A Semantic Lifecycle Approach to Learning Object Repositories

Miguel-Angel Sicilia, Elena García-Barriocanal, Salvador Sánchez-Alonso
University of Alcalá
Ctra. Barcelona, km.33.6 – Alcalá de Henares, 28871 (SPAIN)
{msicilia, elena.garciab, salvador.sanchez}.uah.es

Jesús Soto
Pontifical University of Salamanca
Paseo Juan XXIII, 3 – Madrid, 28040 (SPAIN)
jesus.soto@upsam.net

Abstract

Learning Object repositories play a key role in the vision of reusable learning contents and learning designs, serving as providers for learning-oriented artefacts. Nevertheless, current metadata creation practices result in artefact collections that lack machine-understandable metadata, which seriously hampers opportunities for reuse. Semantic Web ontologies can be used to improve the quality of learning object metadata records, but they are not enough by themselves. In order to respond to requests by returning the adequate resources, the repository is required to be aware of the amount, type and quality of the metadata records it stores. In this paper, the design of a learning object repository approach to "semantic lifecycle" is described and illustrated through the concrete architecture of the prototype of the SLOR repository.

1. Introduction

The concept of *reusable learning object* [21, 13] has become the central notion of a new approach to education and organizational learning that emphasizes reusability as the key characteristic. This new approach may eventually result in mass customization [10] and a general improvement in quality and efficiency when designing learning experiences. Behind such approach lies the requirement for interchange of reusable learning objects (RLO), which has resulted in the concept of *Learning Object Repository* (LOR) as the architectural element responsible for the shared storage and delivery of learning objects. In consequence, a LOR must provide at least some facilities for creating and/or updating RLO metadata, which will be later used to provide querying services to *Learning Management Systems* (LMS) and other kind of tools. The role of a LOR is currently being described in detail in the IMS Digital Repository Interoperability (DRI) specification. A considerable number of LOR have been recently put into functioning, some including built-in support for quality assessment. For example, MERLOT [2] has established a peer-review system for learning objects, while CAREO[1] provides Wikis[2] as a mean to discuss or provide comments associated to learning objects. Nonetheless, existing studies on actual metadata creation [11, 4, 16] have evidenced that delegating to metadata contributors the decision on how to annotate learning objects may result in fragmentary and incomplete metadata records. The main problem is that this kind of metadata is hardly usable for software agents in automated or semi-automated scenarios. In fact, most metadata in current LO repositories are no more than an overall content identification and description, thus providing limited value from the viewpoint of delegating tasks to agents [9]. Surprisingly, metadata are actually being created in the form of natural language, unstructured texts, e.g. in the comments and peer reviews of MERLOT. This points out to the necessity of providing a more disciplined approach to learning object metadata edition oriented to machine-understandability, and connected to clear notions of value that provide a motivation to specify each piece of metadata. This is not in contradiction with adopting metadata standards, but instead a specialized and enhanced application of the basic common ground they provide. The key aspect emphasized in this paper is that of metadata as "function-enabler" [7]. In other words, the actual motivation –from the viewpoint of automation– for the creation of a particular metadata element lies in the value of the functions it enables. Although this is

[1] http://careo.ucalgary.ca/
[2] http://phpwiki.sourceforge.net/

self-evident in elements like *cost* or *title*, clearly understood as enabling functions related to respectively "purchase" and "search", the same does not apply to other elements. For example, the *Educational* section of LOM metadata asks users for describing a usage context which is rarely unique, since many learning objects can be used differently in disparate contexts. This and some preliminary models of reusability [18] call for making explicit the *Context* as a notion in metadata records, which in turn enables a function of "context matching". In addition, metadata elements require some notion of consistency. For example, the *Relationship* category in LOM should have an influence in the structure and properties of the related learning objects [14] to appropriately enable composition functions, but this is neither specified by standards nor enforced by common tools.

The idea of viewing metadata elements as enablers for specific functions can be realized in scenario-oriented terms. Under this view, learning objects are artefacts capable of (a) being used in some *scenarios* including, but not limited to, learning processes, and (b) integrating system-to-system processes like purchasing, selection, composition or exchange. Each scenario requires some metadata elements designated in some specific manners –i.e. the so-called *idioms* in [20]. This way, providing more and better metadata for a learning object broadens the set of scenarios in which it may be used, including also an objective (although partial) notion of reusability. Since the approach described is oriented to the construction of software that exploits metadata, a representation framework far beyond plain metadata records with natural language expressions is required. The use of ontologies in the context of Semantic Web technology [1] has been proposed in previous research as a solution to such requirements –see [19] for an overall view–. Ontologies can be used to provide the machine-oriented semantics required to achieve higher levels of automation and support to decision at the organizational and pedagogical levels.

