Computer Science and Information Systems 12(1):233–255 DOI: 10.2298/CSIS140330088R

Interlinking Educational Resources to Web of Data through IEEE LOM

Enayat Rajabi[1], Miguel-Angel Sicilia[1], and Salvador Sanchez-Alonso[1]

[1] Information Engineering Research Unit, Computer Science Department, University of Alcalá, 28805, Alcalá de Henares, Spain
{ enayat.rajabi, msicilia, salvador.sanchez}@uah.es

Abstract. The emergence of Web of Data enables new opportunities for relating resources identified by URIs combined with the usage of RDF as a lingua franca for describing them. There have been to date some efforts in the direction of exposing learning object metadata following the conventions of Linked Data. However, they have not addressed an analysis on the different strategies to expose Linked Data that could be used as a basis for leveraging the metadata currently curated in repositories following common conventions and established standards. This paper describes an approach for exposing IEEE LOM metadata as Linked Data and discusses alternative strategies and their tradeoffs. The recommended approach applies common principles for Linked Data to the specificities of LOM data types and elements, identifying opportunities for interlinking exhaustively. A case study and a reference implementation along with an evaluation are also presented as a proof of concept of this mapping.

Keywords: educational resources, Linked Data, IEEE LOM, learning metadata, interlinking.

1. Introduction

The purpose of learning object metadata is to support the reusability and discoverability of learning objects and facilitate their interoperability in the context of e-learning. Particularly, it is used to enable information seekers (e.g., teachers and learners) and applications such as repositories, portals and learning environments to search for, evaluate, retrieve and use learning objects. IEEE LOM [1] is a widespread standard for describing educational contents, promoting their reusability and interoperability through the use of a standardized set of descriptors and common conventions to encode descriptive metadata [2]. This standard is a commonly accepted way for describing learning resources in repositories. A recent study [3] has revealed a consistent usage of 20 out of the 50 metadata elements in the standard, considerably more elements than conventionally collected with widespread schemas such as Dublin Core.

Exposing metadata for search and discovery of resources on the Web has always been an important concern for repositories, and the use of standards is a proof of that. However, IEEE LOM does not explicitly promote relating learning objects, even though "Relation" element has been defined. Specifically, it does not recommend relations to be expressed as links, which is the universal approach in the Web of Data to express

relations among resources. The lack of a shared way of linking precludes crawlers and other applications to get the most out of the relations between resources. By analyzing some of the IEEE LOM elements (as we will discuss later), we found that linking several metadata elements (e.g., coverage) to the Linked Open Data(LOD) datasets, makes the learning object enriched and accessible to the other useful information on the Web of Data. In a sample of 815,223 IEEE LOM metadata records gathered from the GLOBE federation [4], 20% of the resources included the "Relation" element in their metadata records. We examined these resources and found that only 95,946 records (about a 12% of the total) were using URIs to express relations and others contain strings an numbers.

The Linked Data (LD) approach [5] relying on the use of RDF links, represents an alternative way of openly exposing metadata fostering interlinking. RDF links allow to interconnect any kind of resource on the Web, allowing to easily link to external datasets or repositories [6] that are already providing URIs for identifying their resources. Many institutions, universities and libraries have adopted the LD principles and have released resources and data as part of the LOD cloud [7]. Notably, DBpedia [8], one of the most used datasets, which exposes a Linked Data version of Wikipedia, makes it possible for anybody to link to general information as well as to extract relationship to other datasets.

The advantage of this new approach to express relationships between resources is making public information linkable and usable for others [9]. This has the benefit of enabling applications to exploit learning object metadata and other information available in the Web of Data. It can also be seen as an extension of open educational resources initiatives [10] in the direction of making them more readily available for discovery.

The exposure of LOM compliant metadata as Linked Data supports functionalities over RDF-defined LOM records that cannot be attained with the human-oriented version of LOM, e.g., triggering queries on SPARQL endpoints [11] no matter where the records are stored. Users can also export their educational metadata in LOD format in the same manner that libraries all over the world are doing in the library field. However, exposing LOM metadata as LOD is not straight-forward and requires a transformation of metadata plus a bootstrapping phase to identify candidate links to other datasets or educational resources, eventually with the aid of interlinking tools [12] [13]. This in turn requires the use of vocabularies to provide some shared semantics that can be exploited for the traversal of metadata across repositories. Given that some semantics are actually encoded in the IEEE LOM standard, there is a need to elaborate some RDF exposure practices for which existing proposals for mapping IEEE LOM to RDF [14] [15] can be useful but are not enough. This includes URI design and the identification of opportunities for interlinking.

This paper reports a complete analysis on the different strategies to expose IEEE LOM as Linked Data, describing how IEEE LOM elements and data types can be represented in RDF based on Linked Data principles [5] and complying with common Linked Data patterns [16]. It also reports on a case study and reference implementation and evaluates its performance. The case is based on the Organic.Edunet repository [17], a IEEE LOM-based repository of learning materials in the field of organic agriculture and agroecology.

The rest of paper is structured as follows. Section 2 briefly describes the background on exposing IEEE LOM elements and the related works in this context. In Section 3, we recommend a URI design for identification of e-learning objects in educational repositories. This section also represents a mapping of LOM elements to Linked Data

format. Section 4 provides an experimental implementation of RDF [18] binding of IEEE LOM. Section 5 presents an evaluation and performance testing over the mentioned implementation. Conclusions are provided in Section 6.

2. Background and Related Works

Work on e-learning metadata standards at the international level has been carried out by a number of organizations including the IEEE, the Dublin Core Metadata Initiative (DCMI), IMS Global [19], and ISO/IEC [20]. Achieving interoperability across these specifications has been recognized as a major challenge since 2000 [21].

IEEE LOM is an internationally-recognized open standard bound up with the history and development of the IMS e-learning interoperability specifications (e.g. IMS Content Packaging [19]), and with the evolution of the ADL SCORM [22] reference model, which supports the IEEE LOM alongside other specifications. Dublin Core (DC) has also been used in many systems and applications as an alternative to other metadata standards (e.g. IEEELOM) or in combination with them to provide wider interoperability.

A recent effort within the ISO community is Metadata for Learning Resources (MLR) [23] which aimed at harmonizing LOM and Dublin Core metadata, as it tries to enable both the "learning object" aspects of LOM and the "entity-relationship" model of the Semantic Web associated with the Dublin Core Abstract Model [24]. Moreover, it is intended to support multilingual and cultural adaptability requirements from a global perspective. The Learning Resource Metadata Initiative (LRMI) [25] has also developed a common metadata framework for describing learning resources on the web. LRMI promoted by popular search engines Google, Bing, and Yahoo, is related to schema.org and supported by Creative Commons. Although the goal of these schemas is to be a complement or alternative to IEEE LOM and DC, a wide variety of learning repositories and federations (notably the GLOBE federation [4]) use IEEE LOM as the base metadata schema and actively aggregates LOM records at a large scale.

