Complete Metadata Records in Learning Object Repositories: Some Evidence and Requirements

Miguel-Ángel Sicilia, Elena García, Carmen Pagés, José-Javier Martínez, José-María Gutiérrez
Computer Science Department, University of Alcalá
Ctra. Barcelona km. 33.6, 28871 Madrid, Spain
{msicilia, elena.garciab, carmen.pages, josej.martinez, josem.gutierrez}@uah.es

Abstract: A learning object can be considered as a unit of instructional content for which a metadata record describing its characteristics and intended educational usage is provided. Metadata records can be used to develop effective search and location of learning objects, and also to develop automated or semi-automated selection and composition tools. In consequence, the quality of metadata records is a critical point for applications, especially if consistent and standardized software tools are desired. Completeness is one of the essential facets of metadata quality, which can be defined in terms of the metadata elements required for each functionality or usage. The eventual emergence of a global space for learning object-based education requires the creation of learning object repositories providing large collections of contents in a form accessible to standardized software. These repositories are called to play a central role in automated approaches to e-learning, since they provide the required support for learning object access and search facilities, oriented not only to humans but to software agents or systems. In consequence, completeness of metadata records becomes a key requirement for learning object repositories. Nonetheless, metadata creation is a time-consuming and laborious process, and current basic standards for learning object metadata allow for a large degree of flexibility in metadata edition. These two factors combined may eventually result in incomplete and poorly structured metadata. In this paper, the completeness of learning object metadata of samples obtained from the MERLOT and CAREO repositories are analyzed from that viewpoint, using the IEEE LOM standard as a reference framework. The paper concludes with a proposal for the specification of completeness levels as compliancy requirements for learning-related services or processes.

Keywords: Learning objects, metadata, learning object evaluation, learning object repositories.

Biographical Notes:

Miguel A. Sicilia obtained a university degree in Computer Science from the Pontifical University of Salamanca, Madrid, Spain, in 1996 and a Ph.D. degree from the Carlos III University in 1999. From 1997 to 1999 he worked as an assistant professor and later on as a part-time lecturer at the Computer Science Department of the same university. He also worked as a software architect in e-commerce consulting firms. From 2002 to 2003 he worked as a full-time lecturer at the Carlos III University, after which he joined the University of Alcalá. His research interests are primarily in the areas of adaptive hypermedia, learning technology, and human-computer interaction, with a special focus on the role of uncertainty and imprecision handling techniques in those fields.

C. Pagés obtained a University degree in Computer Science for the Technical University of Madrid. She worked in private telecommunication business companies as analyst, project manager and consultant between 1987 and 2003, and is at present professor of the Department of

Computer Science at the University of Alcalá, Madrid. She has practical work experience on software process technology and modelling, methodologies for software projects planning and managing, software maintenance and international ERP projects. She is currently working on learning technology, focusing on learning objects and tutorial systems.

Elena García-Barriocanal obtained a university degree in Computer Science from the Pontifical University of Salamanca in Madrid (1998) and a PhD from the Computer Science Department of the University of Alcalá. From September 1998 to February 1999 she worked as a lecturer in the Computer Languages and Information Systems Department of the Pontifical University, and in 1999 she joined the Computer Science Department of University of Alcalá as assistant professor. Starting from 2001, she is associate professor at Computer Science Department of the University of Alcalá and she is a member of the Information System Engineering group of this University. Her research interests mainly focus on topics related to human-computer interaction and knowledge representation; concretely she actively works on ontological aspects in usability and accessibility and supervises several phD works in these areas.

J.J. Martínez-Herráiz obtained a University degree in Computer Science for the Technical University of Madrid. He worked in private telecommunication business companies as analyst, project manager and consultant between 1988 and 1999, and is professor of the Department of Computer Science at the University of Alcalá, Madrid, since 1994. He has practical work experience on software process technology and modelling, methodologies to software projects planning and managing and software maintenance. He is currently working on e-learning technology, focusing on applied intelligent techniques.

José M. Gutiérrez is professor at the Universidad de Alcalá since 1998. At the moment he works on the mobile devices integration possibilities in the information systems, having presented his research results in congresses and having produced a book about mobile phones programming. His previous work deals with representation of information by language XML and the posiblilities of systems automatization by representing the semantics of the systems through this language. For two years, he has been working in several projects helping the information systems development in the Institutions and Universities of various Central American countries. Even though, he participates in lots of e-learning activities, ranging from the publication of investigation results to the creation of new e-learning tools.