It should also be noted that the evolution of the content and structure of the learning object may entail outdated metadata. This raises the need to distinguish between "semantic differential versions" from the rest. Such versions are those that entail that some or all of the metadata elements become invalidated. For example, updating a content piece with the explanation of a new concept will likely make inaccurate the classifications provided in the metadata. Scenarios, versions and metadata edition can all be integrated in a "Semantic Lifecycle" approach on top of the conventional lifecycle expressed by *Lifecycle* and *Meta-metadata* categories in LOM. This is the underlying philosophy to the design of LORs described in this paper, and then implemented in the prototype of the SLOR repository.

The rest of this paper is structured as follows. Section 2 describes the notion of semantic life cycle as a framework for a stepwise, scenario-oriented approach to LORs development. Then, the concrete architecture and life cycle approach adopted in the prototype of the SLOR repository are described as an illustration. Finally, conclusions and future research directions are provided in section 4.

2. Defining a "Semantic Life Cycle" for Learning Objects

Learning objects as digital entities have a lifecycle spanning from their creation to obsolescence which is intended to reflect the chain of versions and uses and their associated information. This is recognized in existing metadata standards, as evidenced by the fact that the IEEE LOM standard [8] provides a metadata category *LifeCycle* that describes versions, status and contributors. In addition, the LOM *Meta-metadata* category enables the description of the history of changes in the metadata record itself. Nevertheless, the creation of learning object metadata, especially in the case of metadata connected to ontologies, is a complex and time-consuming process that requires a stepwise approach. Figure 1 depicts the main states for such a stepwise approach, which involves four states:

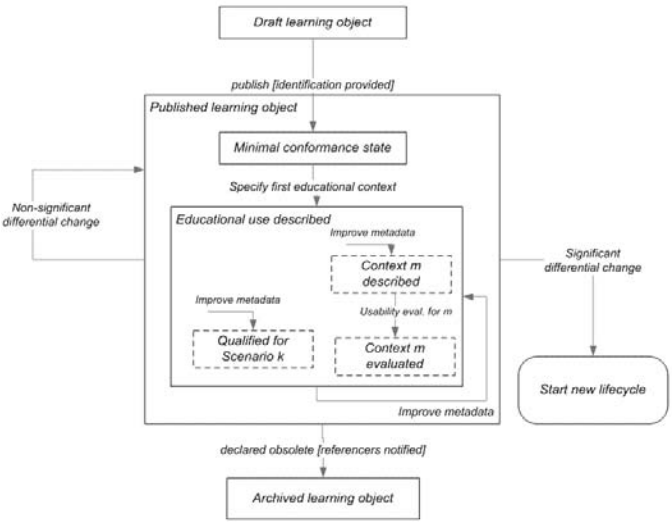

Figure 1. Creation of LO metadata

- *Draft*: learning objects in a "pre-publish" state that lacks the minimal necessary identification and descriptive metadata.

- *Minimal conformance*: objects capable of being referenced as digital entities, which is considered the minimal *availability* runtime commitment in [14].
- *Educational use described*: concrete educational contexts in which the learning object can be used have already been expressed as separate context descriptions. The use of differentiated sections (roughly equivalent to LOM *Educational* category) is consistent with the notion of profile by Downes [3], and fits with the idea of distributed context descriptions and with notions of reusability that are based on evaluating the usability of the learning object for concrete contexts [18].
- Finally, the object would eventually become obsolete, and could be deprecated after all the sources relying on it have been properly notified.

Herein, the notion of "educational context" is used to overcome the apparent inconsistency that could result from specifying different contexts as plain LOM metadata. As an illustration let us suppose that a learning object about Java exceptions is considered to be usable both in an academic setting as a support material, and in a corporate training setting as a test for basic certification. If the two *Educational* LOM metadata records for this description are used, the *5.2.Learning Resource Type* field should have the two values "self assessment" and "exam", while *5.6.Context* would hold the values "higher education" and "training". However, there is not a way to relate these two fields and the field *5.8.Difficulty* which may contain different values, as "medium" and "high", for example. In consequence, the quality or readiness of metadata should be judged on a context basis.

