Navigating learning resources through linked data: a preliminary report on the re-design of Organic.Edunet

Miguel-Angel Sicilia[1], Hannes Ebner[2], Salvador Sánchez-Alonso[1], Fernando Álvarez, Alberto Abián, Elena García-Barriocanal

[1] Information Engineering Research Unit, Computer Science Department, University of Alcalá, Polytechnic Building, Ctra. De Barcelona km. 33.6
28871 Alcalá de Henares (Madrid), Spain
{msicilia, salvador.sanchez, fernandol.alvarez, alberto.abian, elena.garciab}@uah.es
[2] School of Computer Science and Communication, Royal Institute of Technology, Lindstedtsvägen 3, 10044 Stockholm, Sweden
hebner@kth.se

Abstract. Learning objects repositories have grown and matured in the last years, being currently a cornerstone for open education. Several current systems are offering metadata openly through mainstream harvesting protocols or providing standardized query interfaces. Also, the use of standardized vocabularies or ontologies is becoming more common to provide a degree of semantic interoperability. However, learning object metadata is typically not linked across repositories, and it is not providing a way to navigate by using other sources of data available on the Web. The linked open data (LOD) approach provides the framework for the evolution of learning object repositories into a more flexible system of sharing learning resource metadata. This paper describes how linked data has been integrated in the design and re-design of the export mechanisms of Organic.Edunet, a federation of learning repositories in the domain of organic agriculture that uses an RDF store and several ontologies to browse and search resources. The paper focuses on how the existing search and semantic browsing mechanisms can benefit from the use of LOD across repositories.

Keywords: learning objects, repositories, linked data, ontologies, Organic.Edunet

1 Introduction

Learning object repositories have grown and matured in the last years, being considered currently an essential component for e-learning (Ochoa and Duval, 2009). Each of these repositories provide some particular interaction style for searching, browsing and navigating across learning resources, typically including text and metadata search along with some form of topical navigation. Some repositories provide also capabilities that extend its reach to resources or metadata distributed across different repositories (Klemke et al., 2010). This is currently achieved mainly by two mechanisms: harvesting and distributed search. In federated systems using harvesting, the metadata from a number of repositories is collected into a central location and regularly updated. Protocols like OAI-PMH are typically used for that process (van de Sompel et al., 2004). In distributed search, a query is distributed to several repositories and the query results are put together and eventually ranked according to some criteria. The Simple Query Interface (SQI) is an example of a protocol enabling that kind of arrangement (Simon et al., 2005).

While harvesting and distributed search have proved effective for aggregating repositories for the purpose of search, they do not allow actual browsing and navigation across different systems. In the case of systems based on harvesting, the navigation is still local to the centralized metadata store, and thus restricted to the repositories that are inside the federation. In the case of distributed search, browsing is also limited as only metadata records of query results are provided, and not links to categorizations of resources or to other resources not in the same repository, so cross-repository navigation is not allowed. The common problem with these systems is that they restrict navigation either by using a concrete request/response format or by restricting to the metadata already harvested in a central location, which is in most cases not providing navigation to other resources in its original systems. A distributed browsing and navigation system across repositories would require a way to expose metadata or query results that provides links to permanently identifiable resources along with links to elements of terminologies or taxonomies that are able to mediate navigation to other systems.

The collection of practices and recommendations for exposing linked open data (Bizer, Heath and Berners-Lee, 2009) are providing the required infrastructure for allowing that distributed navigation possible by exposing resource information in a common RDF format and enabling navigation through linked open data (LOD) to other systems. However, the deployment of the infrastructure for realizing that concept requires a re-conceptualization of the way repositories are commonly offering metadata and search facilities, along with the development of specialized browsers that are capable of traversing the Web of linked data. This paper provides the ideas and initial design elements for the implementation of such approach, using the Organic.Edunet repository[1] as an example. The research objective of the present work is laying out the foundations and requirements for exposing LOD in learning object repositories that enable cross-repository search and navigation. In consequence, the scope of the proposals presented focuses on how to give support to these scenarios and it is not addressing other infrastructural aspects (e.g. bootstrapping or the practicalities of migrating/extending current repositories).