1. Introduction

The concept of learning object (Wiley, 2001) has become a central element in current approaches to Web-based learning technology. Learning objects are dual entities in the sense that they are made up of two parts: the actual collection of educational contents and a description of them often called "metadata record" including suggestions about appropriate contexts of use for the object (Sosteric & Hesemeier, 2002). Metadata records are descriptions that can be used to search and target learning objects, and are thus a vehicle for sharing and reuse, as described by Downes (2001). Several learning object definitions (Polsani, 2003; IEEE, 2002; Wiley, 2001) have in common the consideration of reuse as a fundamental characteristic. For example, as defined by Polsani (2003), learning objects are "independent and self-standing units of learning content predisposed to reuse in multiple instructional contexts". The predisposition to reuse mentioned in this definition can be considered from two different but complementary perspectives. On the one hand, reuse from the viewpoint of human experts (teachers, course designers and the like) requires, among other design criteria, that the contents were designed with a degree of independence to any particular educational setting, e.g. they should not contain

references to the context of a concrete school or institution, and their granularity should be adequate to fit in several educational contexts. But on the other hand, reuse from the viewpoint of developing software that deals with learning objects critically depends on the quality and machine-understandability of metadata records. From the latter perspective, metadata records should be expressed through standard or shared metadata schemas that enable a degree of consistency in the behaviour of Learning Management Systems (LMSs) and that provide the foundations for building automated software or software-supported tools. But using standardized schemas does not guarantee that metadata properly supports the need for any kind of functionality, since some of the metadata elements may be missing or improperly specified for a given purpose. Metadata definitions are essentially intended to support some activities or processes, as described by Greenberg (2003). For example, a metadata element "cost" is intended to support purchase activities. In consequence, each kind of activity or process can be said to require a number of concrete metadata elements. Examples of such kind of connections between required metadata elements and processes are provided in (Sicilia et al., 2004) in which metadata elements in the LOM standard are associated to processes like learning object acquisition, composition or adaptation. In fact, even the reuse of a learning object achieved by software agents or tools critically depend on the quality of metadata (Sicilia & García, 2003).

The quality of metadata records can be approached from diverse perspectives. Standards-compliancy is a basic quality requirement, and it is easy to check by automated means, e.g. compliance tests yet exist for SCORM[1]. But standard-compliancy does not guarantee that metadata records are complete or appropriate. In fact, the LOM standard (IEEE, 2002) is fully optional, so that LOM-compliant metadata records should not be expected to have all the elements properly described. Appropriateness, understood as the correctness of metadata with regards to the actual contents of the learning object, is a matter of technical or pedagogical judgment that requires the intervention of evaluators or human experts in each instructional domain. Nonetheless, completeness is a concept that characterizes the availability of the "required" metadata items, so that a shared understanding of what a "complete" metadata record is can be easily elaborated and even automated. Within this conceptual framework, a metadata record for a given learning object may be correct (appropriate), but clearly insufficient (incomplete) for a human or a machine to be able to judge about its adequacy to a given educational context.

Since we put an emphasis on the development of LMSs with a high degree of automation, we will consider machine-understandability as a mandatory requirement for completeness of metadata records in what follows. Without this requirement for structured information, automated retrieval processes may be hampered by the inherent difficulty of natural language understanding by means of software. This fact precludes using (only) "free-text" metadata fields like the *Annotation* category of the LOM standard, since it gives room for arbitrary unstructured comments on pedagogical use, which have not a clear semantics for software systems, even in the case that some kind of natural language machine understanding is available.

In this paper, we address the first step of the problem of characterizing completeness of metadata records as a key quality factor for learning objects. This objective requires a previous assessment of the current state of metadata creation practices to determine if it actually exists a need to concern for completeness. In consequence, we have first conducted empirical analysis on two open learning object repositories (MERLOT and CAREO) to gather some initial evidence about the actual practices of metadata annotation. This study represents a first step towards a notion of completeness that could eventually be used as an estimator of the quality of metadata records, and in aggregated terms, of learning object repositories. The empirical study described here

complements the one reported by Najjar et al. (2003) about the actual use of metadata in the ARIADNE project, in which a significant lack of completeness in metadata records is reported.

The results of the empirical study together with recent efforts to state required metadata elements for automated processing have resulted in a functional notion of metadata completeness that can be effectively used to specify the compatibility of learning object metadata of different levels of completeness to consistent LMS behaviours of a diverse kind. Even though some kind of tacit completeness notion is contained in normative specifications like the IMS Simple Sequencing[2], defining completeness as a separate construct provides the benefits of having a tool to assess the degree of quality of learning objects according to the functionalities they are prepared to be used in.

