An Approach for Description of Open Educational Resources based on Semantic Technologies

Nelson Piedra, Janneth Chicaiza, Jorge López
Escuela Ciencias Computación
Universidad Técnica Particular
Loja, Ecuador
nopiedra, jachicaiza & jalopez2
[@utpl.edu.ec]

Oscar Martínez
Centro de Investigación Operativa
Universidad Miguel Hernández
Alicante, España
oscar.martinez@umh.es

Edmundo Tovar Caro
Facultad de Informática
Universidad Politécnica de Madrid
Madrid, España
etovar@fi.upm.es

Abstract—Open Educational Resources are accessed through the web, whose real setting shows an explosion in the use and development of tools and services based on Social Software. However, the growth of this data repository makes it difficult to find information of value, and reduces the possibilities of sharing and exchanging resources. Using semantic technologies to describe educational resources enables any agent (human or software-based) to process and understand its content (applying inference rules on more structured knowledge). Metadata standards can be used to annotate educational resources; they facilitate their interoperability and discovery. In this work, we propose, OER-CC ontology, for the description of Open Educational Resources under Creative Commons Licenses. This approach is based on standard technology and metadata standards. The ontology could be utilized in higher education institutions (and organizations) to facilitate sharing and discovery of their digital content. This electronic document is a “live” template. The various components of your paper [title, text, heads, etc.] are already defined on the style sheet, as illustrated by the portions given in this document. (Abstract)

Keywords-component; formatting; style; styling; insert (key words)

I. INTRODUCTION

Open Educational Resources (OERs) initiatives promote their global exchange with the aim of increasing the human intellectual capacity. The principal idea of the movement toward Open Educational Resources is that knowledge is a public good where technology in general, and specifically Web, provides an extraordinary opportunity for people to acquire key competences in a knowledge society while they share, use, and re-use digital contents.

OER are teaching, learning, and research digital re-sources and tools (see Figure 1), available on the public domain or that have been released under an intellectual property license, that is, permit their free use, re-use or re-purposing by others.

The next level of educational technology infrastructure will need to use social tools and semantic technologies applied to the Web. The application of Social Web in OER projects has demonstrated that regular users can contribute content without specialist skills; any person can participate actively as an author in the Knowledge Society. By other hand, semantic Technologies enable the power of the semantic web for easy sharing, re-using and educational resources discover.

Figure 1. Open Educational Resources Map

In this paper, we propose, OER-CC ontology, for the description of Open Educational Resources under Creative Commons Licenses. This approach is based on standard technology and metadata standards. The ontology could be utilized in higher education institutions (and organizations) to facilitate sharing and discovery of their digital content.

This document is divided into 3 sections. The first section introduces the models and OER drivers and we put forward options for incorporating semantic technologies in OER description. Next, development process of OER-CC ontology is described; at each stage of ontology creation we argue the choice of each standard, tool and applied language. In the last section, we describe the recovery process of knowledge represented in ontology; several queries were designed to obtain information about the properties of the resources generated in the Computer Science School (CSS) of the
II. DESCRIPTION OF OPEN EDUCATIONAL RESOURCES (OERS) BY MEAN OF METADATA AND ONTOLOGIES

The Open Educational Resources Initiatives are based on the Open Access (OA) movement.

Declarations of support for OA have been developed to accelerate efforts to promote open resources, technology and teaching practices in education. The Declaration of the Budapest Open Access Initiative [1] (BOAI, 2002) was among the first to strongly promote the open and free access to academic and research contents.

In the same way, importance of OER has been promoted by international and national organizations initiatives. Some of these initiatives are: Massachusetts Institute of Technology (MIT) in 2001 with its project OpenCourseWare and the OCW Consortium [2], projects from OECD’s Centre for Educational Research and Innovation [4], UNESCO Initiative through International Institute for Educational Planning [5], the Open University’s Open Learn Project1, Universia MIT OCW – Spain2 and IEEE-Signal Processing Society3 will be work with the open-access repository project Connexions. Other initiatives can be found on the Wiki of the UNESCO OER Community4.

OER Projects can be classified according to model (funding, technical, content and staffing) used to ensure their sustainability. Downes in [3] describes each of the models.

In this paper, we focus on staffing models, that is, by the degree to which participants could actually help with resources. Considering this point of view, there are two models: producer-consumer and co-production [4] (we can find a comparison between them in the Wiki community project, WikiEducator5).

In producer-consumer models (or user-producer), an institution or consortium develop materials and release courseware under an open license, which can be reused by other providers. Within this group the MIT OpenCourseWare can be considered.

Co-production models encourage open and unrestricted participation, aimed to leverage the benefits of mass-collaboration and the principles of self-organization. This model, can promote the collaborative production of common resources.