The rest of this paper is structured as follows. Section 2 briefly describes the requirements from the data provider and consumer perspectives. Then, Section 3 describes the main decisions in the way IEEE LOM metadata is exported as linked data. Section 4 describes how the availability of generalized linked data following the conventions exposed could be exploited for browsing across repositories. Finally, conclusions and outlook are provided in Section 5.

1 http://portal.organic-edunet.eu/

2 Requirements

Data provider requirements are no essentially different from those of exposing linked data in other domains in which some metadata schemas are already in place. The most widespread specific standard for providing descriptive metadata of learning objects is IEEE LOM (IEEE, 2002). There are a number of systems providing ontologies or RDF mappings of IEEE LOM, however they are not standardized. Nilsson et al. (2008) published a mapping of IEEE LOM into the Dublin Core Abstract Model that is discussed within the joint DCMI/IEEE LTSC Taskforce[2]. This mapping was also used within Organic.Edunet to annotate resources using the repository tool Confolio (Ebner, 2009). The mapping sets the basis for exposing both Dublin Core and IEEE LOM metadata by using a shared format.

Further, there is a need to consider the particular nature of learning objects and the fact that a LOM metadata record is describing two very distinct types of information: (a) *factual* information on the resource, e.g. identifier, language, technical description, and (b) *contextual* information in which the resource is described for a particular educational context, i.e. difficulty is relying on some particular target user population, and it is possible that another metadata record for the same resource is having a different value as it is considering a different target. Such a contextualisation is supported by the Confolio tool and its backend (Ebner, 2008). One common use case is to harvest metadata from generic repositories and to provide context by adding educational metadata. Separate named graphs with relations between them are used to express metadata (in our case educational metadata) and meta-metadata such as provenance information and access control.

The general availability of learning object metadata as LOD across repositories would enable a number of applications that are beyond the reach of current systems. From the viewpoint of filtering and selection, it would be possible to build systems that mediate search through terminologies or ontologies representing educational systems and competencies required for particular profiles. Particularly, it would be possible to navigate through these representations to browse and search, complementing current distributed query and harvesting mechanisms. Another key application would be that of aggregating metadata descriptions for the same learning object, creating new opportunities for filtering based on quality, and having a way to measure the reusability of the resource across different educational contexts (Sicilia and García-Barriocanal, 2003).

In order to enable cross-repository search and browsing through LOD, there is a need for some conventions on exposing resource metadata and also some descriptions of the repository themselves. These are described in what follows.

3 Exporting IEEE LOM as linked data

Organic.Edunet features two different approaches for storing metadata: (1) the distributed repository tools using a triple-store with an abstraction of named graphs

2 http://dublincore.org/educationwiki/DCMIIEEELTSCTaskforce

and an implementation of the DCMI/IEEE draft using DCAM (Ebner, 2009), and (2) on the federated portal side, an OWL-based repository based on HP Jena with a relational datastore backend, using an OWL representation of IEEE LOM combined with several ontologies, including one with an specific terminology for organic agriculture and agroecology (Sánchez-Alonso, 2009). The repository tools within the federation expose the metadata following the linked data approach and according to the abovementioned DCMI/IEEE draft. In a different way, the process of exporting linked data through the portal required the construction of a new module that uses the existing SPARQL endpoint to translate the native RDFOWL representation to the RDF export described below, combined with a module creating additional RDF links whenever viable in an automated way.

Identifiers should be representing at least two types of entities: the object themselves (i.e. the Web contents) and the metadata records, as the same objects may be described several times in different repositories (Downes, 2004). In our case, all the resources are external to the portal and identified by URIs, so we are exposing only the second kind of objects. The resource identifiers themselves in IEEE LOM are specified in element `General.Identifier`. In our case, the URI is the object of all the resource statements about the resource, with the exception of the `Meta-metadata` category, that will have as object the document requested. In consequence, a fragment expressed in Turtle syntax for the description of an example resource would be as follows (namespace declarations are omitted for brevity):

```
# Meta-metadata information
<>
    lom-meta:schema lom-meta:LOMv1.0
    lom-meta:language < http://dbpedia.org/page/English_language>;
    …
# The description
< http://confolio.vm.grnet.gr/scam/6/resource/271>
    lom-gen:language < http://dbpedia.org/page/English_language>;
    lom-gen:keyword [
     a rdf:Alt;
     rdf:_1 "energy consumption"@en ;
     rdf:_2 "energiforbruk"@no ;
    ];
    lom-gen:structure lom-gen:atomic;
    lom-gen:coverage <http://dbpedia.org/page/Austria>;
    lom-edu:interactivityType lom-edu:expositive;
    lom-life:editor [
         lom-life:contributor [
               foaf:name "Petros Lameras"
               foaf:  workplaceHomepage <http://www.ea.gr>
               …
           ]
    ]
    lom-life:validator [
    …
```