The rest of this paper is structured as follows. Section 2 provides the results and discussion of the empirical analysis. Then, a notion of metadata completeness for learning objects is elaborated in Section 3. Finally, conclusions and future research directions are provided in the last section.

2. Some Empirical Evidence about Metadata Completeness

The empirical study has been designed to find evidence about the actual use of metadata in existing learning object repositories. The LOM standard have been used as a reference framework, since it is widely accepted and it has been adopted or recognized by the major specification bodies in learning technology. Although the study is far from being exhaustive, it provides some evidence that can be added to other empirical studies like that of Najjar et al. (2003) about ARIADNE to have an idea of the actual availability and completeness of metadata. At the best of our knowledge, no public standards-based learning object repository provides a significantly different figure of metadata completeness than ARIADNE and the two repositories reported in what follows.

Two questions were initially issued about actual usage of metadata:
(1) Which amount of metadata is provided for the learning objects?
(2) What part of the metadata provided is specified in structured form?

The first question is aimed at obtaining a figure of the completeness of metadata records in a general sense, since it includes in the definition of "metadata" both standards-based, structured descriptions, and also comments or free text associated to the object. Even though both kinds of metadata are useful to human readers, only the former can actually be used by software agents or systems. This raises the need for the second question, which is intended to indirectly assessing if current metadata creation practices are producing metadata that is useful for automation.

The study on the MERLOT[3] repository was intended to obtain evidence about both questions together. MERLOT was selected due to the fact that it provides three kinds of description levels (details, member comments and peer reviews), resulting in a significant amount of information coming from different sources for each learning object. The CAREO[4] repository was selected as a second target, since it provides an almost complete LOM-based metadata creation form. The study on CAREO is merely quantitative on the number of metadata elements that are actually provided for a selection of several thematic categories of learning objects. It should be noted that each repository is idiosyncratic in what it considers as a "learning object", and several of them do not explicitly adhere to any of the definitions referenced above. Nonetheless, this is largely

irrelevant for our current study, since LOM metadata can in principle being applied to any identifiable Web content, irrespective of its granularity and relative degree of independence.

2.1. Structured and Unstructured Metadata in MERLOT

The MERLOT repository (Cafolla, 2002) is an open learning object database available in the Web, and it was explicitly designed to support collaborative edition of metadata records, by means of informal commentaries and also peer-reviews. Peer-reviews are conducted by experts in the academic field of the topic being reviewed, that actually use Web resources in teaching their academic discipline. Learning objects in MERLOT are categorized both by subject area (Arts, Business, Education, etc.) and by academic discipline communities (Biology, Business, Chemistry, etc.). Our study was aimed at obtaining a portrait of the status of structured and not structured metadata inside MERLOT, as a representative of current annotation practices. The comparison is aimed at gathering some evidence above research question (2) described above.

The study was restricted to learning objects falling into the area of academic teaching of the researchers elaborating the study, namely "Computer Science" and "Information Technology", since the analysis of unstructured text requires a degree of knowledge in the domain. By July 2003, the number of learning objects catalogued in these categories was of 457 and 659, respectively, from which we selected the subset of 52 objects (100%) for which associated peer-reviews were available. This was due to the fact that this subset of objects could be formally considered as appropriate according to MERLOT philosophy, and allow for a better comparison between unstructured and structured metadata. The procedure for the study was simply that of inspecting each of these 52 objects, in search of its compliancy with the LOM specification, which was used as the reference framework. The inspection included the main "Detail View", the "Peer Reviews" and the "Member Comments" (if any) associated to each object.

Table 1 provides the resulting quantitative data about existing metadata, considering only educational related metadata items and not those for format or technical content. The quantities are split in four categories: [a] *structured,* referring to information items that were found as a labelled field in the main section, [b] U-D (*unstructured in description*), referring to information found in the unstructured description of the main section [c] U-MC (*unstructured in member comments*), for items found among member comments, and [d] U-PR (*unstructured in peer reviews*), for information available in unstructured form inside peer reviews. Gathering the data required some degree of interpretation of the intended use of LOM elements, since most of the information elements were found in natural language, narrative descriptions.