To date, there have been some initiatives to expose learning resource metadata as Linked Data. Dietze et al [26] surveyed some high-level approaches aimed towards Linked Education by allowing its exposure as Linked Data and interlinking techniques for the educational domain. Dietze et al [27] also proposed an approach for linking educational resources based on the Linked Data principles by using existing educational datasets and vocabularies. Its aim was to exploit the wealth of existing technology-enhanced learning (TEL) data on the Web by exposing it as Linked Data. The approach has been implemented in the context of the mEducator project [28] where data from a number of open TEL data repositories has been integrated, exposed and enriched making use of the Linked Data approach.

Fernandez et al [29] presented a work on linking educational resources across universities through the use of Linked Data principles by focusing on extracting and structuring information of video lectures produced by 27 different educational institutions according to some vocabularies, e.g. FOAF. As a result of this work, a new media educational dataset was released.

There exist some other projects such as LinkedUp [30] and Linked Universities [31] which aim at sharing learning data or metadata related to educational Linked Data on state-of-the-art Linked Data principles.

In particular, Linked Data exposure of IEEE LOM is not a new subject though, as the work was initiated in 2000 in the context of the IMS Global Learning Consortium [19] (together with the ARIADNE Foundation [32]) that developed a XML binding and RDF binding of LOM elements and, as a result, some RDF documents were produced as IMS RDF Bindings. The Dublin Core Metadata Initiative (DCMI) [33] also provided recommendations for expressing DC metadata as RDF and described how the features of the DCMI are represented based on LOM to DCAM mapping document [34]. The recommended document described how to use the definitions of metadata terms defined by the IEEE LOM Standard, for RDF binding of IEEE LOM Elements together with DCMI metadata terms.

A mapping from LOM to RDF model (defined by Nilsson et al. [15]) described advantages of expressing learning object metadata as RDF. Nilsson also discussed some problems encountered in the process of producing the RDF binding for LOM elements and focused on some specific futures of the binding, although this early work was discontinued [26] and did not cover all the LOM elements.

Some other tools and IEEE LOM editors also export LOM elements as RDF. For example, ocw2rdf [35] harvests metadata from the MIT Open Course Ware web site [36] and transforms it into an RDF representation of IEEE LOM. Kunze et al [37] developed and implemented a browser-based editor in which the author can choose the type of metadata using any kind of RDF-schema available on the WWW to annotate learning resources in a specific repository (OLR3). Balog-Crisan and Roxin [38] proposed an on-line tool called RDF4LOM, to edit metadata in RDF. The proposed tool creates RDF documents according to the LOM standard.

Our work continues and completes Nilsson et al [15] approach for exposing learning object metadata as Linked Data. To this aim, we consider all the IEEE LOM elements, data types and vocabularies and provide a mapping to RDF. We also present a complete and unified solution for exposing learning object metadata and implement this approach on an educational repository so that this repository can link its (meta)data to Linked Open Data by following clear guidelines. As the RDF implementation is not straightforward and the decisions for the transformation of several items often have 2 or more possible alternatives. We tried to base our decisions and recommendations on good practices, but even so, our decisions are subject to debate and can evolve in the future.

3. Exposing IEEE LOM as Linked Data

In this section, we highlight on the exposure of the IEEE LOM elements as RDF, represented here in XML format. Initially, we discuss how e-learning objects are identified in LOM elements. The recommendation presented in this study is the outcome of a long authors' discussion with both Linked Data and e-learning experts. A complete mapping of all the IEEE LOM elements is available at http://data.organic-edunet.eu/ODS_LOM2LD/ODS_SecondDraft.html.

3.1. URI Design

In IEEE LOM, identifiers are defined as "globally unique label that identifies a learning object" and are to be provided in:

- Element 1.1: General.Identifier as the identifier of the resource
- Element 3.1: Meta-Metadata.Identifier as the identifier of the metadata record
- Element 7.2.1: Relation.Resource.Identifier as the identifier of a related resource

In a general case, the dereferenceable URIs that deliver RDF descriptions, are actually identifying metadata records and not the actual resources. In consequence, the identification in Element 3.1 is represented as the dereferenceable URI from which the RDF metadata is exposed, and there is no need to expose it again in the RDF representation. In the case of the "Relation" element, the recommended practice is using the dereferenceable URI of the resource pointed by this one, if available, in the form of a RDF link. If the URIs of learning objects are considered to form a natural hierarchy, then a patterned URI can be assigned to them [16].

In terms of technical design, World Wide Web Consortium (W3C) published some guidelines in order to define a well-formed URI [39] [40]. The document we used as a basis for our solution to define learning object identifiers, stated two approaches based upon the HTTP URI scheme and protocol which fulfils the following requirements:

Description of the identified resource should be retrievable with standard Web technologies.

A naming scheme should not confuse things and the documents representing them.

3.2. Binding Simple and Structured Elements

Two types of LOM elements exist: simple and structured (or aggregation). The following sub-sections discuss the RDF representation of each type. One metadata example of a learning resource (e.g., http://youtube.com/example_resource), represented in an XML format, is used throughout this paper and, therefore, we will avoid repeating the resource identifier in each example. As simple elements do not contain other LOM elements and mostly include one value (e.g., String) at the target, they have been represented plainly as subject, predicate and object. This RDF binding have been recently followed by many datasets in the LOD cloud (e.g., DBpedia, Factbook). As an illustration, technical format of learning objects in LOM (Technical.Format) is expressed in Turtle [41] (Consider Table 1).

Structured elements included other LOM elements (either simple or structured elements) are often realized using intermediate nodes, but there exist various options for exposing structured elements as LD depending on maximum number of entities they include (multiplicity) and their order. Several LOM elements (e.g., "General.Title") with structural format were considered with multiplicity one in the IEEE standard and given that their order is not significant in the metadata, the simplest way of representation and already compatible with a wide range of existing software, is leveraging the repeated properties in RDF. In the repeated properties, the user can assign many predicates to one subject regardless to its objects' order and thus can be applied to appropriate elements.

Table 1 illustrates how a structured element ("General.Title") is expressed in different languages in RDF.

Table 1. RDF binding of IEEE LOM elements

Binding type	XML representation example	RDF Binding
simple element	<technical> <format> "application/x-shockwave-flash" </format> </technical>	<http://youtube.com/example_resource> dcterms:format "x-shockwave-flash".
element with multiplicity 1	<title> <string language="en"> What is organic. </string> <string language="de"> Was ist biologisch. </string> </title>	<http://youtube.com/example_resource> dcterms:title "What is organic"@en; <http://youtube.com/example_resource> dcterms:title "Was ist biologisch"@de.
structured element using blank nodes	<general><keyword> <string language="en"> Certification </string> <string language="de"> Zertifizierung </string> </keyword> <keyword> <string language="en"> Farming </string> <string language="de"> Landwirtschaft </string> </keyword></general>	<http://youtube.com/example_resource> lom:keyword _:node1,_:node2. ---------- _:node1 rdf:value "Certification" @en, "Zertifizierung" @de. _:node2 rdf:value "Farming" @en, "Landwirtschaft" @de.

