

# Hybrid System for Generating Learning Object Metadata

Olivier Motelet<sup>1</sup> – Nelson Baloian<sup>1,2</sup>

{omotelet,nbaloian}@dcc.uchile.cl

<sup>1</sup>*DCC-Universidad de Chile, Blanco Encalada 2120 Santiago, Chile*

<sup>2</sup>*GITS, Waseda University, 1011 Okuboyama, Nishi-Tomida, Saitama 367-0035, Japan*

## Abstract

*Generating LOM for learning material is a complex and tedious task to be completed “by hand”. Therefore, the current trend is to automate this process. However, there are several important issues restricting this approach. In order to solve these issues, this article presents a hybrid system in which both human and computer collaborate for instantiating LOM attributes. This system is implemented in LessonMapper2, a graphical interface designed to generate LOM-based graphs of learning resources.*

## 1. Introduction

IEEE LOM (Learning Object Metadata)([1]) is generally accepted as the standard for providing metadata to multimedia learning resources. The aim of using metadata for describing learning objects is to promote the sharing of learning material. Nevertheless, studies made about the real use of LOM shows that metadata is often misused or not instantiated ([2]). Such results are mostly due to the complexity of the specification: LOM includes almost 60 attributes. Such amount of elements is an obvious obstacle to the generation of LOM because many authors refuse to fill all the metadata. On the other hand, some of the values for the metadata are subjective and make it difficult to assign a value to them.

Many authors have already proposed the automatic generation of metadata as a mean to solve at least partially the problem of the provision of metadata values for learning objects, thus alleviating the work of the author or the learning object “librarian” ([3,4]). However, automatic generation of LOM is still limited. In particular, the educational attributes of LOM are very difficult to produce. However, educational metadata are a key topic for building systems allowing retrieval of learning objects in a pedagogically sounded manner. Indeed, such systems should not only focus on the retrieval process based on the content of the learning objects as a search engine like Google, but

also on the pedagogical context and characteristics for which they were done. Considering this issue, it appears still necessary to count on human intervention in the process of instantiating metadata.

Therefore, this article explores the development of hybrid systems in which both human and computer collaborate for instantiating LOM attributes. Our approach considers the lesson syllabus authoring as a favorable context for such collaboration, in which LOM values may help making the organization of the lesson more coherent. In particular, we consider a lesson as a graph in which nodes are learning objects and links define relations between them as defined in the LOM specification. In order to explore this proposal, we have developed LessonMapper2, a graphical front-end for building LOM-based graphs of learning resources.

This article discusses the characteristics of a hybrid system for generating LOM and illustrates these characteristics with their implementation in LessonMapper2. In section 2, the progresses and issues of automatic generation of LOM values are discussed. Next, section 3 deals with the problems of human-based generation of the LOM values. Then, graph-based lesson authoring is presented as a favorable context for easily instantiating LOM values. Finally, section 4 discusses the advantages of a hybrid system putting together the expertise of education professionals (authors or librarians) with the potential of automatic systems for generating LOM values.

## 2. Automatic Generation of LOM

Most authors agree on the fact that dealing with metadata cannot be a human task ([5]). On one hand, metadata creation is too difficult and time consuming work for content authors and the benefit of using them is not immediately appreciated. On the other hand, expert metadata creators are considered too expensive to be employed in most educational institutions.

## 2.1. Automatic Instantiation of LOM

Most metadata providing technical information (e.g., size, format, creation date, duration, etc.) about an educational resource can be automatically generated without problems by the same applications used to create them ([4]). Also some other type of metadata like the author's name and language can be sometimes automatically deduced from the context in which the application is running. Approaches for generating metadata using document contents are based on the process of extracting information identifying the subject of the educational resource (title, keywords, description, classification). However, these techniques are based on text mining and not all educational resources are text-based. In fact, rich educational resources make extensive use of multimedia, which is still difficult to analyze with the same content-based methods. Moreover, most of the useful information for the educational-related metadata remains implicit in the learning material and its usage.