Table 1. Quantitative data for a sample of peer-reviewed learning objects in MERLOT

Metadata item	Structured	U-D	U-MC	U-PR	Overall
LOM:General:Coverage	0	0	0	0	0
LOM:General:Structure	0	15	2	34	51
LOM:General:AggregationLevel	0	17	0	34	51
LOM:Educational:InteractivityType	0	10	0	41	51
LOM:Educational:LearningResourceType	18	5	2	45	70
LOM:Educational:InteractivityLevel	0	1	2	45	4

LOM:Educational:SemanticDensity	0	0	4	30	34
LOM Educational:Intended End User Role	0	11	1	29	41
LOM:Educational:Context	45	0	1	6	52
LOM:Educational:TypicalAgeRange	0	0	0	1	1
LOM:Educational:Difficulty	0	0	3	41	44
LOM:Educational:TypicalLearningTime	0	1	1	0	2

The main conclusion that can be extracted from Table 1 is that peer reviews are the vehicle for most of the information about metadata elements, while structured descriptions are scarce. The paradox raised by this result is that the effort to providing educational information has actually be spent, but the narrative form of peer-reviews make such information not machine-understandable. It can be argued that a structured, detailed variant for producing structured peer-reviews would end up with high-quality metadata, since it is the format and not the contents that is lacking.

The quality, validity and reliability of the information included in comments and peer reviews of MERLOT learning objects are endorsed by the area-responsible instructional team. Nonetheless, a number of improvement areas can be pointed out:

1. The amount of metadata information is low, relative to the large amount of learning objects in Merlot. Concretely, only a 4,7% of the objects in the categories addressed by this study have been peer reviewed. About a 60% is only described by basic data, and the rest is provided with personal comments, but not peer reviews.
2. There exist some incoherencies and inconsistencies when comparing judgments coming from comments and peer reviewers. These divergences become more evident when analyzing judgments and issues given regarding the estimations of the difficulty of learning for a given learning object. This rests value to this particular information item since a large amount of subjective or contextual considerations appear to be considered by reviewers. Perhaps more strict commenting guidelines would come up with more convergent annotations.
3. In general terms, most data is not structured but provided in less concrete comments. This of course entails that the understanding of such information requires careful reading from human beings.

The inspection of metadata records carried out resulted in a number of metadata descriptions in unstructured form that are not easily mapped to LOM elements, including the following:
1. "Depth" of the educational content, defined as the "extent of penetration" in the proposed ideas. This concept is different to others that exist in LOM like Difficulty or Semantic Density. Difficulty refers to the effort required to master contents, while Semantic Density indicates the degree of concreteness and compactness of descriptions.
2. "Timeliness" of the educational content, defined as the utility of the information at the time it is being assessed, i.e. it refers to the fact that some contents, especially in technology, become outdated and obsolete, while others remain current.
3. "Pedagogical use", defined as the specific utility of the learning object with respect to the global training of the educational content. The content of this element is more subtle and complementary to the *Context* LOM metadata item, since the ultimate goal of the later is

to inform about the principal environment within which the learning and use if this learning object is intended to take place, while the "pedagogical use" element can be used to explain that the learning object is only useful as a strengthening resource, as a self-study guide and other kinds of specialized uses.

4. "Pedagogical preference", defined as the most effective use of the learning object among the existing instructional theories. This concept is linked with the LOM metadata *Learning Resource Type*, which points out the specific kind of the learning object: exercises, graphics, text, etc, but the pedagogical preference can be used to prescribe specific methods and techniques. For example, it can be used to prescribe the use of top-down or bottom-up methodologies, or to give preference to a kind of support, e.g. textual or diagrammatic.

2.2. Cross-domain quantitative analysis in CAREO

The CAREO Repository has as its primary goal the creation of a searchable, Web-based collection of multidisciplinary teaching materials for educators. As showed in the screenshot in Figure 2, the repository provides facilities for browsing, searching and editing of learning objects. The addition of an object to the repository can be carried out through the interface showed in Figure 3, in which a form for IMS-based metadata elements is provided, along with on-line help to aid in the edition process.

We have used the Learning Commons[4] version of the CAREO interface, which is actually the one showed in Figures 2 and 3. The repository is organized in over seventy disciplines, including languages, computer science, arts, economics and many others. This second study was intended to have a comparative account for the presence of metadata in several different disciplines. The domains of Computer Science (CS), Education (E), Sciences (S), Accounting (A), Medical Science (MS) and Mathematics (M) were selected for the study, as a representative range of disciplines. The amount of learning objects registered at the time of the study was of 21, 46, 97, 43, 21 and 47, respectively for CS, E, S, A, MS and M, but only a sample with the first twenty one appearing at the repository for each discipline were analyzed. Using again the LOM standard as a reference framework, the quantitative results are provided in Table 2, which shows the amount of metadata specified in the learning objects.

Table 2. Quantitative metadata analysis of six disciplines in Learning Commons repository.