Intermediate nodes, also called blank nodes due to the absence of a name (or a dereferenceable URI) to a triple, are used to indirect referencing to a element with unspecified name. Although intermediate nodes are considered as problematic approach in terms of implementation of RDF and users readability [42], their usage is unavoidable when there exists a deep hierarchy (more than two) of elements in a model or the multiplicity of structural elements is "Many". The "Keyword" element in "General" category is a good practice, as can be expressed repetitive in more than one language and thus using repeated properties here is not applicable, as it would mix the translations of the different values. In the table above we showed that how two intermediate nodes have been used for representation of the keyword element.

Figure 1 also portrays a simple guideline for the RDF binding of simple and structured elements of IEEE LOM according to foregoing discussion.

Alternatively, RDF containers e.g., RDF:Alt and RDF:Seq [16] are applied to describe a group of values in RDF representation and they are appropriate when the element hierarchy is limited in two levels. RDF:Seq suits particularly when the order among elements is important (see Table 2). As this representation becomes more complicated in deep hierarchical structures of the IEEE LOM elements (e.g., classification.taxon), using the RDF containers for the elements that are not explicitly required to be ordered, is not recommended.

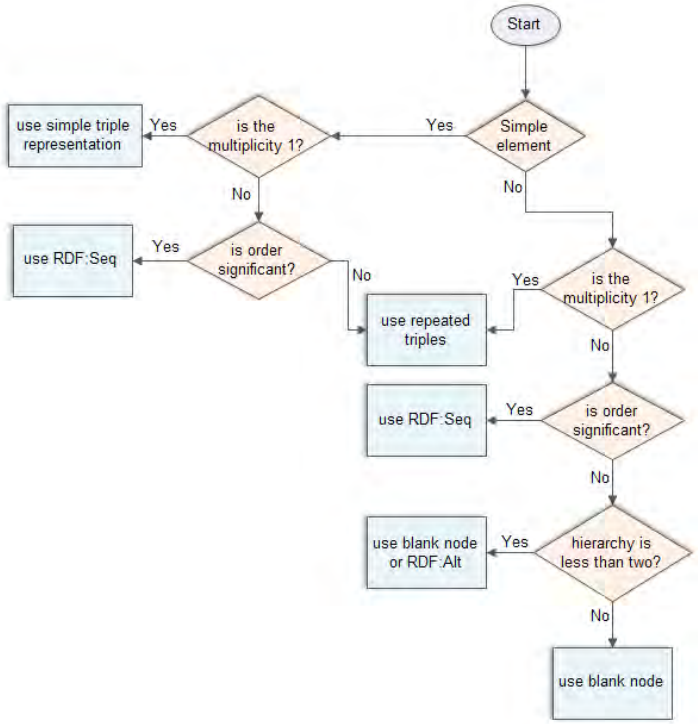

Fig. 1. The workflow of RDF binding of IEEE LOM elements

Table 2. RDF binding of a structured element using RDF containers

XML representation	RDF Binding
<classification> <keyword> <string language="en"> Organic <string> </keyword><keyword> <string language="en"> Farming <string> </keyword> </classification>	<http://youtube.com/example_resource> lom:classificationKeyword _:node1. _:node1 a rdf:Seq; rdf:_1 "Organic" @en. rdf:_2 "Farming" @en.

3.3. DataType Mapping and Reusing Vocabulary

The following sub-sections provide a description of data type mapping of the IEEE LOM elements.

3.4. CharacterString

Simple elements in String format are represented as plain literals in RDF, e.g., “Technical.Format” in the LOM standard, whose type is “CharacterString”, would be represented as follows:

```
<http://youtube.com/example_resource>
    dcterms:format "x-shockwave-flash".
```

3.5. LangString

Several IEEE LOM elements use the “LangString” data type which binds together multiple literals with equivalent expressions in different languages. The literal is expressed as a plain literal in RDF along with a language tag (e.g., en) conformed to RFC1766 [43]. The “LifeCycle.Version”, as a good practice, has multiplicity one and is therefore, as mentioned earlier, represented as a direct property pointing to a plain literal with a language tag:

```
<http://youtube.com/example_resource>
       lom:version
            "It is not available" @en,
            "No está disponible" @es.
```

3.6. DateTime

The International Standard for the representation of dates and times, ISO 8601 [20], describes a large number of “DateTime” formats. IEEE LOM standard defines at least four digits for year, two for month and two for day. For representing the time, it states two digits for hour, two for minutes, two for seconds and one or more digits representing a decimal fraction for a second. IEEE LOM elements that represent “DateTime” values can be exposed in the following format:

```
<http://youtube.com/example_resource>
   lom:contributionDate  "2011-05-17T05:53:31.00Z"
```

IEEE LOM allows “DateTime” elements to be expressed as literal with language (e.g., {“en”,“circa 1300 BCE”}). For those elements, we recommend “LangString” representation as follows.

```
<http://youtube.com/example_resource>
   lom:contributionDate "circa 1300 BCE" @en
```

3.7. Duration

Duration, as an interval data type, is recommended to be expressed as follows:

```
<http://youtube.com/example_resource>
   lom:technicalDuration  "PT0.25S"
^^<http://www.w3.org/2005/xpath-
datatypes#dayTimeDuration>.
```

In the above example, duration ("PT0.25S") shows that technical duration of the learning object is 25 seconds based upon ISO8601 [20], although the represented format is not human readable. As "DateTime" data type, elements with String value representing Duration (e.g., {"en", "Fall Semester 1999"}) are expressed as LangString.

3.8. Boolean, Integers and other Simple Data Types

In the RDF exposure, it is encouraged to reuse the XML schema data types [44]. For example for "Boolean" values, the data type of the element is indicated as true or false:

```
<http://youtube.com/example_resource>
   lom:cost false
   ^^<http://www.w3.org/2001/XMLSchema#boolean>.
```

Likewise, for expressing other simple data types such as integer, long, float, etc. using the XML schema data type is recommended.

3.9. vCard:

vCard [45] is a standard for electronic business cards. To capture a vCard, an intermediate node is recommended together with properties such as vCard:FN, vCard:ORG and vCard:Email. The entity value of contribute element in the LifeCycle category, the Table 3, is a vCard record represented in XML.

Table 3. RDF binding of vCard

XML representation	RDF Binding
BEGIN:VCARD FN:John Smith EMAIL;TYPE= INTERNET: John@example.org ORG: http://www.example.org N:John;Smith VERSION:3.0 END:VCARD	<http://youtube.com/example_resource> lom:contributionEntity _:bnode1247. _:bnode1247 vcard:FN "JohnSmith". _:bnode1247 vcard:N "John;Smith". _:bnode1247 vcard:EMAIL _:bnode1248. _:bnode1247 vcard:ORG "http://www.example.org". _:bnode1247 vcard:VERSION 3.0. _:bnode1248 rdf:value "John@example.org"; rdf:type "http://www.w3.org/2001/vcard-rdf/3.0#internet".

3.10. Undefined Data Type

The IEEE standard states "Undefined" as one of the data types of LOM elements, although most date types are expressed explicitly and can be represented in RDF. For example "xsd:dateTime", is used for "DateTime" format and "xsd:boolean" for "Booleans" and so forth. However, if an element cannot be defined in any specific date type in the LOM schema, "xsd:anyType" is recommended, which does not restrict the data content [44].