It should also be noted that, from the semantic lifecycle perspective, two kinds of changes exist. "Significant" changes entail that the metadata describing the object may have become invalid, which requires the inception of a new lifecycle instance. Moreover, such changes should result in the creation of a new learning object with a differentiated metadata record (even though it should trace back to the originating object through the appropriate information in the LOM *Relationship* category). Once the "minimal conformance state" is reached, we have an identifiable digital asset with essential descriptive data. From this point, another lifecycle dimension called "semantic conformance" starts. This new dimension is orthogonal to the one where contexts are described. The semantic conformance lifecycle represents the automatic classification process of the learning object into categories that make it qualified to engage in some specific and predefined semantic conformance profiles [20]. Such classifications are called "automatic" in the sense that they are defined concepts inside the underlying ontology of learning objects, i.e. the qualifying conditions are stated in description logics as implicit (non-primitive) classifications. One last aspect that should be tracked is that of evaluation. Usability evaluation, as a way of checking the effectiveness to attain stated learning goals [18] is an additional aspect that influences the way learning objects are considered inside the LOR. This is also an important source from which global reusability indicators may be obtained. Scenario-classification and usability evaluation are the two key aspects considered regarding the value of each object. The former represents a formal statement of completeness (i.e. presence of certain metadata description in some required form) and the latter complements it with an evaluation of the actual contents and instructional design for a particular context.

3. The SLOR Prototype Architecture

The SLOR prototype uses OWL as the underlying representation language. Even though its design is intended to respect existing guidelines for the RDF representation of LOM metadata [12], a considerable number of extensions have been required for metadata to fit in this scenario-oriented approach. In this paper, some of the prototype functionalities are sketched as an illustration of a semantic-enabled learning object repository. The architecture of the prototype is structured in three layers as shown in Figure 2:

- The persistence of ontological models is stored inside MySQL instances, using the built-in persistence capabilities of the *Jena* framework[3].
- The *Joseki* RDF server[4] is used to provide an intermediate layer that allows the distribution of servers with transparent persistent capabilities. This could be used to distribute persistence in different nodes, possibly with different database server technologies.
- The Web presentation layer is built with *Struts* technology[5] to obtain a Model-View-Controller separation from the Web logic. The use of properly modularized Struts actions provides better reuse opportunities for the cases in which domain-specific learning object repositories would be constructed having the code of the prototype as a basis. In addition, the ontology-aware capabilities of Jena are used for processing queries and updates, relying on the HTTP based update model of Joseki and the use of RDQL queries.

[3] http://jena.sourceforge.net/
[4] http://www.joseki.org/
[5] http://struts.apache.org/

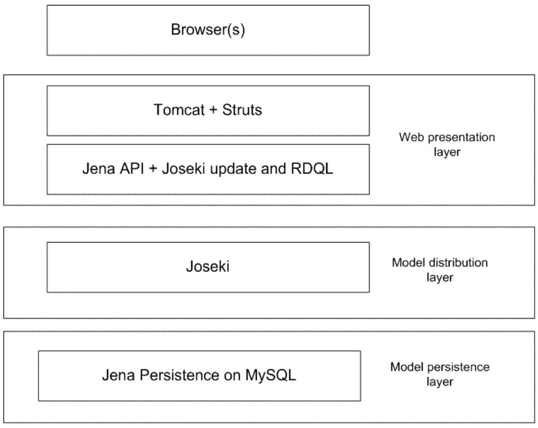

Figure 2. Main components of the prototype

Using Joseki in this version represents a restriction whenever the built-in reasoning capabilities in Jena are required. This is due to the fact that current information interchange protocols consist in HTTP-based transmissions of RDF trees, which produces massive data transfers to the Web Server when complex queries on the whole repository are invoked. The prototype is called SLOR and is intended as a proof of concept for the system. This is why well-known ontologies are used instead of those more specific to certain domains. Figure 3 shows the initial learning object creation facility that would lead to the state of "minimal conformance", where two elements are worth to be discussed:

- The LOM *Description* and *Coverage* categories allow for the standard language-string specifications prescribed in the standard, but also for ontology-based annotations that will be discussed later (prefixed in Figure 3 by the term "ontology" in parentheses).