LOM	Name	CS	E	S	A	MS	M	Total
1.1.1	Catalog		5			5		10
1.1.2	Entry		5			5		10
1.2	Title	20	20	20	20	20	20	120
1.4	Description	20	19	20	20	20	20	119
1.5	Keyword	19	15	13	20	15	20	102
2.1	Version		1	1				2
2.3.1	Role	20	11	10	20	5	20	86
2.3.2	Entity	20	16	17	20	20	20	113
2.3	Date	19	6	8	20		20	73
3.1.1	Catalog		5					5
3.1.2	Entry		5					5
3.2.1	Role	20	5	6	18		20	69
3.2.2	Entity	20	5	6	18		20	69
3.2.3	Date	19		6	18		20	63

4.1	Format	20	20	20	20	20	20	120
4.2	Size		6	6		20		32
4.3	Location	20	20	20	20	20	20	120
4.4.1.1	Type	5		5	10			20
4.4.1.2	Name	5		7	13		20	45
4.4.1.3	Minimum version						20	20
5.2	Learning resource type	19	6	10	19	15	20	89
5.5	Intended end user role	19	5	6	20		20	70
5.6	Context	19	6	7	20	5	20	77
5.10	Description		1	3		11		15
6.1	Cost	19	5	6	20		20	70
6.2	Copyright and other restrictions	19	5	6	20		20	70
6.3	Description	19	6	11	14	15	20	85
7.1	Kind		6	5		5		16
7.2.2	Description		7	5				12
9.1	Purpose		6	4		5		15
9.3	Description		1					1
9.4	Keyword	20	20	20	20	20	20	120
Total		361	238	248	370	226	400	1843

The overall amount of metadata in each discipline has a fairly similar volume across the disciplines of CS, A and M, and also across the disciplines of E, S and MS. Nonetheless, the differences between the two groups are not significant enough to draw any conclusion from them.

The numbers in Table 2 suggest that there are some "preferred" metadata elements, in the sense that they are filled most of the times. These elements include title, description and keywords, the field "entity" describing contributor(s) to the metadata record, the technical format of the learning object (e.g. video, image, HTML), the location of the learning object (i.e. its URL), and keywords for the learning object as part of the element "Classification". Given that the contributor field is filled by default with the username of the author of the metadata record, it becomes clear that the preferred elements follow the intention of identifying the object (4.3), and describing its format (4.1) and its overall subject (1.2, 1.4, 1.5, 9.4).

Surprisingly, educational metadata (category number 5 in LOM) does present a considerable degree of incompleteness, except for the learning resource type element, which is intended to describe the kind of learning object, e.g. exercise, narrative text, exam, etc. Elements like interactivity level or type, semantic density and difficulty are not used, and the educational context – used to describe the principal intended usage environment, e.g. school or training - is missing in more than thirty percent of the objects. It should be noted that educational elements in LOM are essential for the process of learning object selection and they are closely connected to notions of reusability described elsewhere (Sicilia & García, 2003), so that improved metadata at least in this area should be subject to further study.

The overall results of this study support the intuition that providing free edition capabilities to learning object metadata may not be a wise approach for the sake of completeness, and stricter creation interfaces could be worth the effort if complete metadata records for certain usages is required. In the following section, this idea is elaborated in terms of levels of compliancy with processes.

3. Completeness as a Requirement for Process Compliancy

The study described in the previous section suggests that stricter metadata edition interfaces and practices could end up with metadata of a much better quality. Nonetheless, the justification of making edition more constrained (and consequently, more demanding and time-consuming) requires a framework in which metadata creators and users clearly understand that their extra efforts will result in learning objects that are better equipped to be used by software tools to carry out concrete automated activities or processes.

Such notion of completeness connected to the enablement of processes has been proposed recently (Sicilia et al., 2004) as part of the specification of processes of acquisition, composition, publication and selection of learning objects.

In consequence, we propose the following working definition for learning object metadata completeness: "the extent to which the metadata record of the learning object provides the necessary metadata to properly support a given process or activity (capable of being automated, partially or totally)". The reference to automation may be arguable, but in our view, a strong commitment with software development needs is required from the perspective of making learning technology evolve. This definition can be extended to include metadata profiles, in the sense described by Downes (2004), in which a learning object may have several, possibly even conflicting, metadata records.