3.11. Reusing of Existing Vocabularies

Several well-known vocabularies are used in Linked Data to describe things such as people, places, and locations. By reusing known vocabularies, data publishers increase their chance of being interoperable with other parties as well as avoid the time consuming process of defining and documenting own vocabularies. In consequence, we mention a brief guide of reusing the vocabularies as an example:

- To describe simple data, use the basics of RDF and RDFS
- To name things, use "rdfs:label", "dcterms:title", and "foaf:name"
- To describe people, use FOAF and vCard
- To describe Web pages and other publications, reuse Dublin Core properties, for example "dcterms:creator" and "dcterms:description"
- To describe addresses, use vCard

4. Interlinking to other Datasets

Linked Data (LD) approach unlocks e-learning resources away from learners and enables enriching, navigation, casual discovery and improved resource seeking. Linking resources using LD also makes it easy for intelligent processing of data, as several operations e.g., integration of experiment data, consumption, and publication of experiment data are doable using the related tools. Particularly, the LD exposure of educational materials became a general approach specially for enrichment of learning resource as well as interlinking them to useful datasets on the Web of Data. To this end, some institutions have emerged their educational materials as LD. For example Europeana dataset [46], as European Union flagship digital library project, links the data providers metadata to other datasets such as DBpedia and Geonames [47] as well. As Figure 2 depicts, interlinking LOM elements to one dataset, makes the metadata global to access other valuable information over the LOD.

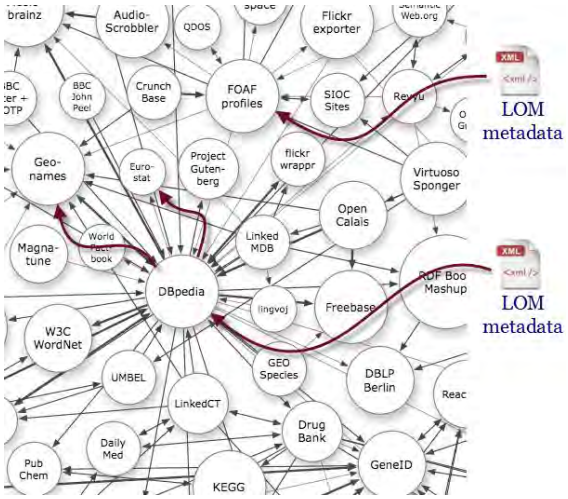

Fig. 2. Linking educational data to LOD

To this end, we examined all the IEEE LOM elements to discover the linkable elements to the LOD cloud. Figure 3 portrays the workflow we followed for the interlinking analysis. In the first step, we looked over those elements that potentially cannot be linked due to their specific values, and thus, they have been filtered out (e.g., "DateTime", controlled vocabularies). In the second step and being precise on the metadata records, we discovered that several elements (e.g., identifiers, vCard) contains local values defined by each repository according to its policy. Particularly, the values did not follow a specific rule for interlinking purpose. As a consequence of analysis, we found various elements such as "General.Title" and "Technical.Format" include string values that can be linked to the related datasets. However, after running an interlinking tool to link the data to a specific dataset, the outcomes were a few and in the most cases were not useful.

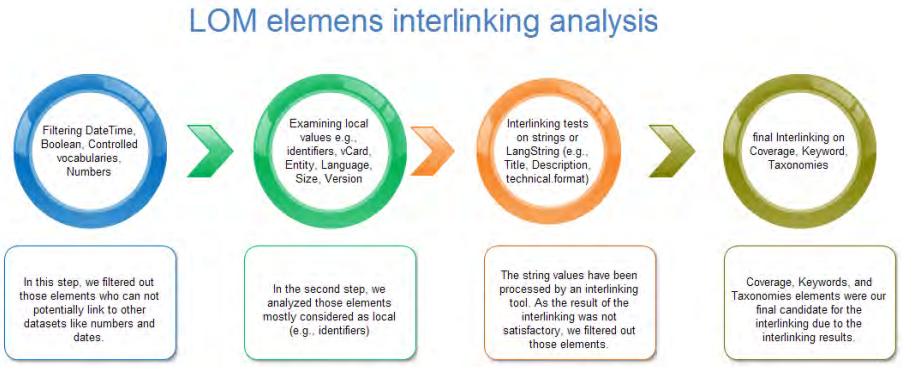

Fig. 3. Linking educational data to LOD

To be specific, linking coverage of a learning object to DBpedia, as a good practice, not only adds more geographical information about the place, but also allows metadata to be connected to other statistical sources (e.g., population, history) as well. In the

following sub-sections, we will summarize a couple of important elements of IEEE LOM, which can be linked to the LOD as well.

4.1. Linking Elements to DBpedia

The DBpedia dataset includes structured information about persons, places and organizations. It features labels and abstracts for 10.3 million unique things in 111 different languages [8]. This dataset has been recently identified as a hub in the LOD cloud [48], as it connects a wide variety datasets together with high centrality. Particularly in eLearning context, Lama et al. [49] presented an approach that automates the classification of learning objects and improves its search in repositories by annotating the learning objects with DBpedia ontology. As we will discuss later, DBpedia is also significant place for linking coverage of educational materials ("General.Coverage" in LOM) to regions, countries and cities of DBpedia Other datasets such as GEMET [50] and Eurostat [51] can be used for this purpose, as they include useful information about statistics of public places. The "Keyword" element of learning objects ("General.Keyword" in LOM) can be also linked to DBpedia, as we found a lot of similarities between the keywords of aggregated e-learning resources and the DBpedia labels.

4.2. Linking the "Classification" Category of IEEE LOM to LOD

The IEEE LOM provides an area for annotating and classifying educational resources and makes them discoverable specially when a learning resource is accessible based upon the classification it belongs. It expresses the classification of a learning object in the classification category that each classification includes purpose and taxonpath. The taxonpath states the structure of the taxonomy. One of the possibilities of classification interlinking, for example, is linking the taxonomy of a learning object to the LOD taxonomy dataset [52]. This dataset as a knowledge base provides informative LOD URIs for species concepts that improve the quality and stability of links between a species and the related data. There exist around 108,175 species concepts and 1,000 records for species occurrences [53], interlinked with the GeoNames dataset [47]. The following example illustrates part of Organic.Edunet metadata linked to the LOD taxonomy dataset through the classification category.

```
<http://youtube.com/example_resource> lom:classification
_:classification1.

_:classification1 lom:purpose lomvoc:discipline;
        lom:classificationDescription
"This classification provides many examples of Organic
Principles and Ontologies. @en";

lom:taxonPath _:taxonpath1.
```

```
_:taxonpath1 lom:taxonSource "LOD taxonomy" @en;
             lom:taxon  _:taxon1, _:taxon2.
_:taxon1 lom:id

"http://lsd.taxonconcept.org/describe/Organic_farming";
       lom:entry  "Organic farming category" @en.
_:taxon2 lom:id
"http://lsd.taxonconcept.org/describe/Certification";
lom:entry  "Certification" @en.
```

