Interoperability issues between learning object repositories and metadata harvesters

Ricard de la Vega[1], Jordi Conesa[2], Julià Minguillón[2]

[1] Centre de Supercomputació de Catalunya, Barcelona, Spain,
rdelavega@cesca.cat
[2] Universitat Oberta de Catalunya, Barcelona, Spain,
{jconesac, jminguillona}@uoc.edu

Abstract. In this paper we describe an open learning object repository on Statistics based on DSpace which contains true learning objects, that is, exercises, equations, data sets, etc. This repository is part of a large project intended to promote the use of learning object repositories as part of the learning process in virtual learning environments. This involves the creation of a new user interface that provides users with additional services such as resource rating, commenting and so. Both aspects make traditional metadata schemes such as Dublin Core to be inadequate, as there are resources with no title or author, for instance, as those fields are not used by learners to browse and search for learning resources in the repository. Therefore, exporting OAI-PMH compliant records using OAI-DC is not possible, thus limiting the visibility of the learning objects in the repository outside the institution. We propose an architecture based on ontologies and the use of extended metadata records for both storing and refactoring such descriptions.

1. Introduction

Nowadays, most educational institutions use learning content management systems in order to provide learners with additional support in their learning process. Books have been the traditional content used in such institutions, but new learning theories more oriented towards competence acquisition and development rather than content consumption are changing this, promoting the concept of activity in front of the concept of content. In this sense, learning objects are small pieces of content that are supposed to help learners to achieve a specific learning goal, as part of the activities performed during the learning process (Wiley, 2000). Ideally, learning objects are self-contained, that is, can be taken independently; reusable, that is, context independent; can be aggregated and are tagged with metadata.

Obviously, digital repositories are a way to organize learning objects (and their parts that can be processed separately) in collections, although there are several specific issues that must be firstly addressed. For example, an exercise (which is basically a text defining a problem and, optionally, its solution, another text which may include references to the use of software or tables with data, for instance) is a typical learning object. But, differently to classical items in a collection of a digital repository, exercises may have neither a title nor even an author, the two main fields used for finding a book. Other typical learning objects can be data sets, mathematical proofs, equations, simulations, and so. Usually, learners search through these kinds of resources not by title or author, but by keyword or, even better, using a hierarchical taxonomy specially designed. Therefore, it becomes necessary to rethink the traditional way of describing learning resources, using criteria related to the learning process but maintaining a minimum description for archiving purposes. The use of metadata standards for describing learning objects such as IEEE LOM (instead of Dublin Core) is also a possibility, although there is not a single solution to be found superior than the rest. On the contrary, some authors point out that several standards and specifications will (or should) converge in a near future for improving the description of learning resources (Currier, 2008).

2. Learning object repository on Statistics

This work is part of a large project that takes place in a higher education institution, the Open University of Catalonia (UOC), with the aim of promoting the development and acquisition of competencies through the use of learning object repositories. The UOC is an online distance university with more than 40,000 students and more than 2,500 staff including instructional designers, teachers, tutors, academic and technical staff. The UOC uses a virtual campus as an integrated e-learning environment that allows students to pursue their studies purely online. We intend to design and develop a learning object repository that is not only useful as a mere repository but, at the same time, its use becomes an active element of the learning process, so students using the repository will achieve a set of competences.

As part of a pilot experience, an open learning object repository with resources on Statistics has been built using DSpace as platform. This repository, named OER, has been designed with the aim of providing learners with a comprehensive vision of the whole knowledge domain of Statistics, trying to make of browsing and searching a true learning experience (Ferran et al., 2009). In order to do so, learning objects in the repository have been tagged according to the following minimum criteria: every resource is an element part of several taxonomies (one for describing the Statistics domain of knowledge, another for describing the kind of resource and a third one for identifying the course or degree the resource was created for), and it is described by one or more keywords. Therefore, according to their nature, learning objects may have or not title, author, creation date, etc., so they cannot be accessed by classical retrieval mechanisms used in digital libraries or repositories. In fact, DSpace had to be customized to change the basic fields used for searching and browsing, as well as the workflows related to the process of adding new resources to the repository.