An important consequence of the definition provided is that completeness is a concept relative to a process or activity (or a collection of them) that handles in some way some metadata elements. A derived definition for "overall completeness" could be considered as the aggregate completeness for a set of processes that are considered to be the important ones for a specific technological context. It should be noted that completeness only reflects a basic aspect about metadata quality, since it may happen that a *complete* metadata record is actually not *appropriate*, i.e. it provides metadata values, but they do not describe properly the contents of the learning object. In this case, an automated process will be able to use the object, but it will probably produce meaningless results. Nonetheless, completeness can be considered as a prerequisite to appropriateness, and it provides a way to easily identify not appropriate metadata records, since incomplete records for a given activity are in any case not appropriate, due to the lack of essential information. The checking of completeness can be easily automated through software tools, while the checking of appropriateness will typically entail the assessment of humans, so that the completeness checking may be justified in terms of saving assessment efforts.

The measure of completeness in the above described studies has been described as the counting of the metadata elements that were present, without considering the actual form of that metadata values. An extension of the concept may also include what has been called as "metadata *idiom*" (Sicilia et al., 2004), which could be defined as "a constraint placed on the acceptable values of a concrete metadata element that does not break the semantics of the original definition of the element". A trivial example of idiom is restricting the values of a vocabulary for a given element in LOM, so that the resulting metadata is still LOM-compliant.

In order to illustrate the concept of completeness, we will describe in what follows several types of composition processes (CMP) for learning objects. We will assume that the problem of

composition takes as inputs a collection of learning objects {LO} and a set of concepts {C} that are intended to describe the intended outcomes of the aggregated learning object. Even though this definition ignores several facets of the problem of composition, it can be used to specify a simple scenario for learning object selection. We can specify, for example, three different composition processes as showed in Table 3, where ID-1 stands for basic identification data for the learning object, i.e. location and identifier.

Table 3. Example composition process specifications.

Process	Required metadata	Metadata idioms	Intention
CMP-A	ID-1 Classifications	Classifications should refer to a term inside a shared, open ontology used by the elements in {LO}.	Compose learning objects to cover all the required concepts in {C}.
CMP-B	ID-1 Classifications	Classifications should refer to a term inside a shared, open ontology used by the elements in {LO}.	Compose learning objects to cover all the required concepts in {C}, including sub-concepts of concepts in {C} connected through "part-of" relationships.
CMP-C	ID-1 Classifications Relationships	Classifications should refer to a term inside a shared, open ontology used by the elements in {LO}. All the prerequisite relationships of the learning object must be defined as a set of concepts {C2}.	Compose learning objects to cover all the required concepts in {C}, including also the declared prerequisites in {C2}.

The three CMP process sketched in Table 3 respond to three levels of functionality, resulting in three compliance levels. Learning Management Systems will select the processes to which they comply, as a competitive advantage or positioning with respect to automation. Of course that other requirement on the learning objects being composed will be needed, as the proper technical separation of presentation and content to be able to merge their appearance into a consistent one, but we omit such details here for brevity. It should be noted also that some processes are extensions of others, for example, CMP-B and CMP-C cover also CMP-A, but they do not cover each other.

The assessment of completeness requires an additional extension to existing metadata schemas: the possibility of including null values. For example, in CMP-C, the provision of a null value in the Relationships element will have the meaning that the learning object does not have any kind of relationship, while the absence of value for the same metadata element should be interpreted as missing metadata, which can not be interpreted as "no relationships".

Edition interfaces designed for completeness to process specification will require the automatic checking of the presence of certain metadata elements. In addition, learning object repositories would be able to provide search facilities in which the user specifies the "desired processes", and the repository filters out the objects that do not comply with them.

4. Conclusions and Future Work

Existing evidence about the actual usage of metadata in learning object repositories point out to a significant lack of structured information, which seriously hampers the development of software tools that search, select, aggregate and manipulate learning objects.

The concepts of completeness of learning object metadata records and completeness of learning object repository are required to serve as quality indicators about the descriptions of reusable learning contents. Such indicators would be connected to the evaluation of reusability of learning objects, since reusability requires precise enough descriptions to be able for a human or a software module to retrieve the appropriate items, and also to be able to decide its appropriateness for the usage context at hand.

This paper reports a first step towards a realistic characterization of completeness given the existing resources. Concretely, two studies in open repositories have been reported, and a concept of metadata completeness for learning objects has been provided, justifying stricter metadata creation as a way to comply with concrete processes or activities that could eventually be carried out in an automated or semi-automated way.

Much empirical research is still needed with regards to the metadata that is being actually created inside learning object repositories. Empirical evidence can be used to hypothesize about the appropriateness of concrete metadata creation practices or user interfaces, which would eventually result in elaborated best practices that improve existing free editing approaches that actually result in a significant lack of completeness. In addition, the notion of completeness of learning object metadata requires extensive reflection and debate before being considered as a key quality factor for metadata.