4.3. Linking the "Relation" Category of IEEE LOM to LOD

The "Relation" category of IEEE LOM groups features that establish the relationship between the learning object and other related learning objects. As learning objects may include different relations, they can be exposed in RDF in different intermediate nodes. The following example shows the relation of our sample learning object to DBpedia represented in RDF. In particular, the learning object is linked to many related resources exist in the DBpedia dataset through the Relation category.

```
<http://youtube.com/example_resource>
             lom:relation _:relation1.

_:relation1lom:relationKind  dcterms:isPartOf
          lom:relatedResource _:resource1;
            lom:resourceDescription
_:resourceDescription1.

 _:resource1 lom:relatedResourceCatalog "URI";
             lom:relatedResourceEntry
             "http://live.dbpedia.org/page/Agriculture".

_:resourceDescription1 rdf:value "Organic farming is kind
of agriculture that has been explain" @en.
```

5. Architecture and Implementation

The RDF binding of LOM elements is not sufficient for exposing educational materials as Linked Data, as Linked Data principles [5] should be covered by an educational repository in order to have the learning resources in a linkable way. To this aim, the repositories cater a service or API which users are able to make queries via SPARQL endpoint [11]. Repositories can also provide an RDF dump [53] which makes the whole dataset to be accessible through the repository website. Here, we propose an architecture along with a software prototype implemented on the Organic.Edunet [17] repository as our case study.

5.1. Exposing Organic.Edunet Resources as Linked Data

As previously mentioned, Organic.Edunet is a learning portal that provides access to digital learning resources as well as their metadata on Organic Agriculture and Agroecology and aims to facilitate access, usage and exploitation of such content. Several types of e-learning resources including reports, handbooks, presentations, experiments and lesson plans are available through the portal [17]. The LD exposure of the Organic.Edunet metadata [54] was performed by taking the following steps:

Initially, educational metadata were stored in the Organic.Edunet repository in XML format based upon an IEEE LOM Application Profile [55]. We transformed the XML files into a relational database by developing a transformer tool. In consequence, we exposed the metadata as Linked Data by using a mapping tool (e.g., D2RQ [56] as a mapping tool for mapping relational databases to RDF). In particular, we represented the educational metadata in a complete uniform dataset and made them accessible via a SPARQL endpoint and a RDF dump. The proposed architecture is presented in three layers as Figure 4 depicts.

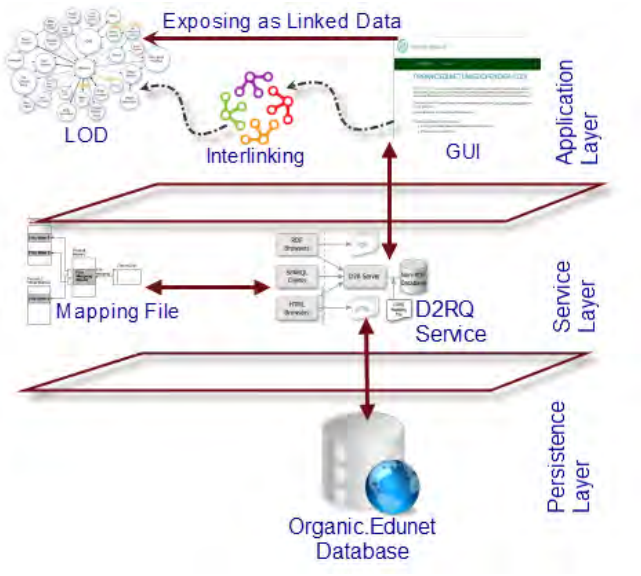

Fig. 4. The LD architecture of Organic.Edunet

In the persistence layer, the metadata are collected in the Organic.Edunet repository and converted the XML files into a relational database by developing a Java program. In the service layer, a D2rQ service mapped the relational database to the RDF format. We created a mapping file in order to express the relational structure to the RDF triples. In the application layer, we implemented an interface in front-end to depict the educational metadata in a graphical user interface (GUI) along with a search interface for users. Particularly, the SPARQL Endpoint and RDF dump of dataset made the data to be available through the GUI. We established a link between the Organic.Edunet dataset and DBpedia by mapping around 11,093 metadata records in the relational database to

RDF. This was performed by running a simple code to find the similarities between the metadata elements and DBpedia. Table 4 illustrates the matches between Organic.Edunet and the DBpedia dataset for "Keyword" and "Coverage" elements (although the interlinking analyzed have been based upon equal string match without any consideration of polysemy and lexical variants). Finally, around 73% of coverage of the learning objects (e.g., countries and cities) and 23% of keywords matched to the DBpedia concepts. Upon this finding, it is reasonable to conclude that the IEEE LOM elements include latent potential for linking to other datasets on the Web of Data.

Table 4. Interlinking Organic.Edunet to DBpedia

Metadata element	Total number	Matched number
Keyword	99,506	22,087 (23%)
Coverage	11,906	8,585 (73%)

5.2. Performance Testing over the Case Study

Regarding the performance testing of implementation, we used JMeter [57], as a testing tool for performance measurement and selected three queries to simulate the work as well. The queries became more complex from query 1 to query 3 according to Semantic Publishing Benchmark (SPB)[1], as a LDBC[2] benchmark for testing the performance of RDF engines (consider the appendix). SPB defines a set of "choke points" to evaluate the reliability of RDF database and address the complexity of queries. In particular, "join ordering" as one the choke points, tests the ability to consider cardinality constraints and decide which type of join should be used in a query, as it has been pointed out by other studies as well (e.g., [58]). We simulated as well as evaluated the queries on the same machine over D2RQ service and a triple store for 1, 5 and 10 users to compare the performance between them. Each query was repeated for five times in order to examine the results precisely. The Linked Data version of the Organic.Edunet was evaluated by making use of the query pages of D2RQ service. As Table 5 shows, the performance of queries decreases when they are run by more users. Obviously, the response time increases when they become more complex. As it can be seen from the table, there is a huge difference between response time of RDB and D2RQ services, as D2RQ performs both mapping the queries to SQL and running them over the relational database at once.

We also examined the implementation on a triple store in order to analyze the performance of executing the queries on a triple store directly. To this end, we imported the RDF dump of Organic.Edunet dataset in a Jena-Fuseki triple store [59] and evaluated the queries via its SPARQL page, as Table 6 illustrates the result of JMeter.

[1] http://ldbcouncil.org/benchmarks/spb
[2] http://ldbcouncil.org/

Table 5. Performance testing on D2RQ mapping service (5 times running of each query)

Query #	1 user	5 users	10 users
Query 1	661 ms	700 ms	746 ms
Query 2	526 ms	1773 ms	3632 ms
Query 3	1356 ms	4317 ms	9778 ms

Table 6. Performance testing on a triple store (5 times running of each query)

Query #	1 user	5 users	10 users
Query 1	8 ms	14 ms	13 ms
Query 2	10 ms	15 ms	44 ms
Query 3	67 ms	72 ms	200 ms

Executing the queries on a relational database (RDB), we realized that the analysis of executing the queries on RDB and triple store is comparable, as Figure 5 depicts the difference between these two data stores in terms of response time. The queries in both cases are shown on the x-axis (for 10 users), while the y-axis illustrates the runtime in milliseconds. Comparing these results with the mapping approach mentioned earlier, we can conclude that the D2RQ mapping service is not scalable when the number of users increases and queries become more complicated.