The co-production strategies to OER production can be deployed using Social Web tools. The Web 2.0 or Social Web is focused in the contents, relationships and knowledge but not specifically in technology. As Will Richardson [6], [7] writes: “The good news for all of us is that today, anyone can become an all life student. (Yes, even you.) Those technologies are easy to use in a way that was not possible in the past”.

The use of wikis is a good example of transition from static and restricted web sites to social and collaborative participation (co-authors). O’Reilly refers to Wikipedia as “a radical experience of confidence” [8]. Wikipedia, blogs and other tools allow the user to publish and then the community determines the relevance and pertinence of the content; mistakes are solved by social correction and with the support of content configuration management (version control).

The core strength of the social software is its easy use, however a OERs co-production model based on this kind of services have some weaknesses; Redecker in [13] identifies the main risks, barriers and obstacles to the implementation of Web 2.0 into teaching and learning practice. Besides those points, we also believe that one of the greatest difficulties to adopt this approach is that the use of social tools is generating an impressive growing in the quantity of contents that are available in the internet. This situation is hindering the organization, retrieval and resources interoperability and its “intelligent” processing is required. Using semantic technologies we can provide a more explicit meaning to information, so that computers can understand and generate new knowledge by applying rules of inference to a better structured knowledge.

A. Incorporating semantics in OER Production

These days, the use of semantic technologies and applications development of the Semantic Web (SW) in the learning context, are the focus of researchers, organizations, universities and consortiums.

Semantic technologies will automate or semi-automate certain educative tasks through “synergy between human and machines” (see Figure 2). Thus is explained by Gruber in [14] “Clearly, there are different roles for people and machines. People are the producers and customers: they are the source of knowledge, and they have real world problems and interests. Machines are the enablers: they store and remember data, search and combine data, and draw mathematical and logical inferences”. With the incorporation of web semantic technologies in processes of creation, storage, retrieval and educational resources remix we expect a reduction in workload of those who run the learning process.

Figure 2. Synergy between human and machines

The Koper’s study [15] showed application areas of the SW in education, he talk about will help to educators to perform some of their tasks such as settings, assessment and
management and courses administration. We also think that the Semantic Web will enable to:

• identify more easily resources with particular properties or that have a kind of specific relationship with others;
• use of semantic rules, it can be a good support for interpretation and argumentation of several issues;
• suppliers of educational contents that use semantic technologies improve the administration, exchange and integration of resources with other information suppliers;
• bear a great potential of providing a robust and extensible base for emerging Social Web applications; interchange, distribution, and reuse of OER can be greatly facilitated by the infrastructures that the Semantic Web offers.

To achieve these objectives, educational resources should be described by a standard schema in a way that any agent (human or software-based) can understand and processing its content

1) Metadata Standards for Educational Material

Al-Khalifa and Davis as said in [24] “Metadata standards are formal specifications used to semantically annotate educational material of any kind. They have been developed to support both machine interoperability (information exchange) and resource discovery by human users”. Some metadata schemes allow the description of resources in XML, RDF, RDFa, or even OWL.

For the learning materials “these various standards and specifications have been developed to meet different requirements and to support the needs of different communities” [16]. Three are most recognized metadata standards:

• The IEEE LTSC (Learning Technology Standards Committee) developed IEEE LOM6 (Learning Object Metadata).
• The IMS GLC (Global Learning Consortium) proposes IMS Learning Resource Metadata7.
• Dublin Core Metadata Initiative8.

The structure proposed by these standards is similar; each one has categories and each category contains a set of metadata. Depending on application, we can choose a subset or extend9 the set with more metadata (the outcome of this adaptation process is called "metadata application profile"). Given the effort required to describe an educational resource using each characteristic proposed by those standards, we could create an profile for describe OERs, considering the particularities of this type of resource, such as:

- Some projects are based on social software, therefore, metadata about the social tools, type of re-source and resource annotations (using social classification systems, such as, folksonomies) are required.
- Open educational resources are licensed under copyleft licenses such as Creative Commons (CC) or GNU/GPL, therefore, metadata about licensing is needed to determine what is permitted, required or prohibited in a work.

2) Ontologies

According to Gruber’s definition [11], an ontology specifies the conceptualization of a specific domain in terms of concepts, attributes, and relationships [12]; also it should be expressed in a _formal language_ “so that a given ontology expression can be interpreted and processed” [18]. Unlike other knowledge structures (e.g. library classification or thesaurus), an ontology allows applying logic, inheritance and other issues [17].