References

Cafolla, R. (2002). Project Merlot: Bringing Peer Review to Web-based Educational Resources. In *Proceedings of the USA Society for Information Technology and Teacher Education International Conference*, pp. 614– 618.

Downes, S. (2001). Learning Objects: Resources For Distance Education Worldwide. *International Review of Research in Open and Distance Learning* 2(1).

Downes, S. (2004). Resource Profiles. *Journal of Interactive Media in Education*, 2004 (5).

Greenberg, J. (2003). Metadata and the World Wide Web. In Dekker, M. (ed.): *Encyclopaedia of Library and Information Science*, pp. 1876-1888.

IEEE Learning Technology Standards Committee (2002). *Learning Object Metadata (LOM)*, Final Draft Standard, IEEE 1484.12.1-2002.

Najjar, J., Ternier, S. and Duval, E. (2003). The Actual Use of Metadata in ARIADNE: an Empirical Analysis. In *Proceedings of ARIADNE Conference* 2003.

Polsani, P.R. (2003). Use and Abuse of Reusable Learning Objects. *Journal of Digital Information*, 3(4).

Sicilia, M.A. and García, E. (2003). On the concepts of usability and reusability of learning objects. *International Review of Research in Open and Distance Learning* 4(2).

Sicilia, M. A., Pagés, C., García, E. y Sánchez, S., Rius, A. (2004). Specifying semantic conformance profiles in reusable learning object metadata. In *Proceedings of the 5th International Conference on Information Technology Based Higher Education and Training (ITHET05)*, Istanbul, Turkey.

Sosteric, M., and Hesemeier, S. (2002). When is a Learning Object not an Object: A first step towards a theory of learning objects. *International Review of Research in Open and Distance Learning* 3(2).

Wiley, D.A. (editor) (2001). *The Instructional Use of Learning Objects*. Association for Educational Communications and Technology, Bloomington.

Notes

[1] Compliancy tools for the SCORM specification can be found at http://www.adlnet.org/.
[2] The IMS Simple Sequencing specification can be found at http://www.imsproject.org/.
[3] The MERLOT repository is at http://www.merlot.org/.
[4] The CAREO repository is at http://www.careo.org/, and the Learning Commons variant is at http://careo.ucalgary.ca/.