On the other hand, this repository is also used as a starting point for creating a new visual user interface, so users will be able to browse and search for learning resources without using the interface provided by DSpace. The main goal of this project is to promote the integration of learning resources into the learning process, by allowing learners to tag, rate, make comments, favorite and ask questions on learning objects which they consider useful, thus creating a social network among learning resources and learners according to their interests (Córcoles et al., 2009). These services allow learners to use their own words to describe learning objects, creating folksonomies that need to be further analyzed in order to ensure and improve a certain level of quality of the metadata (Kim et al., 2010). The main problem here is that current standards and specifications do not provide full support to the information generated during the use of a given learning object, although such information can be stored using RDF or any other XML-based schema.

3. Interoperability with metadata harvesters

On the other hand, in order to give more visibility to these learning objects and encourage its reuse beyond the university itself, the learning object repository can be exposed to external metadata harvesters using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). Then, learning objects can be integrated by other repositories (or harvesters) with a greater scope of potential users, such as MDX[1] (Learning Materials Online). MDX is a cooperative repository that contains digital materials and resources resulting from teaching activities carried out in member universities, which currently are 10 (Huguet et al., 2007). This metadata harvester also has been built using DSpace as platform and with an Old Dominion University OAI-PMH plug-in. Currently now, MDX does not harvest metadata from OER, so the latter is virtually invisible to the whole academic community.

[1] http://www.mdx.cat

Interoperability between the learning objects repository and the OAI-PMH metadata harvester is based in the 15 fields of the non-qualified Dublin Core (nqDC) schema, which include title, author, type, etc. This schema is worldwide used by metadata providers, such as institutional or thematic repositories, on the one hand, and for metadata services, on the other. Nevertheless, with research repositories and harvesters (thesis, eprints, etc.), the use of nqDC brings forward semantic problems because, for example, there is only one "date" element, without distinguishing between "created date" and "issued date". In the case of learning object repositories, as mentioned before, there is the difficulty of completing a minimum description for archiving purposes, such as adding a title element.

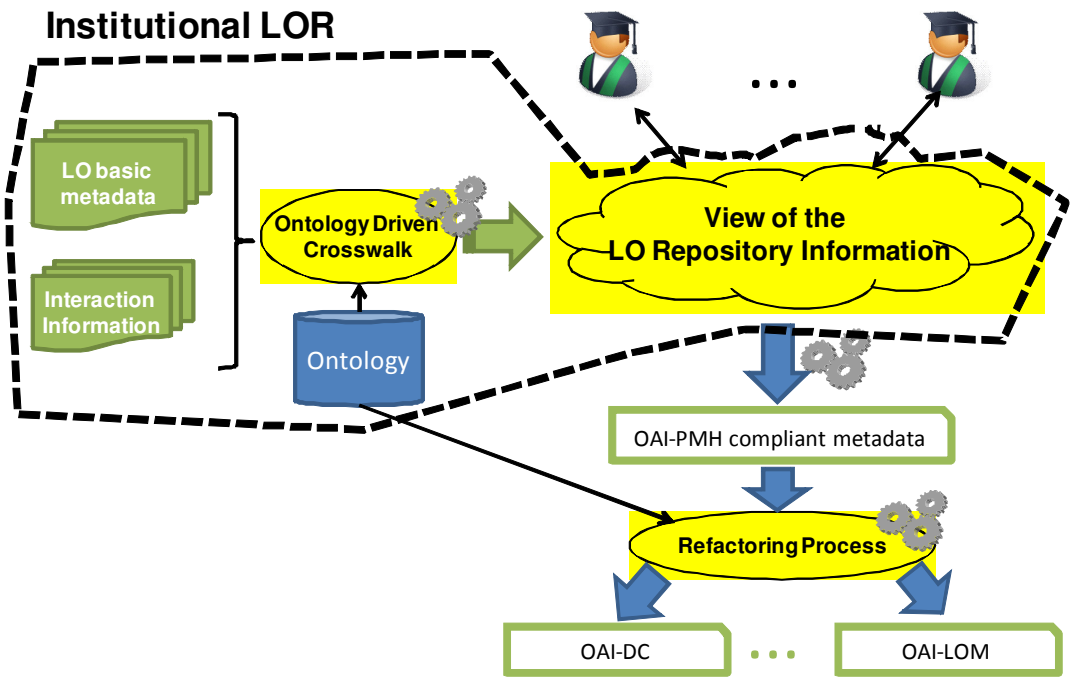

Figure 1. System architecture of the Learning Object Repository.

