the SLO metadata will be derived from CLO, which is related with SLO by the "translation of" relationship type. Only the language dependent metadata are considered basic.

Each conceptual LO can be distributed in two different formats: audio and pdf. Therefore, four new LO representing the existing and accessible objects for SLO and CLO are created: CLO_p, CLO_a, SLO_p and SLO_a (created as instances of Existing LO and related with their conceptual LO by the "instance of" relationship type). This schema results in six occurrences of LO. CLO_a and CLO_p derive most of their metadata from their conceptual LO (CLO). Metadata to be considered basic are the implementation-dependent metadata as, for example, it would be the case of technical category metadata elements. The same happens to SLO_p and SLO_a regarding to SLO.

It is important to note that not only translations trigger new copies of a conceptual LO. For example, imagine that in "Introduction to Databases" students must practice SQL over a relational DBMS as PostgreSQL. In this case, we can also differentiate among the common characteristics associated to PostgreSQL (conceptual LO), the properties associated to specific PostgreSQL versions (copy conceptual LOs), and the available instances of PostgreSQL (existing LO). It is obvious that the Oracle DMBS will correspond to a different conceptual LO.

Our interpretation provides several advantages to the current ambiguity in LOM semantics:1) it provides a compressible guideline to annotate LOM. 2) The annotation process of LOM may be partially automated and therefore is less error prone. 3) All metadata are defined according to different conceptual levels. 4) No redundancy and easy data maintenance, thanks to a clear separation of concerns. 5) Some inferences are possible due to the semantic assumptions implicit in the model. And 6) the fact that the existing and the conceptual LO are split may improve search mechanisms.

The main disadvantage of our interpretation that we have to distribute the metadata through different LO, which implies more metadata register. Even though, this is a relative drawback because most of the metadata are derived. Therefore, we can say that the number of basic metadata is at least the same in our interpretation versus the others. In other cases, we have less basic metadata because other interpretations may tend to repeat metadata and do not allow metadata derivation.

3.3 Mapping in OpenCyc

This subsection presents a possible mapping of our semantic interpretation of LOM to Opencyc (see figure 4). We also sketch some suggested mappings between LOM

	IEEE LOM Interpretation	Opencyc
LO root concepts	Learning Object	Thing
	Conceptual Learning Object	Conceptual Work (=CW)
	Copy Conceptual Learning Object	Textual Specific Work (=TSW)
	Existing Learning Object	Information Bearing Thing (=IBT)
LO root relationships	Conceptual LO <->Copy LO	translationOfPCW(CW, TSW)
	Conceptual LO <->Existing LO	instantiationOfWork(CW, IBT)
Some LOM Metadata Examples	12_tittle	tittleOfWork(CW, String)
	13_language	languageOriginallyWrittenIn(CW, Language) languagePublishedIn(IBT, Language)
	16_coverage	timeIntervalInclusiveFn culturalRegionOfOrigin Other cultural information derived from the applicable microtheories

Figure 4: Excerpt of the mapping between LOM Standard and Opencyc

metadata and Opencyc.

A mapping between a concept of Opencyc and a concept of LOM means that all the instances of the Opencyc concept and its subtypes may be seen as instances of the LOM concept. Obviously, the inheritance relationship of Conceptual LO and Copy Conceptual LO also exists in the corresponding concepts of Opencyc (`CW` is supertype of `TSW`). Hence, the predicate `instantiationOfWork` defined in the context of `CW` is also applicable to `TSW`. One of the subtypes of `CW`, for example, is `DatabaseProgram`. So, this reinforces our statement that a computer program as a DBMS may be seen as a Conceptual LO. In that example, we would be able to create an instance of `DatabaseProgram` for the PostgresSQL. There are also subtypes of `IBT` that deals with running computer programs, such as `ComputerIBT`.

Some mappings depend on the level of abstraction of LO where are applied. For example, when dealing with the `CW`, the predicate `languageOriginalWrittenIn` describes its native language. In the other cases the language is not the original one and therefore it is represented by `languagePublishedIn`. The selected Opencyc predicate relates LOs with `Language` instead of `Human Language`. That makes possible to represent a LO that use formal (but not human) languages such as the mathematical language.

All derived data of our semantic interpretation may be also represented within Opencyc by using forward and backward rules. As an example, in the following we present the backward Opencyc rule that derives the title of existing LOs from their conceptual version:

```
(#$implies
        (#$and
                (#$isa ?IBT #$InformationBearingThing)
                (#$isa ?CW #$ConceptualWork)
                (#$instantiationOfWork ?IBT ?CW)
                (#$titleOfWork ?CW ?T)
        )
        (#$titleOfWork ?IBT ?T)
)
```

4 Conclusions and Future Work

Under our point of view, this work answers the research questions stated in the introduction by using Opencyc in order to study LOM standard possible improvements. As a result, we have categorized LOs and their metadata according to different conceptual levels. That categorization improves the semantics of the LOM standard and make possible to know which kind of LO we are dealing with.

A possible mapping between Opencyc and an excerpt of LOM standard also has been presented in this paper. That mapping may improve the interoperability between heterogeneous learning systems, helping to deal with some disambiguation and to establish a conceptual framework. The fact of using a very large ontology such as Opencyc allows further inferences, such as reusing the taxonomy that this ontology has. For example, the taxonomy that may be extracted from Opencyc that deals with Conceptual LO (`CW`) is of more than 4,000 concepts.

In near future, we plan to complete the mapping between LOM standard and Opencyc in all levels. That means, to rewrite the derivation rules within Opencyc and try to see how we can reuse the taxonomies of Opencyc in order to derive other relevant specializations of LOs, such as for example one based on the kind of learning resources (LOM 5.2 *learning resource type* metadata element).

References

[Aroyo et al. 2006] Aroyo, L., Dolog , P., Houben, G.J., Kravcik, M., Naeve, A., Nilsson, M., and Wild F.: “Interoperability in Personalized Adaptive Learning”; Educational Technology & Society, 9, 2 (2006), 4-18.

[Duval et al. 2002] Duval, E., Hodgins, W., Sutton, S.A., and Weibel, S.: “Metadata Principles and Practicalities”; D-Lib Magazine, 8, 4 (2002).

[Guarino 98] Guarino, N.: “Formal Ontology and Information Systems”; Proc. FOIS'98, IOS Press (1998), 3-15.

[Lenat 95] Lenat, D.B.: “Cyc: A Large-Scale Investment in Knowledge Infrastructure”; Communications of the ACM, 38, 11 (1995), 33-38.

[IEEE LOM 2002] LTCS WG12: “IEEE Learning Technology Standards Committee. Draft Standard for Learning Object Metadata”; Technical Report 1484.12.1, IEEE Inc, (2002).

[Polsani 2003] Polsani, P.R.: "Use and abuse of reusable learning objects"; Journal of Digital Information, 3, 4 (2003).

[Sicilia et al. 2004] Sicilia, M.A., García, E., Sánchez-Alonso, S., Rodríguez, M. E.: “Describing learning object types in ontological structures: towards specialized pedagogical selection”. Proc. ED-MEDIA’2004, (2004), 2093-2097.

[Sicilia 2006] Sicilia, M.A.: "Metadata, semantics and ontology: providing meaning to information resources"; International Journal of Metadata, Semantics and Ontologies, 1, 1 82006), 83-86.

[Wiley 2002]]Wiley, D.A.: “Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy”; In D. A. Wiley (ed.), The Instructional Use of Learning Objects: Online Version, (2002).