- The learning object type is selected from among an ontology of types that extends the elementary classification in LOM. The rationale for this is that different learning object types could prescribe different properties and structure. The most typical example is that of learning objects of type *Questionnaire*, for which the detailed structure and metadata of specifications like IMS QTI should be applied. But even objects of different granularity or structure should be provided with different metadata elements that go beyond the scope of the "common set" provided by LOM.

As a general design criterion, natural language or enumerated labels inside metadata elements can be complemented with predicates that assert some descriptive or prescriptive aspect of the learning object. Such predications are shown in Figure 3 in the form `(ontology) predicate [Class]:instance`. This allows for the connection

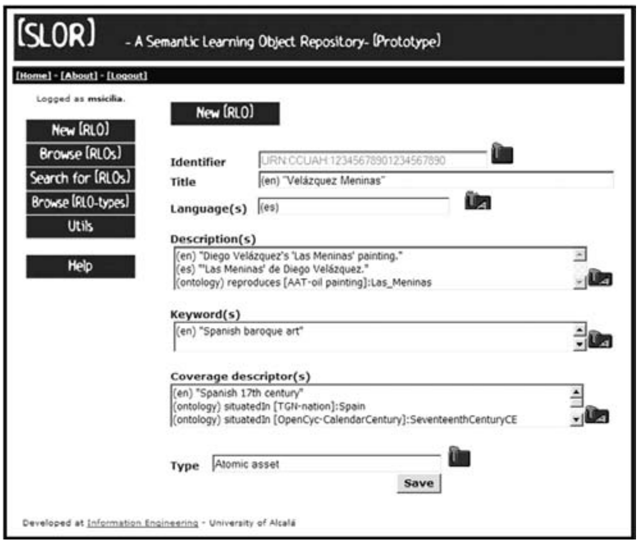

Figure 3. SLOR creation facility

of instances of `LearningObject` to instances of any available domain ontology, providing the benefit of having the objects represented as elements inside the ontology, and thus "visible" for reasoners or any other kind of inference support. Although currently used predicates are specific of this prototype, a common set of "learning object annotation predicates" could be agreed in the future for enhanced metadata interoperability. Learning object search is now implemented as an ontology-based seeking interface based on previous work [6]. The browser's role in this model is to allow that any metadata category in LOM (excluding lifecycle and meta-metadata, not related to educational purposes) can be used as top guiding criteria. This way, the ontology terms attached as descriptions would be either selected or not, what will result in a query expression consisting of a collection of ontology terms. The navigational affordances and the translation approach from the query to the retrieval of learning objects are again based on previous work [5]. However, queries here are by default interpreted on a contextual basis, i.e. all the requirements selected by the user should match the same learning context. Figure 4 shows an example of navigation. It should be noted that both semantic conformance profiles and a number of metadata elements in LOM can also be used as filters.

Figure 4 shows a query construction step in which several elements are included as requirements. These elements are terms (like "Art_Period" from AAT) or instances of terms (like "Spain") that characterize a particular aspect of the metadata, what is shown in the second column. All the underlined elements are actually links to additional information for the ontology element, and the "Example LO" functionality works like a partial search in the sense

Process	*Required metadata*	*Definitions required*	*LMS behaviour*
CMP-A	ID-1 Classifications	$ID1 \equiv LOM\_Id = 1$ $CMP - A \equiv ID1 \cap \exists competency.C$	Given a goal expressed as a set of arbitrary concepts $K = \{C_i\}$, find a set of instances of LearningObject so that $\forall C_i, \exists competency(lo, c_i)$ Compose learning objects to cover all the required concepts in $\{C\}$.
CMP-B	ID-1 Classifications	$CMP - B \equiv ID1 \cap \exists competency.C$ $CMP - A \supseteq CMP - B$	Compose learning objects to cover all the required concepts in $\{C\}$, including sub-concepts of concepts in $\{C\}$ connected through "part-of" relationships, i.e. the goal becomes $K' = \{ C_i \,\vert\, K \vee \; part-of \; (C_i, x_j \,)$ where $x_j \in K\}$
CMP-C	ID-1 Classifications Relationships	$CMP - C \equiv ID1 \cap \forall prerrequ.C$ $CMP - A \supseteq CMP - B$	Compose learning objects to cover all the required concepts in $\{C\}$, including sub-concepts of concepts in $\{C\}$ connected through "part-of" relationships, and also the declared prerrequisites, i.e. $K'' = \{ C_i \,\vert\, K' \vee \; prerreq \; (C_i, x_j \,)$ where $x_j \in K'\}$