There are several cases in which links to other linked data sites can be provided, including `language` of the metadata record and of the resource, which can be mapped to DBPedia resources. The lifecycle of the resources in Organic.Edunet requires metadata on the content provider (editor) and on the party that underwent quality assessment (validator). Both are using internally vCards, however in the exporting through linked data, these are converted into FOAF sentences linking to organizations where available. Another interesting case is `lom-gen:coverage` in which spatial coverages are linked to the DBPedia, and can also be linked to other LD datasets representing geopolitical information. In general, the vocabulary values used (e.g. lom-edu:expositive) are represented using the (namespace, value) construct proposed as a general mechanism in Nilsson et al. (2003).

Linking to ontology/terminology elements is achieved by exposing the Organic.Edunet ontology as linked data by using the following convention: <http://www.organic-edunet.eu/ont/*element*>. Where *term* is one of the terms, instances or properties in the ontology. The RDF description enables the navigation through the is-a hierarchy but also laterally by following predicates. For example:

```
< http://www.organic-edunet.eu/ont/biological-fertilizer>
 oe:is-made-from
   < http://www.organic-edunet.eu/ont/plant-origin-processed-product>;
```

A special case is that of element Classification which is used for various purposes in Organic.Edunet. IEEE LOM provides a vocabulary of purposes for this field, but these have been extended in the project with additional ones that were selected to provide a better description of classifications. The following fragment shows an example of IEEE LOM and extended classification purposes.

```
oe:details <http://www.organic-edunet.eu/ont/soil-fertility>
```

The oe:detail property is an rdfs:subPropertyOf of the vocabulary property lom-cls:educationalObjective. This is one of the IEEE LOM purpose values refined currently in the portal, a similar approach could be used for others as for example, prerrequisites.

The abovementioned conventions are the basis for exposing metadata in LOD format, including links to arbitrary terminologies. An additional step is that of providing links to other resources. This is implemented by extending the meanings of the IEEE LOM Relation.Kind element (which has a value space based on Dublin Core). As there are no data for this element in Organic.Edunet, descriptions are extracted from other parts of the portal. Currently and for demonstration purposes, links based on co-ocurrence of user-created tags are automatically added.

4 Browsing across repositories by using IEEE LOM

Architecting a browser that works over repositories publishing linked data following the abovedescribed conventions requires components to start the search process and to traverse links across metadata to reach related resources. We limit our discussion to reaching resources that are related through common reference to entities or concepts in terminologies that are also published as linked data. Organic.Edunet features a form of "berrypicking" search as described by Bates (1989). The existing implementation was based on the traversal of ontology relations, namely subclass for the main navigation and arbitrary properties for an alternative traversal. The selection one or several terms is done by navigation, iteratively and search processes are done on the central metadata repository that harvests regularly the provider systems. A REST-based interface is invoked from a Flash-based visualization, using JSON for the transfer of data for the rendering of the traversal tree. The re-design to exploit linked data consumption entails two main aspects:

1. Changing the interfaces serving terms during navigation to provide ontology terms across distributed linked terminologies. This allows the visualization to render any kind of relationship that is provided as backlinks in the description of the terms.
2. Changing the invocation from a local search to a distributed one. This can do in two flavors: (i) using text descriptions to go through conventional distributed query systems and (ii) selecting repositories or aggregators depending on the self-declaration of the terminologies used by repositories.