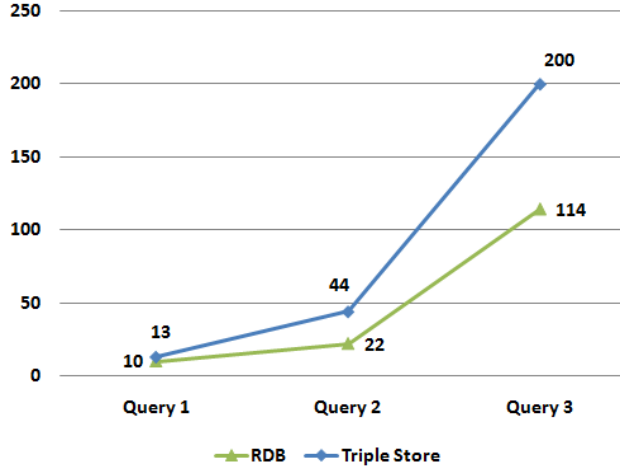

Fig. 5. Response time comparison between relational database and triple store in

6. Evaluation of the Case Study

Evaluating the interlinking results between the Organic.Edunet and DBpedia datasets, we realized that the "Coverage" element of e-learning resources in the Orgainc.Edunet repository includes information about countries and places that can be connected to the

DBpedia places. In particular, a user in Organic.Edunet can explore the dataset and obtain useful knowledge about the country or place of resources. To take a scenario about the advantage of such interlinking through the "Keyword" element, a teacher in agricultural science might explore the contents to find an article about "organic farming". In one of the results, a relevant book catches the teacher's attention and thus follows the keywords of article to find out the exact context of the learning resource. The researcher has never come across the specific terms which do not yield any more relevant data. As the learning resources in Organic.Edunet are interlinked to DBpedia, more information on topic including different translations are presented to him, when he is redirected to the DBpedia pages.

As a consequence of quality control of interlinked data in Organic.Edunet repository, we selected 20 random resources enriched by DBpedia over the "Coverage" and "Keyword" elements and presented to five end users. The Organic.Edunet resources included a full metadata information and we asked the users to explore especially the "Keyword" element linked to the DBpedia pages. In particular, the users were asked to answer 4 questions regarding the interlinked metadata elements. The questionnaire included the following statements regarding the linked items:

1. Was the link available to evaluate?

Here we asked whether the user can reach the target by clicking the provided URL or not? (As some links might not be available either the link is broken or it does not respond in a reasonable time).

2. Was the link information related to the term?

The relatedness of information to the term is evaluated by the user in the question above. It is possible that the provided information in the target semantically is not the same as source due to e.g., polysemy or ambiguity between them. For instance, there exist several abbreviations in the "Keyword" element (e.g., TOF, SDW...) which might refer to different terms.

3. Did the link information help you to find more useful data regarding the resource?

The most important question, from the authors perspective, was the usefulness of provided link. Overall, the users were asked if the link information in the target included useful knowledge about the resource and particularly could help learners to obtain usable data.

4. What do you recommend for improving the quality of interlinking?

Finally, we asked users to write their comments regarding the improvement of this experience.

Table 7. users' answers to the questions (5 times running of each query)

Question #	User 1	User 2	User 3	User 4	User 5
Q1	20	18	20	19	19
Q2	16	15	14	15	16
Q3	12	13	13	12	14

As Table 7 illustrates, almost all of the links were available for the evaluation. Also, the average number of resources that were relevant to the terms, was around 15 resources (76%). This amount of resources implies that a few number of them included

ambiguity or not informative for users to examine. To gauge the responses reliability of Question 3, we applied intra-class correlation coefficient (ICC) [60], as one of the popular reliability statistics, to determine the internal consistency of multiple raters. In this approach, the accepted value for describing internal consistency is defined by an alpha greater than 0.6 and the results is highly coefficient when value is more than 0.9. We later imported the users' answers into SPSS to analyze the responses and run the reliability statistics. Accordingly, the software output for our data was 0.726 that shows the users agreed on the results and the average number of questions determined by the users as useful was around 13 (65% of all questions).

Regarding the Question 4, one of the users commented that interlinking Organic.Edunet to the DBpedia dataset gives general information about the terms to readers, but if users want to obtain more information about the resource (e.g., relevant books or articles), they have to explore the Web. Interlinking Organic.Edunet to more educational and scientific datasets (e.g., universities) was also recommended by the user. Other users did not mention any important comments.

7. Conclusion

The widespread adoption of the Linked Data approach has led to the availability of huge amount of data ranging from public domain such as DBpedia to domain-specific space, for example Europeana which includes data about cultural heritage. Connecting e-learning resources to the LOD makes educational materials linkable to other useful datasets as well as enriches the contents as well.

To this aim, we discussed mapping and linking of the IEEE LOM elements, as an accredited metadata schema for describing educational materials, to the Linked Data based upon its principles. We developed an implementation of this approach for the Organic.Edunet repository, as our case study, so that the metadata of the e-learning resources became accessible through a graphical user interface. The metadata were also linked to other datasets in LOD (e.g., DBpedia). At the time of this research, the SPARQL endpoint of the Organic.Edunet dataset is available for users to make queries. Likewise, other educational datasets can foster their data to the released dataset. Eventually, some selected queries passed a performance testing on both relational database and triple store considering their complexity. The analysis of performance testing states that providing a powerful triple store on top of the Linked Data exposure of e-learning repositories dramatically improves the performance than using a mapping tool to convert the data as Linked Data format.

Acknowledgments. The work presented in this paper has been part-funded by the European Commission under the ICT Policy Support Programme CIP-ICT-PSP.2011.2.4 - elearning with project No. 297229 "Open Discovery Space (ODS)". The authors also would like to thank Dr. Hannes Ebner and Matthias Palmér for their assistance and helpful suggestions about the Linked Data exposure of Organic.Edunet.