## 2.2. Suggesting and Restricting LOM Values

From another perspective, we can try to find in the context in which the learning material has been used more opportunities for generating the metadata related to the educational description. In particular, techniques exploiting the semantics of the context in which the learning object is used seem to offer interesting opportunities for generating educational metadata. For example, Hatala and Richards ([6]) relate the semantics of the already existing metadata in order to generate new ones. In particular, they defined a set of specific rules concerning the inheritance between educational resources (from parent to children), the accumulation (from children to parent), and the content similarity between educational resources. Nevertheless, the correctness of the metadata generated with such techniques remains only approximate. Consequently, the authors claim that such values should be used as suggestions or for validations. Motelet and Baloian ([7]) also propose the usage of inference rules based on the relations between a certain educational resource and other material. This work is based on the principle that a semantic relation between two educational resources may impose a certain mutual influence between their metadata values. For instance, let's consider a first learning object which is explained by a second learning object, i.e. it exists a relation of type *explainedBy* between the first and the second object. We may say that this relation implies that the keywords of the first learning resource are quite similar to the keywords of the second. Such a rule enables a

deduction process for generating possible metadata values, which may be presented to the user as a suggestion. Furthermore, this principle is extended in order to generate rules formulating restrictions on metadata values. For example, if two learning objects are related with the semantic relation *introducesTo*, then the value of the attribute *semanticDensity* of the first resource is certainly lower than the value of the *semanticDensity* of the second. However, this assumption may not be valid for all potential users. For that reason, a domain specific language enables the creation of the rules. Thus, each community may be able to define their own rules according to their needs.

## 3. Human-Based Generation of LOM

Generating LOM "by hand" implies filling forms listing the numerous attributes of this specification. Human-based generation tools (e.g. Reload Editor) and interfaces of learning object repositories (e.g. Merlot, or Ariadne) are based on the form metaphor. Such systems frequently provide some additional support to the user: they explain the meaning of the attributes and suggest choosing them within a vocabulary defined by a particular community of practice. However, this process remains tedious and is generally accomplished separately from the authoring process. LOM are usually generated for packaging, classifying and publishing the learning material when there is a wish to share it. Nevertheless, we believe that LOM has potential for other usages. In particular, it may facilitate the characterization of the material during lesson authoring. To allow such practice, human-based generation of LOM may be supported by other metaphors, different from form-based user interfaces.

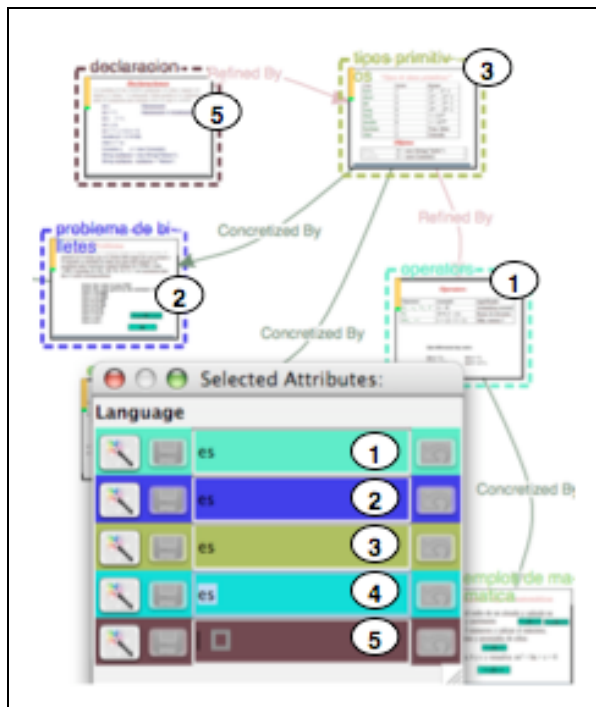
### 3.1. Packaging with LOM

Research work about LOM usage point out to the fact that most learning objects available in repositories are coarse grained. In fact, LOM are generally instantiated for packaging and sharing master educational pieces as an entire course or a complex resource like a simulation system or a flash animation. However, fine-grained learning objects are easier to reuse in various contexts than the coarse ones [8]. Learning Object Repositories are intended to facilitate the reuse of pedagogical material. However, most of them do not yet contain important quantity of varied fine-grained resources. Since education professionals and institutions are still reluctant to spend time and resources to pack and instantiate the metadata of only a few learning material items, it seems quite difficult to expect for human-based LOM generation of tens of

fine-grained learning material items. Until now, LOM as a packaging tool has shown too few benefits for LOM creators. Other benefits for LOM should definitely be found in order to stimulate their use.