In both cases, repositories need to expose some form of standardized query interface as the SQI. The second approach would requires knowing which terminologies are in use by each repository, so that RDF links as applicable in `Classification` are used for the selection. A kind of self-declaration from repositories may follow the linked data approach by exposing information about themselves using linked data. An example is as follows:

```
< http://www.organic-edunet.eu/lor>
     dc:requires <http://www.organic-edunet.eu/ont/>;
     dc:requires <http://aims.fao.org/aos/agrovoc/>;
      ...
```

The W3C VoID[3] vocabulary for expressing metadata about RDF datasets can be used for the same purpose, concretely using the `void:vocabulary` property. Also, a property similar to VoID `sparqlEndpoint` could be used to provide the SQI endpoint of the repository. While many repositories are not using any kind of controlled vocabulary there are some relevant cases starting to use them as the LRE Thesaurus[4], the Organic.Edunet ontology or the different classification systems used in MACE[5]. In other cases, they can be generated automatically from metadata, as has been experimentally tested with AGROVOC in Organic.Edunet by using KEA[6] keyphrase extraction mechanisms. As an example, let's consider the current semantic navigational interface in Organic.Edunet showed in Figure 1. The OE ontology is providing mappings to terms in the AGROVOC thesaurus via `owl:sameAs`, which has very recently started to be exposed as linked data[7]. This enables the navigation to move to AGROVOC for the collection of query terms. Once the user has finished the query formulation, the terms selected would be sent via SQI to the underlying repository (in this case Organic.Edunet) or to any other that is exposing their description. The `requires` statements serve the role of selecting the more appropriate repositories according to terminologies used. Repositories can be discovered by crawling or by trying with the prefixes of URIs of resources included in the `relation` statements of records obtained.

The results snippets in Figure 1 would then potentially correspond to resources in several repositories without the need for central harvesting or registries of repositories. Also, relations between learning objects and to other linked data sets (e.g. as in `coverage`) can be exploited for query expansion.

---

[3] http://www.w3.org/TR/void/
[4] http://lre.eun.org/node/87
[5] http://portal.mace-project.eu/
[6] http://www.nzdl.org/Kea/
[7] http://aims.fao.org/website/Linked-Open-Data/sub

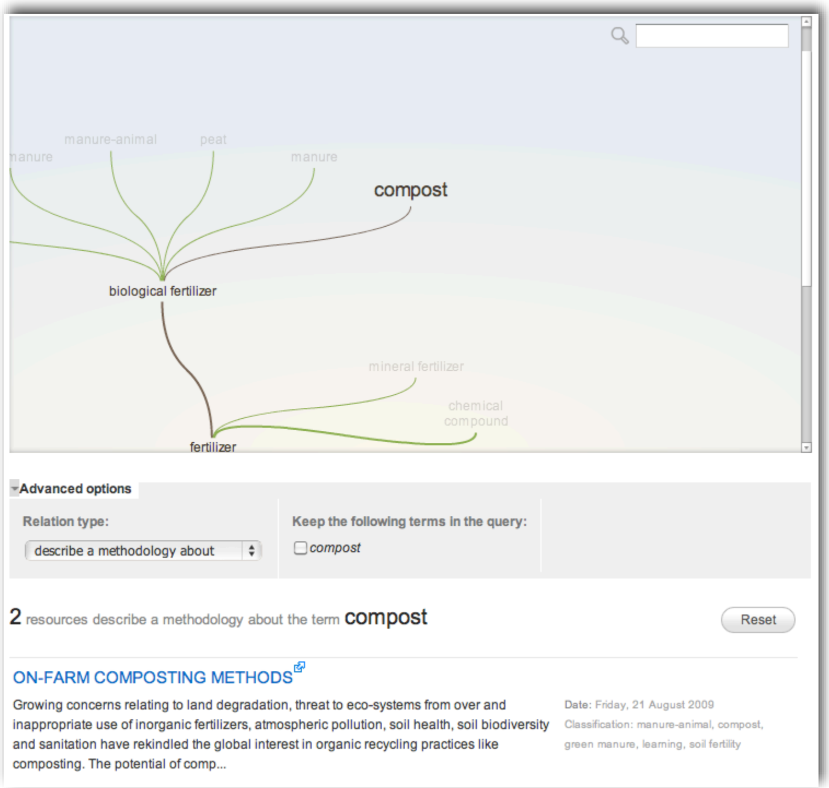

Fig. 1. The "berrypicking" interface exposed in Organic.Edunet

5 Conclusions and outlook

The linked data approach represents an opportunity to evolve existing learning object repository technology towards the open exposure of metadata in a form that enables novel approaches to search and navigation that are not restricted to centralized metadata stores and that enable navigating across repositories without a need for integrating them beyond the provision of links to other linked data sources.