Figures

Figure 1. Screenshot of a peer-review page included in the MERLOT Learning Object Repository.
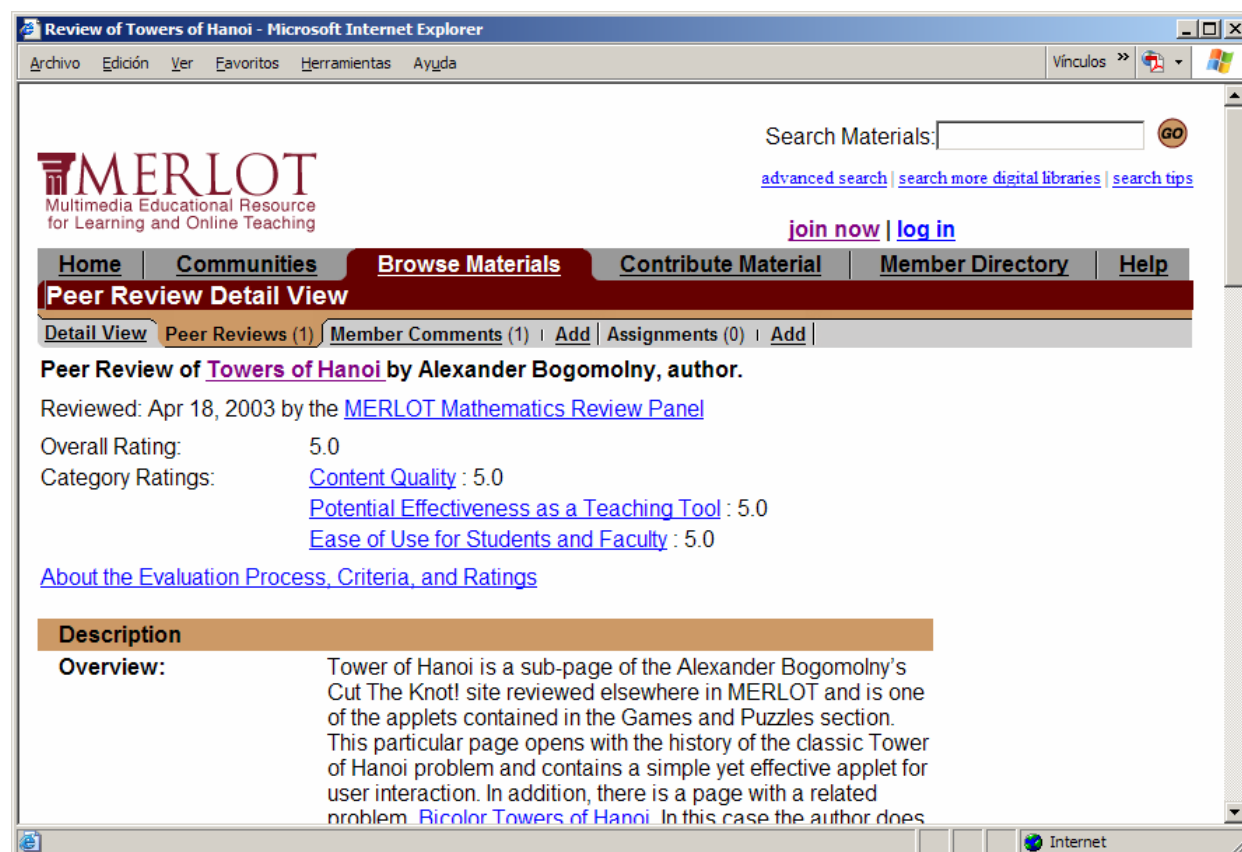
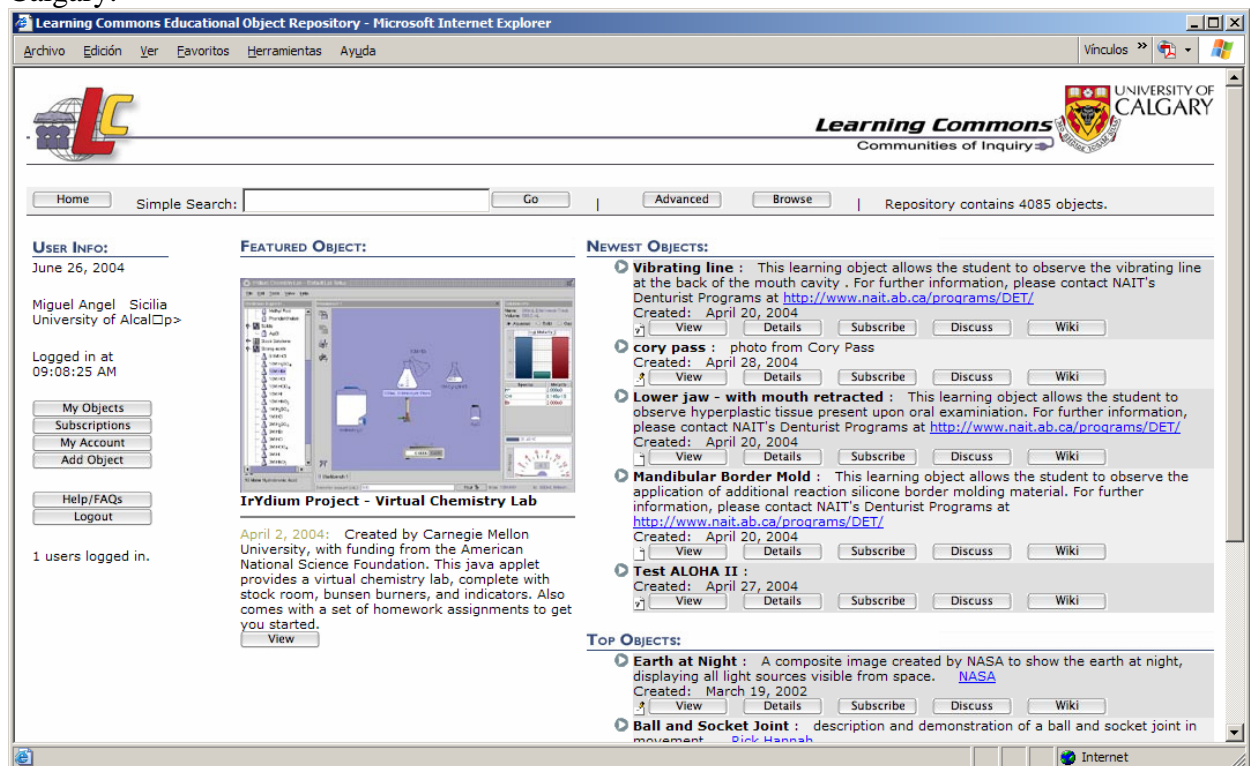Figure 2. Screenshot of the Learning Commons Learning Object Repository at University of Calgary.

Figure 3. Screenshot of the CAREO Learning Object Repository at University of Calgary: Metadata edition interface.