### 3.2. Characterizing with LOM

From the beginning of the development of intelligent multimedia for learning, authors have proposed to structure learning material through graphs [9,10]. Indeed, lesson graph is a key structure for achieving flexibility. The LOM specification includes the notion of relations between two learning resources and so makes it possible to formulate graphs of learning resources. In such graphs, each node corresponds to a piece of material characterized with LOM. We have build LessonMapper2, a graphical application for building LOM-based graphs of learning resources. This tool was developed in Java and uses a Piccolo toolkit for building zoomable interfaces. It also uses JDOM so that the metadata of each learning object is directly modified in its XML form. LOM RDF is used to define a thesaurus for the attributes.



**Figure 1.** Simultaneous edition of a same metadata attribute for various learning objects with LessonMapper2.

In a lesson syllabus graph, a the graph nodes correspond to learning objects and they generally have several characteristics in common. For example, the language used in the different learning objects of a

same lesson is normally the same. In such situation, we believe that human-based LOM generation may benefit from instantiating simultaneously the *general/language* attribute of various nodes of the graph instead of independently repeating this task for each piece of material. In Figure 1, the *general/language* attribute is edited simultaneously for various learning objects. Here the value is the same for each piece of material and replicated by dragging and dropping it.

Editing simultaneously the same attribute for various resources also facilitates the comparison between their values. For example, instantiating the interactivity level of a resource makes sense when comparing this level with other resources. Instantiating this attribute in an isolated manner and without guidelines is a tedious task and does not directly benefit the education professionals. On the other hand, if it is instantiated during the authoring of lesson syllabus, this information may serve the coherence of the learning design of the lesson. This same conclusion may be drawn for most subjective attributes of LOM, that is to say those attributes that are difficult to instantiate in an objective and rational manner. This category concerns almost all the *educational* attributes. However, the instantiation process of the objective attributes like most technical attributes, offers much less potential benefits for the instructor than the latter ones. In this case, automatic generation systems are really necessary.

## 4. Hybrid Systems for LOM Generation

In a study about metadata generation methods, Greenberg [11] concludes, “*best metadata generation option is to integrate both human and automatic processes*”. According to this principle, metadata for educational resources should remain instantiated by content authors or professional metadata creators, but this instantiation process should be supported by the suggestions and restrictions of various automatic generation methods. A tool supporting such a process is a hybrid engine between editors and generators. Typically, it will support both LOM Instantiation and LOM Validation.

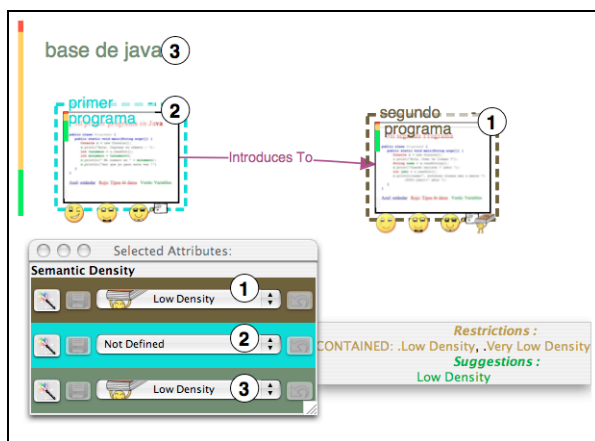
### 4.1. LOM Instantiation

The information produced by the automatic systems exposed in the second section may be classified in three groups:

- *the very probable values*: such values are the results of automatic generation systems like in [3]. They concern objective attributes like the technical format or size

- *the probable values*: these values concern the suggestions done by intelligent systems. Typically, these suggestions are extracted from the context like learning management systems or related learning material ([5,6,7]). These values are generally not enough reliable to serve automatic instantiation purposes, but may be interesting for creators of metadata.
- *the restriction of possible values*: analyzing the context around a learning object allows to deduce some restrictions about the LOM values of this learning object ([6]). In the context of a community of practice in which the sets of possible values for LOM are well defined, these restrictions may serve to reduce the scale of possible values.