References

1. IEEE LTSC, IEEE Standard for Learning Object Metadata 1484.12.1-2002, Final Draft version, [Online]. Available: http://ltsc.ieee.org/wg12/20020612-Final-LOM-Draft.html. [Accessed: 22-February-2014].
2. Ochoa, X. and Duval, E.: Towards automatic evaluation of learning object metadata quality. In Proceedings of the 2006 international conference on Advances in Conceptual Modeling: theory and practice, Berlin, Heidelberg, 372–381. (2006)
3. Ochoa, X., Klerkx, J., Vandeputte, B. and Duval, E.: On the use of learning object metadata: the GLOBE experience, Proceedings of the 6th European conference on Technology enhanced learning: towards ubiquitous learning, Berlin, Heidelberg, 271–284. (2011)
4. GLOBE | Connecting the World and Unlocking the Deep Web. [Online]. Available: http://globe-info.org/. [Accessed: 09-Jun-2013].
5. Bizer, C., Heath, T., and Berners-Lee, T.: Linked Data - The Story So Far. International Journal of Semantic Web Information System, vol. 5, no. 3, 1–22, 33. (2009)
6. Sicilia, M.-A., Ebner, H., Sanchez-Alonso, S., Alvarez, F., Abian, A., and Garcia-Barriocanal, E.: Navigating learning resources through linked data: A preliminary report on the re-design of Organic.Edunet, presented at the 1st International Workshop on eLearning Approaches for the Linked Data Age. (2011).
7. The Linking Open Data cloud diagram. [Online]. Available: http://lod-cloud.net/. [Accessed: 14-May-2013].
8. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., and Hellmann, S.: DBpedia - A crystallization point for the Web of Data, Web Semantics, vol. 7, no. 3, 154–165. (2009).
9. M. D'Aquin. Linked Data for Open and Distance Learning, Common Wealth of Learning, 2012. [Online]. Available: http://elearningeuropa.info/en/directory/Linked-Data-for-Open-and-Distance-Learning. [Accessed: 22-February-2014].
10. OER Commons. [Online]. Available: http://www.oercommons.org/. [Accessed: 17-Jun-2013].
11. SparqlEndpoint Description. W3C Wiki. [Online]. Available: http://www.w3.org/TR/sparql11-service-description/. [Accessed: 17-Jun-2013].
12. Simperl, E., Wölger, S., Thaler, S., Norton, B., and Bürger, T.: Combining human and computation intelligence: the case of data interlinking tools, International Journal of Metadata and Semantic Ontologies, vol. 7, no. 2,77–92. (2012)
13. Hausenblas, M., Halb, W., and Raimond, Y.: Scripting User Contributed Interlinking, in Proceedings of the 4th workshop on Scripting for the Semantic Web (SFSW2008), co-located with ESWC2008. (2008)
14. Dublin Core metadata initiative, Dublin Core Metadata Element SetVersion 1.1, Dublin Core metadata initiative, Dublin Core Metadata Element SetVersion 1.1. [Online]. Available: http://dublincore.org/documents/dces/. [Accessed: 22-February-2014].
15. Nilsson, M., Palmer, M., and Brase, J.: The LOM RDF Binding - Principles and Implementation, Proceeding of third annual ARIADNE Conference. (2003)
16. Dodds L., and Davis, I.: Linked Data Patterns: A pattern catalogue for modeling, publishing, and consuming Linked Data, Linked Data Patterns Pattern Cat. Model. Publ. Consum. Linked Data. (2010)
17. Organic.Edunet Project. [Online]. Available: http://wiki.organic-edunet.eu/index.php/Main_Page. [Accessed: 22-February-2014].
18. Klyne, G., and Carroll, J. J., Resource Description Framework (RDF): Concepts and Abstract Syntax. W3C Recommendation. (2004)
19. IMS Global Learning Consortium, [Online]. Available: http://www.imsglobal.org/. [Accessed: 22-February-2014].

20. Representation of dates and times. International Organization for Standardization, ISO 8601:2004, 2004. [Online]. Available: http://www.iso.org/iso/catalogue_detail?csnumber=40874. [Accessed: 22-February-2014].
21. Flack, I., and Evan, B.: Metadata for the uninterested. [Online]. Available: http://conferences.alia.org.au/online2011/papers/paper_2011_B4.pdf. [Accessed: 22-February-2014]. (2011)
22. Advanced Distributed Learning (SCORM). [Online]. Available: http://www.adlnet.gov/scorm. [Accessed: 13-May-2013].
23. Currier, S.: Metadata for Learning Resources (MLR): An Update on Standards Activity for 2008 | Ariadne: Web Magazine for Information Professionals. 2008. [Online]. Available: http://www.ariadne.ac.uk/issue55/currier/. [Accessed: 22-February-2014].
24. Hoel T., and Mason, J.: Expanding the Scope of Metadata and the Issue of Quality, in Proceedings of the 19th International Conference on Computers in Education, Chiang Mai, Thailand. (2011)
25. Learning Resource Metadata Initiative :: About the LRMI. [Online]. Available: http://www.lrmi.net/about. [Accessed: 22-February-2014].
26. Dietze, S., Sanchez-Alonso, S., Ebner, H., Yu, H. Q., Giordano, D., Marenzi, I., and Nunes, B. P.: Interlinking educational resources and the web of data: A survey of challenges and approaches, Program Electron. Libr. Inf. Syst., vol. 47, no. 1, 60–91. (2013)
27. Dietze, S., Yu, H. Q., Giordano, D., Kaldoudi, E., Dovrolis, N., and Taibi, D.: Linked education: interlinking educational resources and the Web of data. in Proceedings of the 27th Annual ACM Symposium on Applied Computing, New York, NY, USA, 366–371. (2012)
28. Mitsopoulou, E., Woodham, L., Balasubramaniam, C., Poulton, T., Protopsaltis, A., and Dietze, S., mEducator: multi type content repurposing and sharing in medical education," Acad. Subj. Cent. Med. Dent. Vet. Med. Newsl. 01, vol. 22, 26–28. (2010)
29. Fernandez, M., Aquin, M. d', and Motta, E.: Linking data across universities: an integrated video lectures dataset, in Proceedings of the 10th international conference on The semantic web - Volume Part II, Berlin, Heidelberg, 49–64. (2011)
30. LinkedUp: Linking Web Data for Education. [Online]. Available: http://linkedup-project.eu/. [Accessed: 13-May-2013].
31. Linked Universities :: Home. [Online]. Available: http://linkeduniversities.org/lu/. [Accessed: 13-May-2013].
32. The ARIADNE Foundation, [Online]. Available: http://www.ariadne-eu.org/. [Accessed: 22-February-2014].
33. DCMI Metadata Terms. [Online]. Available: http://dublincore.org/documents/dcmi-terms/. [Accessed: 22-February-2014].
34. LOM to DCAM mapping document. [Online]. Available: http://dublincore.org/moinmoin-wiki-archive/educationwiki/attachments/LOM-DCAM-newdraft.pdf. [Accessed: 22-February-2014].
35. MARC/MODS RDFizer - SIMILE. [Online]. Available: http://simile.mit.edu/wiki/MARC/MODS_RDFizer. [Accessed: 22-February-2014].
36. MIT OpenCourseWare | Free Online Course Materials. [Online]. Available: http://ocw.mit.edu/index.htm. [Accessed: 17-Jun-2013].
37. Kunze, T., Brase, J., and Nejdl, W.: Editing Learning Object Metadata: Schema Driven Input of RDF Metadata with the OLR3-Editor. (2002)
38. Alog-Crisan, R. B., and Roxin, I.: RDF Editor for LOM. in Proceedings of International Conference e-Learning IADIS 2007, Lisbon, Portugal, 2007.
39. Cool URIs for the Semantic Web. [Online]. Available: http://www.w3.org/TR/cooluris/. [Accessed: 22-February-2014].
40. GoodURIs - W3C Wiki. [Online]. Available: http://www.w3.org/wiki/GoodURIs. [Accessed: 22-February-2014].