Table 1. Learning object compliancy with respect to profiles.

that it retrieves from the knowledge base a learning object that fulfils that requirement. This mode of fulfilment is an adaptation of that described in [5], in which resources linked by relations to the selected terms or instances are retrieved, with a relevance that depends in the form they are linked. For example, the selection of "Spain" in coverage would pre-select all the instances connected to that instance with any arbitrary relationship as part of the "coverage" section like "situated in". In addition, the "Suggested relation" section shows arbitrary terms or instances connected to the ones selected in the requirements area, and the associated elements can be added as requirements. The "refine one level" link traverses one level through the primitive subsumption hierarchy as a way to narrow down the search criteria. This navigational structure allows the use of the ontology as a driver for guided search.

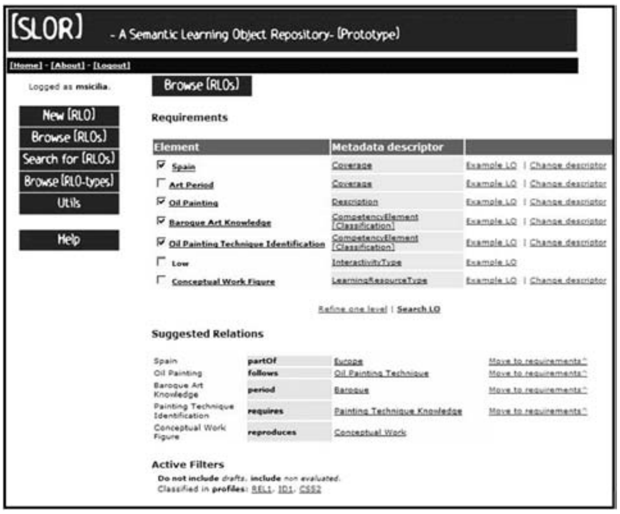

Figure 4. An example of navigation

Classifications in conformance profiles are expressed as regular logic ontology constraints, except for some descriptions like external tests (e.g. accessibility conformance) that are defined as primitive classes. For example, the three simple conformance profiles sketched in [20] can be translated as shown in Table 3, provided that the expected outcomes of the learning objects are expressed in terms of competency definitions [17], namely `Knowledge`, `Abilities` and `Attitudes` that are required as components of a `CompetencyDefinition` (expressed generically as *C* in what follows). Classifications in CMP-A and CMP-B should refer to a term inside a shared, open ontology. In CMP-C, all the prerequisite relationships of the learning object must be defined as well. The definitions shown in Table 3 illustrate the notion of learning object compliancy with respect to LMS behaviours (i.e. profiles of functionality). For example, the requirements of $CMP - B$ and $CMP - A$, as expressed in description logics, are the same, but the LMS behaviours informally expressed in predicate logic are not. A quality characteristic that should be taken into account in $CMP - C$ is that all the prerequisite relationships are actually declared. In other words, $CMP - C$ somewhat follows a "closed world" assumption in considering that only the prerequisites declared for a given learning object exist. The concept of conformance profile embodies an approach to precise definitions of LMS behaviour based on well-defined requisites on metadata. Such profiles can be composed and extended through logical means, resulting in standardized, predictable complex LMS behaviour. In this context, the LOR plays the role of classifying objects to allow for advanced filtering targeted at functional interoperability.

4. Conclusions and Future Research Directions

The design of LORs requires a consideration of the status and degree of description of learning object metadata records, which determines the behavior of the LOR when fulfilling requests. Without such consideration, queries issued to a LOR may result in learning objects that are poorly or inadequately described for the automation needs of learning systems. A basic "semantic life cycle" has been defined, which may serve as a basis to define various "semantic conformance levels" that are enabled as a result of the specification of certain metadata elements in previously arranged ways. This makes the repository useful for automation of search tasks in which an automated agent or Web Service looks for learning objects with certain metadata characteristics, as illustrated in contract-based architectures [15]. Future work should investigate which are the semantic conformance profiles that are useful for concrete "intelligent" processing scenarios, extending the work initiated in [20]. In addition, further study and consensus reaching is required in a concrete and commonly agreed technique for creating LOM-based annotation as linked to ontologies. This paper has provided some example sentences, but without a shared idiom for such annotation, achieving actual semantic interoperability across repositories is difficult.