On the point of view of a human-based instantiation of LOM, all these groups of information are relevant. First, the very probable values may be automatically instantiated without human intervention. Second, suggestions may be displayed to help the user in the process of metadata value instantiation. Third, the restrictions may also be used to speed up the instantiation and limit the cognitive overload related with the large scale of possible value for a same attribute ([2]).



**Figure 2.** Suggestions and restrictions for the possible values of a LOM attribute with LessonMapper2.

We intended to implement part of these concepts in our LOM-based lesson graph builder, LessonMapper2. In particular, LessonMapper2 integrates the LOM semantic diffusion system presented in [6]. As described in section 2, this system uses the semantic of LOM-based lesson graphs for generating suggestions and restrictions for LOM values. Figure 2 shows the semantic density of three learning objects: (from top to bottom) “segundo programa”, “primer programa” and “base de java”. The two first are contained in the third.

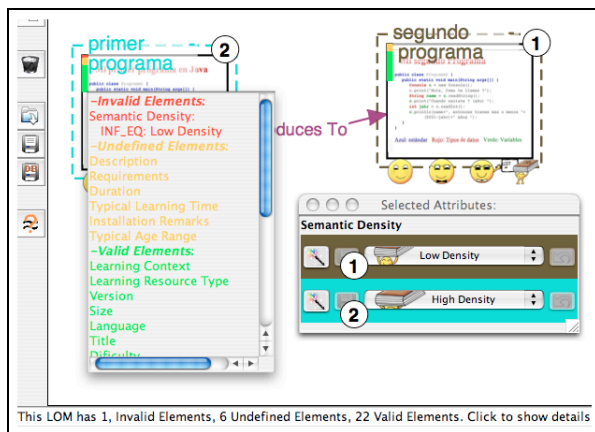
The semantic density of “primer programa” is not defined. In this example, suggestions and restrictions were asked for this attribute.

Since “primer programa” is part of “base de java”, the system suggests the value of the semantic density of “base de java” for instantiating the semantic density of “primer programa”. Since “primer programa” introduces to “segundo programa”, the system applied the restriction that the semantic density of “primer programa” should be inferior or equal to the semantic density of “segundo programa”. Therefore, a set of possible values is proposed according to this restriction. In LessonMapper2, simple drag and drop allows to adopt the values deduced as suggestions or restrictions.

## 4.2. LOM Validation

In order to provide full access to an educational resource, it has to be ensured that all the information is annotated with valid metadata. Valid metadata should satisfy a minimum level of completeness and correctness. The analysis of completeness of LOM simply consists in checking the number of instantiated attributes. Complete LOM have all their attribute instantiated. Correctness evaluation is more complex because it deals with the semantics of the metadata values. A first method consists of checking the validity of the data type. Basically, this test can be done using the XML Schema validation mechanism. Another method consists in checking whether the given value belongs to a set of vocabulary terms. The Reload editor, e.g., enables to check the vocabulary corresponding to a different version of the IMS Metadata definition. Nevertheless, none of the available editors provides a deep checking of the meaning validity, because this topic has to do with both the semantics of the metadata and the semantics of the educational resource. For this purpose, Ochoa et al. [3] suggest a framework for using various automatic metadata generation methods for cross-validating metadata. In LessonMapper2, we propose a different method. It consists of using the restrictions deduced from the analysis of the graph semantics for checking the correctness of the values. Even if this method does not provide sufficiently accurate results to safely validate the correctness of the metadata, it may provide relevant data about metadata coherence for a set of educational resources. As shown in figure 3, in LessonMapper2, all the learning material items of the graph are decorated with tricolor bars representing the proportion of invalid, undefined, and valid elements for the LOM attributes of each learning material item. Invalid elements hold values not satisfying with the restrictions deduced by the system. Undefined

elements have not yet any value assigned. Finally, valid elements hold values successfully passing both tests of completeness and correctness.