41. Turtle - Terse RDF Triple Language. 2011. [Online]. Available: http://www.w3.org/TeamSubmission/turtle/. [Accessed: 22-February-2014].
42. Miličić, V.: Problems of the RDF model: Blank Nodes, [Online]. Available: http://milicicvuk.com/blog/2011/07/14/problems-of-the-rdf-model-blank-nodes/. [Accessed: 22-February-2014].
43. Alvestrand, H.: Tags for the Identification of Languages, RFC 1766, Mar-1995. [Online]. Available: http://www.ietf.org/rfc/rfc1766.txt. [Accessed: 22-February-2014].
44. Fallside, D. C., and Walmsley, P.: XML Schema Part 0: Primer Second Edition, W3C Recommendation, 28-Oct-2004. [Online].
Available: http://www.w3.org/TeamSubmission/turtle/. [Accessed: 22-February-2014].
45. Representing vCard Objects in RDF, W3C Member Submission 20 January 2010. [Online]. Available: http://www.w3.org/Submission/vcard-rdf/. [Accessed: 22-February-2014].
46. Europeana Professional - Linked Open Data. [Online].
Available: http://pro.europeana.eu/web/guest/linked-open-data. [Accessed: 18-Jun-2013].
47. GeoNames. [Online]. Available: http://www.geonames.org/. [Accessed: 18-Jun-2013].
48. Rajabi, E., Sanchez-Alonso, S., and Sicilia, M.-A.: Analyzing Broken Links on the Web of Data: an Experiment with DBpedia, Journal of the Association for Information Science and Technology, Vol 65, 1721–1727, doi: 10.1002/asi.23109.
49. Lama, M., Vidal, J. C., Otero-Garcia, E., Bugarin, A., and Barro, S.: Semantic Linking of a Learning Object Repository to DBpedia. 2011. In 11th IEEE International Conference on Advanced Learning Technologies (ICALT), 460–464. (2011)
50. GEMET Dataset. [Online]. Available: http://www.eionet.europa.eu/gemet/. [Accessed: 18-Jun-2013].
51. Eurostat - Linked Data. [Online]. Available: http://eurostat.linked-statistics.org/. [Accessed: 18-Jun-2013].
52. TaxonConcept - Search Species. [Online]. Available: http://www.taxonconcept.org/search-species/. [Accessed: 18-Jun-2013].
53. DataSetRDFDumps - W3C Wiki. [Online].
Available: http://www.w3.org/wiki/DataSetRDFDumps. [Accessed: 18-Jun-2013].
54. Exposing Organic.Edunet eLearning resources as Linked Open Data. [Online]. Available: http://data.organic-edunet.eu/. [Accessed: 13-May-2013].
55. Organic.Edunet Metadata Application Profile - Organic.Edunet. [Online]. Available: http://wiki.organic-edunet.eu/index.php/Organic.Edunet_Metadata_Application_Profile.
[Accessed: 22-February-2014].
56. Bizer, C.: D2RQ - treating non-RDF databases as virtual RDF graphs. In Proceedings of the 3rd International Semantic Web Conference (ISWC2004), (2004).
57. Apache JMeter - Apache JMeterTM. [Online]. Available: http://jmeter.apache.org/. [Accessed: 22-February-2014].
58. Schmidt, M., Hornung, T., Meier, M., Pinkel, C., and Lausen, G.: SP2Bench: A SPARQL Performance Benchmark. In Semantic Web Information Management, R. de Virgilio, F. Giunchiglia, and L. Tanca, Eds. Springer Berlin Heidelberg, 371–393. (2010)
59. Apache Jena - Fuseki: serving RDF data over HTTP. [Online]. Available: http://jena.apache.org/documentation/serving_data/. [Accessed: 13-May-2013].
60. Johnson W. D., and Koch, G. G.: Intraclass Correlation Coefficient, in International Encyclopedia of Statistical Science, M. Lovric, Ed. Springer Berlin Heidelberg, 685–687. (2011)

Appendix: SPARQL Queries

Query 1:Title and description of resources for higher education for 10 resources with the following complexity: Join ordering

```
PREFIX lom:
 <http://data.organic-edunet.eu/lom_ontology.owl#>
PREFIX dcterms: <http://purl.org/dc/terms/>

SELECT ?s ?f ?desc ?r WHERE {
   {?s dcterms:title ?f.}
   {?s lom:description ?d.?d dcterms:description ?desc. }
   ?s lom:educational ?e. ?e lom:educationalContext ?r.
   FILTER regex(str(?r),"highereducation", "i").
}
limit 10
```

Query 2: Resource format category along with count of them for the resources related to organic with the following complexity: Aggregation, Ordering, Join ordering, Search.

```
SELECT ?format (count(?format) as ?count) WHERE {
   {?s dcterms:format ?format.}
   {?s lom:description ?d.
    ?d dcterms:description ?desc.
   FILTER regex(str(?desc),"organic", "i").}
}
GROUP BY (?format)
ORDER BY DESC(?count)
```

Query 3: Title and web address of courses that are in html or pdf formats with the following complexity: Search, Ordering, Join ordering, Optionals with filters, Complex filter conditions

```
Select ?title ?location
WHERE {
 {?s lom:educational ?edu.
  ?edu dcterms:type ?r. Filter Regex(str(?r),"course","i").}
OPTIONAL {?s lom:technicalLocation ?location. }
OPTIONAL {?s dcterms:title ?title. }
OPTIONAL {
       ?s dcterms:format ?format.
       Filter ((?format="application/pdf") ||
(?format="text/html")) .}
}
Order by (?title)
```

Enayat Rajabi is a PhD researcher in the Computer Science Department of the University of Alcalá since 2012. He holds his MSc (2004) degree in Computer Engineering from the department of Ferdowsi University of Mashhad (Mashhad, Iran). He is presently a full time researcher in the Information Engineering research unit at the University of Alcalá (Alcalá de Henares, Madrid, Spain). The aim of his PhD is to interlink educational data using Semantic Web and Linked Data.

Miguel-Angel Sicilia is currently full professor at the Computer Science Department of the University of Alcalá and the head of the Information Engineering research unit, where he leads several European and national research projects in the topics of learning technology and Semantic Web. He is the Editor-in-Chief of the International Journal of Metadata, Semantics and Ontologies, published by Inderscience, and serves as a member of editorial board of many other scientific journals in the area of Semantic Web, Computational Intelligence and Information Systems. He has been active in metadata research and was awarded the 2006 Cyc prize for the best research paper.

Salvador Sanchez-Alonso is an associate professor in the Computer Science Department of the University of Alcalá, Spain. He has participated in and coordinated several EU-funded projects over the last few years related to TEL, learning object repositories and federations, notably eContent+ Organic.Edunet, CIP-PSP Organic.Lingua and VOA3R,. His current research interests include TEL, Learning repositories, Semantic Web and Computer Science education.

Received: March 30, 2014; Accepted: October 01, 2014