4. Towards a semantic description of learning objects

Therefore, although learning object repositories may use different metadata schemes created from a educational perspective, it is important to make them able to export metadata records in a "standard" way, in order to become more visible. In order to solve this problem we propose a prototype (see Figure 1) that extends regulars LOR in the following aspects:

1. It provides a XML-based scheme for storing the information generated during the interaction between learners and the learning objects stored in the repository. This information is needed by the user interface in order to recommend the most appropriate learning resources according to learner profile and previous usage history.

2. It uses an ontology for providing semantic support to partially incomplete descriptions based on keywords and taxonomies. For each learning object type, the ontology will define the required metadata (LO Basic Metadata) and its exact semantics. For example, in the case of exams, the ontology will specify that the title attribute is not necessary and that the meaning of date is when the exam was done, instead of when the document was created. The ontology will also define some derivation rules in order to infer new metadata of a learning object from its basic metadata, such as the creation of the title of an exam as the concatenation of the word "Exam" with its semester and its subject, or

defining as extra keywords the name of the terms of the taxonomies related to the LO and their supertypes. Therefore, the ontology will be used to expand the basic information of the repository with derived information. The union of the basic and derived information will compose the view the users will have of the institutional repository.

3. It provides an extended OAI-PMH compliant metadata format for such scheme, in order to make it available through *ListMetadataFormats*. Even though this repository may have some metadata that is inexistent within other standards, it would be convenient to have a metadata schema that allows providing all the available information thru an OAI service, in order to provide better interoperability (within different repositories of the same institution for example or with similar repositories in other institutions).

4. It provides an OAI-DC compliant metadata format for any record in the repository by means of a refactoring process. The refactoring process will use the information of the ontology and a set of rules, different for each type of learning object, in order to create views of the institution metadata according to different formats or standards of representation (IEEE LOM, Dublin Core…).

Acknowledgements. This work has been partially supported by Spanish Government funded projects PERSONAL(ONTO) ref. TIN2006-15107-C02 and E-MATH++ ref. EA2008-0151.

References

Córcoles, C., Minguillón, J. and Lamb, B. (2009) “Building an open social learning community around a Dspace repository on Statistics”, In: Proceedings of the 4th International Conference on Open Repositories (posters), Online Version, http://hdl.handle.net/1853/28480

Currier, S. (2008), “Metadata for learning resources: an update on standards activity for 2008”, ARIADNE, Vol. 55. Online Version, http://www.ariadne.ac.uk/issue55/currier/

Ferran, N., Guerrero-Roldán, A., Mor, E., and Minguillón, J. (2009) “User Centered Design of a Learning Object Repository”, In: Proceedings of the 1st international Conference on Human Centered Design: Held As Part of HCI international 2009 (San Diego, CA, July 19 - 24, 2009). M. Kurosu, Ed. Lecture Notes in Computer Science, vol. 5619, pp. 679-688.

Huguet, M., Anglada-i-de-Ferrer, Lluís M. and de-la-Vega, R. (2007) "Catalan Policies and Experiences on Cooperative Repositories", In: 1st Iberian Grid Infrastructure Conference Proceedings (IBERGRID). Santiago de Compostela, Spain. May 14-16, 2007. Santiago de Compostela: Centro de Supercomputación de Galícia. p. 63-75. Online Version, http://hdl.handle.net/2072/4083

Kim, H.-L., Decker, S. and Breslin J. G. (2010) “Representing and sharing folksonomies with semantics”, Journal of Information Science, 36 (1), pp. 57–72.

Sicilia, M.A. and García-Barriocanal, E. (2005) “On the Convergence of Formal Ontologies and standardized e-Learning”, Journal of Distance Education Technologies, 3(2), pp. 12-28.

Wiley, David A. (2000), "Connecting Learning Objects to Instructional Design Theory: A Definition, A Metaphor, and A Taxonomy", in Wiley, David A. (DOC), The Instructional Use of Learning Objects. Online Version, http://reusability.org/read/chapters/wiley.doc