Linked data support in the Organic.Edunet portal is currently in a prototype stage and in ongoing development, so it may still be subject to change. It has been used in a different way from the beginning within the repository tools of the providers to the portal based on Confolio, complying with DCMI/IEEE drafts for encoding the metadata. The support for distributed browsing is currently based only on pilot prototypes, as there are no other learning object repositories exposing linked data that could be used for integration testing. However, the re-design of the browsing interfaces show that the approach is feasible to implement. Future work should deal with the evaluation and refinement of the conventions used to provide semantic relations between learning resources and their metadata, moving to a formal specification that could be used in the many repositories currently exposing IEEE LOM metadata through OAI-PMH or SQI interfaces. It should be noted that the success of the approach presented here relies on the adoption of a shared set of linked data exposure conventions for repositories, which can only come from a kind of community consensus.

Acknowledgments

The work presented in this paper has been funded with support by the European Commission through the projects Organic.Edunet and TEL-Map.

References

Bates, M.: The Design of Browsing and Berrypicking Techniques for the Online Search Interface. Online Review 13 (October 1989): 407-424.

Bizer, C., Heath, T. and Berners-Lee, T.: Linked data - the story so far. International Journal on Semantic Web and Information Systems (IJSWIS), 5(3): 1-22 (2009).

Downes, S. (2004) 'Resource profiles', Journal of Interactive Media in Education, Available at: http://www-jime.open.ac.uk/2004/5.

Ebner, H., Palmér, M.: A Mashup-friendly Resource and Metadata Management Framework, In: Wild, Kalz, Palmér (Eds.): Mash-Up Personal Learning Environments, Proceedings of the 1st Workshop MUPPLE, European Conference on Technology Enhanced Learning (EC-TEL), Maastricht, The Netherlands, 17-19 September, 2008.

Ebner, H., Manouselis, M., Palmér, M., Enoksson, F., Palavitsinis, N., Kastrantas, K., Naeve, A.: Learning Object Annotation for Agricultural Learning Repositories, IEEE International Conference on Advanced Learning Technologies, Riga, Latvia, 15-17 July, 2009.

IEEE LOM, "Draft Standard for Learning Object Metadata", IEEE Learning Technology Standards Committee, IEEE 1484.12.1-2002, 15 July 2002.

Klemke, R., Ternier, S., Kalz, M. and Specht, M. (2010), Implementing infrastructures for managing learning objects. British Journal of Educational Technology, 41: 873–882

Nilsson, M.:Draft Recommended Practice for Expressing IEEE Learning Object Metadata Instances Using the Dublin Core Abstract Model. Draft IEEE P1484.12.4^{tm}/D1 (2008).

Ochoa, X. and Duval, E.: Quantitative analysis of learning object repositories," IEEE Transactions on Learning Technologies 2(3), 226–238 (2009).

Van de Sompel, H., Nelson, M.L., Lagoze, C. and Warner, S.: Resource harvesting within the OAI-PMH framework. D-Lib Magazine, 10(12), 1082-9873 (2004).

Sánchez-Alonso, S.: Enhancing availability of learning resources on organic agriculture and agroecology. The Electronic Library 27(5): 792-813 (2009).

Sicilia, M.A. & García, E. (2003). On the Concepts of Usability and Reusability of Learning Objects . International Review of Research in Open and Distance Learning 4(2).

Simon, B., Massart, D., van Assche, F., Ternier, S., Duval, E., Brantner, S.,Olmedilla, D., Miklos, Z.: A Simple Query Interface for Interoperable Learning Repositories. In: Proceedings of the 1st Workshop on Interoperability of Web-based Educational Systems, pp. 11–18 (2005)