In 1997 emerged the first _language for ontologies_ SHOE10 (Simple HTML Ontology Extensions), so far some languages have been proposed to be more complete and less complex languages than their predecessors. Nowadays the languages most used are RDF and OWL; one of limitations of RDF is that it is not expressive enough to represent complex ontologies, while that OWL extends the possibility of using logical expressions to describe complex concepts and relationships.

The _modeling of domain knowledge_ can be done using traditional paradigms and tools, UML (Unified Modeling Language) or database technologies. But as stated Gomez et al. [19] “…the formalisms used to model the domain knowledge and the languages that implement these techniques limit the kind of knowledge that can be modeled and implemented”, i.e. “only allow the representation of lightweight ontologies”. Consequently to modeling of large or complex domains have been proposed several approaches that allow us develop ontologies (“ heavyweight ontologies”) based on Artificial Intelligence (AI) formalisms; so we have better control over the vocabulary and the semantic. In [19] a comparative study between these three paradigms can be found.

There are several _methodologies_ for building ontologies, two of them could be regarded as more mature a) METHONTOLOGY, proposed by Gómez-Pérez et al. in 1996; and b) On-To-Knowledge, proposed in 2001 by Staab et al.

Regardless of the methodology used to create the ontology, its design should be guided by the Gruber’s principles: clarity, coherence, extendibility, minimal encoding bias and minimal ontological commitment; the idea is to ensure the “knowledge sharing” [21]. To fulfill these criteria, _ontology evaluation_ should be performed at the same time as the development-oriented activities. Different type of evaluation could be made [23]: a) _Syntactic evaluation_, to check languages specification b) _Semantic evaluation_, focused on detecting if the ontologies have inconsistencies and redundancies c) _Lexical evaluation_, it refers to the vocabulary used to represent concepts and domain relationships. Other...
issues related to evaluation are presented in the Brank's et al. [25] paper.

B. Related Work

We have found several works that involve the development of ontologies for a specific issue of e-learning or to modeling knowledge related to a specific problem.

1) Ontology for e-learning systems

Ghaleb et al. in [12] developed an ontology for web-based e-learning system of Qatar University. End-user can annotate learning resources by means of Web form; the system converts this information to a set of RDF statements using the RAP API (Semantic Web toolkit for PHP developers). According to the authors “the ontology can be used for adaptive learning to retrieve the context of a course and to structure the contents”. Their model includes a service of simple semantic search.

This work cannot be reusable in OER domain because it isn’t metadata standards-based (or there isn’t information) and the ontology model hasn’t been published.

2) Application ontologies for assembling learning objects

Santacruz in [26], [22] put forward OntoGlue, “an ontology-based mechanism for assembling learning objects”. In order to assemble two learning objects, the requirements of one must be covered by the competencies of the other. The problem resolved by OntoGlue is the semantic comparison (considering that one requirement and one competence can’t be syntactically equal, but the semantics is the same) among concepts, since mappings between ontologies are established. The standard LOM is used to describe ELOs (Electronic Learning Objects) although an extension that has been included.

3) Metadata Standards for describing Objects and Learning Resources

For semantic representation of Learning Objects. Fermoso et al. in [20] put forward the ontology LOM2OWL11, its structure allows to describe LOs using IEEE LOM standard. Later, we will refer to this work.

For description of Learning Resources. Brown and Thomas explain in [9] metadata approach within the OER project "The Open Polytechnic of New Zealand". They created an application profile that contains necessary elements to ensure metadata interoperability “with both a national and an international audience”, thus, 15 metadata was chosen of those 5 belongs to Educational Category. However, this work doesn’t consider projects social software-based neither the licensing issue.

For description of works licensed under Creative Commons (CC). In 2008, CC published the metadata standard ccREL12 (Creative Commons Rights Expression Language), it that aims to “make licensed works more reusable and easy to find”. To date ccREL metadata, as encoded using RDFa or XMP.

III. DEVELOPMENT OF OER-CC ONTOLOGY

In this section, authors introduce details dealing with the implemented ontology to model the OER and CC domains. Thus, research team tried to achieve the following main goals: (i) to describe OER and CC resources using a common vocabulary by users and producers, i.e., an implemented ontology offered to students and lecturers within an educational context respectively, and (ii) to automate execution of tasks for selected domains such as information retrieval using semantic techniques.

A. Process and tools to build OER-CC ontology

Then, development process of our OER-CC ontology was divided into three general phases:

1) OER ontology development. At this point, we started with an initial ontology (LOM2OWL13) adapted to the metadata requirements for OERs (based on IEEE LOM).

2) CC ontology development. In this case, CC ontology was designed using METHONTOLOGY [19] guidelines.

3) Merging of CC and OER ontologies. Finally, we unified concepts, terminology, definitions and constraints from the two source ontologies.