**Figure 3.** Displaying validation state of the metadata of a learning object with LessonMapper2.

The validity bar permits to spotlight the elements forgotten by the instantiation process and prevent from the incoherences as far as they can be detected by the rule-based system generating the restrictions. The bar is automatically refreshed when a change occurs in the metadata values of graph elements. Figure 3 shows the validity results of the learning resource “primer programa” in their detailed form. This view exhibits the list of reasons for eventual issues. In this example, we attached a high semantic density to the element “primer programa”. Nevertheless, as described before, the semantic density should be inferior or equal to the low semantic density of “second programa”. Since the current value does not satisfy this restriction, the element *semanticDensity* is tagged as invalid.

## 5. Conclusion

This article discussed the issues of both automatic and human-based generation of LOM. In order to bypass the problems at stake in these approaches, we suggested an original hybrid system for LOM generation, LessonMapper2. The authoring of the lesson syllabus as a graph of learning resources is shown as a favorable context for such a system. LOM is used not only for packaging the learning material items, but also for characterizing them and potentially helping the design of the lesson syllabus. Our approach also attempts to integrate the results of automatic LOM generation tools in order to assist the instantiation process. In particular, the system provides suggestions and restrictions on LOM values. These features may speed up the instantiation process and limit its

cognitive overload. Another advantage of this hybrid system is to evaluate the validity of LOM values. This evaluation is used to highlight the forgotten elements and the incoherent values. In fact, LOM instantiation state for each learning material items is shown permanently and updated automatically when a change occurs. On the one hand, the omnipresence of LOM during lesson authoring may be seen as an additional noise disturbing the complex process of designing a lesson. On the other hand, LOM may sound useful if it is considered as a relevant tool for characterizing learning objects and helping the design of the lesson syllabus. We hope that the development of hybrid systems such as LessonMapper2 is a step forward in making this second statement a reality.

## 6. References

- [1] IEEE LOM specification. <http://ltsc.ieee.org/wg12>. Last accessed on February 4, 2006.
- [2] Friesen, N. International LOM Survey: Report. ISO/IEC JTC1/SC36 sub-committee, 2004
- [3] Ochoa, X. et al. Frameworks for the Automatic Indexation of Learning Management Systems Content into Learning Object Repositories. Proc. of ED-MEDIA 2005, Montreal, pp. 1407-1414.
- [4] Pinkwart, N. et al. Partial Generation of Contextualized Metadata in a Collaborative Modeling Environment. Proc. of 2nd International Workshop on Applications of Semantic Web Technologies for E-Learning - AH 2004, Eindhoven.
- [5] Duval, E., Hodgins, W. (2004). Metadata matters. Proc. of DC-2004, Shanghai, China.
- [6] Hatala, M., & Richards, G. Value-Added Metatagging: Ontology and Rule Based Methods for Smarter Metadata. Proc. of RuleML 2003, LNCS 2876, pp. 65-80.
- [7] Motelet, O., & Baloian, N. A. (2005). Taking advantage of LOM semantics for supporting lesson authoring. Proc of WOSE – OTM 2005 Workshop, LNCS 3762, pp. 1159-1168.
- [8] Wiley, D. The Reusability Paradox, Connexions Web site. <http://cnx.rice.edu/content/m11898/1.18/>, last accessed on February 4, 2006.
- [9] McCalla, G.: The search for adaptability, flexibility, and individualization: Approaches to curriculum in intelligent tutoring systems. In: Foundations and Frontiers of Adaptive Learning Environments. Springer (1992), pp. 91–122.
- [10] Baloian, N.A., Pino, J.A., Motelet, O.: Collaborative authoring, use, and reuse of learning material in a computer-integrated classroom. Proc. of CRIWG. 2003, LNCS 2806, pp. 199–207.
- [11] Greenberg, J. Metadata Extraction and Harvesting: A Comparison of Two Automatic Metadata Generation Applications. Journal of Internet Cataloging: The International Quarterly of Digital Organization, Classification, and Access, 6(4), 58-82.