5. Acknowledgements

The research reported in this paper was partially supported by University of Alcalá project UAH-PI2005-070 and Spanish Ministry of Education project TSI2004-21263-E.

References

[1] T. Berners-Lee, J. Hendler, and O. Lassila. The semantic web. *Scientific American*, 284(5):34–43, May 2001.

[2] R. Cafolla. Project Merlot: Bringing peer review to web-based educational resources. In *Proceedings of the USA Society for Information Technology and Teacher Education Int. Conference*, volume 2002, pages 614–618. AACE, 2002.

[3] S. Downes. Resource profiles. *Journal of Interactive Media in Education*, 2004(5), May 2004.

[4] N. Friesen. International LOM survey: Report. Technical report, Digital Library of Information Science and Technology, July 2004.

[5] E. Garca and M. Sicilia. Designing ontology-based interactive information retrieval interfaces. In *Proceedings of the Workshop on Human Computer Interface for Semantic Web and Web Applications*, pages 152–165, 2003.

[6] E. Garca and M. Sicilia. User interface tactics in ontology-based information seeking. *Psychology Journal*, (3):242–255, 2003.

[7] J. Greenberg. *Encyclopedia of Library and Information Science*, chapter Metadata and the World-Wide-Web, pages 1876–1888. Marcel Dekker Inc., New York, second edition, 2003.

[8] LTSC. *IEEE Standard for Learning Object Metadata, 1484.12.1-2002*. IEEE LTSC, 2002.

[9] M. Lytras, A. Pouloudi, and A. Poulymenakou. Dynamic e-learning setting through advanced semantics: The value justification of a knowledge management oriented metadata schema. *International Journal of e-Learning*, 1(4):49–61, 2002.

[10] M. Martinez. Mass customization: Designing for successful learning. *International Journal of Educational Technology*, 2(2), 2001.

[11] J. Najjar, S. Ternier, and E. Duval. The actual use of metadata in ariadne: an empirical analysis. In *Proceedings of the third Ariadne Conference*, pages 1–6, 2003.

[12] M. Nilsson, M. Palmer, and J. Brase. The LOM RDF binding - principles and implementation. In *Proceedings of the 3rd Annual Ariadne Conference*, pages 291–294, 2003.

[13] P. R. Polsani. Use and abuse of reusable learning objects. *Journal of Digital Information*, 3(4), February 2003.

[14] S. Sánchez-Alonso and M. A. Sicilia. Relationships and commitments in learning object metadata. In *Proceedings of ITHET 2004 - 5th Int. Conference on Information Technology Based Higher Education and Training*. IEEE, 2004.

[15] S. Sánchez-Alonso, M. A. Sicilia, and J. Lpez-Cobos. Design contract-based selection and composition of learning objects. In *Proceedings of the 6th Int. Simposium on Computers in Education - SIIE 2004*, 2004.

[16] M. Sicilia, E. Garca-Barriocanal, J. J. Martínez, C. Pagés, and J. M. Gutiérrez. On the evaluation of completeness of learning object metadata in open repositories. In *Proceedings of the 2nd Int. Conference on Multimedia and Information and Communication Technologies in Education*, pages 1760–1764, 2003.

[17] M. A. Sicilia. Ontology-based competency management: Infrastructures for the knowledge-intensive learning organization. In M. Lytras and A. Naeve, editors, *Intelligent learning infrastructures in knowledge intensive organizations: A semantic Web perspective*, pages 302–325. IDEA Group Inc., 2005.

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

[19] M. A. Sicilia and E. García. On the convergence of formal ontologies and standardized e-learning. *Journal of Distance Education Technologies*, 2(4), 2004.

[20] M. A. Sicilia, C. Pagés, E. García, S. Sánchez-Alonso, and A. Rius. Specifying semantic conformance profiles in reusable learning object metadata. In *Proceedings of the 5th Int. Conference on Information Technology Based Higher Education and Training*. IEEE, 2004.

[21] D. A. Wiley. Connecting learning objects to instructional design theory: A definition, a metaphor and a taxonomy. In *The instructional use of learning objects*, pages 3–24. Bloomington, Indiana, 2002.