B. Development process

As introduced before, development process was based on METHONTOLOGY [19] guidelines. This method proposes an ontology building life cycle based on evolving prototypes. That is, it allows adding, changing, and removing terms in each new version (prototype). Next, figure 3 depicts the development process and used tools through each phase respectively.

11 http://www.cc.uah.es/ie/ontologies.html
12 http://wiki.creativecommons.org/CcREL
13 http://www.cc.uah.es/ie/ontologiaLOM2OWL/LOM2OWL.owl
Therefore, we started the development process creating concept maps for both knowledge domains respectively. For this purpose, we depicted main points of OER-CC ontology with a general study of references to describe metadata of selected domains. As a consequence, we evaluated some well-known references to describe OER metadata like IEEE-LOM [30] and Dublin Core (DC) [31]. At this point, we understand that IEEE-LOM introduces enough metadata to characterize our educational purposes for this work. In addition, IEEE-LOM was used because we think that it includes large amount of structured metadata into different categories, allowing greater semantic flexibility for our selected domains. Another important issue was that it is a leading standard within e-learning environments [16] usually.

Then, all metadata associated with CC licenses was represented considering the ccREL standard [10] which provides information related to Licenses Properties, e.g., permissions, prohibitions, requirements and general metadata. Once we had decided how to describe metadata, we selected well-referenced tools to model the OER-CC ontology successfully. Concretely, we used the following:

- **CMapTools** [28] was used to represent and interconnect both knowledge domains, i.e., OER and CC respectively. In addition, **CmapTools Ontology Editor** [14] (COE) was used for constructing, sharing and viewing modeled ontology based on CmapTools. Besides, this allowed preparing OER-CC ontology to be modeled using **OWL** language.

- **Protégé** [29] was used to implement our ontology formally. For this purpose, all prior results were imported to this ontology editor. Then, we used **SWRL** to provide deductive reasoning capabilities and **SPARQL** to retrieve the educational resources metadata.

Each tool and language was applied at different stages of process.

Additionally, each prototype of OER-CC ontology was validated through evaluation items like: (i) Syntactic evaluation, by mean of Pellet reasoner [19] and syntax checker used for this work, (ii) Taxonomy evaluation, at this point we validated standard rules in taxonomies like inconsistency, incompleteness and redundancy in concepts respectively. Once we finished the development process successfully, we were ready to deploy OER-CC ontology. Therefore, next section introduces details that give preliminary evidences of our contribution in this work.

**IV. OER-CC ONTOLOGY INSTANTIATION AND RETRIEVAL**

Then the formalization and implementation activities are carried out, in this last section, we attempt to provide a piece of the true potential of ontologies use to represent domains in the context of Education and particularly to describe OERs, i.e.

- The machines can understand and interpret meaning of educational content, thus facilitate or automate execution of certain tasks, such as, accessibility or retrieval with the contents of both domains.

- Users and OER producers [19] can use a common vocabulary to describe their resources; it enables to reuse and share application domain knowledge.

- Represent a metadata structure through ontology; it will enable interoperability among different do-mains (vocabularies). This promotes accessibility and OERs sharing between universities and organizations around the world.

One of the ways we use to demonstrate that potential is defining and executing queries and inference rules on OER-CC ontology to retrieve and to exploit its knowledge. Before discussing the results, we explain the filling the ontology with resources metadata produced in a higher institution.

**A. Instantiation process and tools**

Once the ontology is built, the process of ontology population starts, and then it allows extracting relevant information from the resources hosted by ontology. The population process “consists in creating a knowledge base containing instances of the ontology concepts and instances of the ontology relations” [19].

This process can occur in 3 ways, depending on the level of automation is required or that it can be applied. This process can occur in 3 ways, depending on the level of automation is required or that it can be applied.

1. Manual. This mechanism may be viable when we have a small number of instances for each concept and it can be done through ontology development tools and tool suites; they provide basic support for the ontology population so that ontology developers can create instances with their ontology editors [19].

2. Semi-automatic, applying automatic learning techniques on sources of unstructured or data semi-structured.

3. Automatic, creating specific applications (using ontology management APIs) to generate in-stances from structured data sources and directly insert them in each concept.

For the OWL ontology population, Protégé offers certain plugins.

- **DataMaster** [20], to import schema structure and data from relational databases into Protégé.

14 http://coe.ihmc.us/groups/coe/
15 http://www.w3.org/TR/owl-ref/
16 http://www.w3.org/Submission/SWRL/
17 http://www.w3.org/TR/rdf-sparql-query/
18 http://clarkparsia.com/pellet/
19 The role of users and OERs producers change depending on adopted production model. In a user-producer model, students consume re-sources and institutions generate them; while a co-production model anyone participant (including students and self-learner) can contribute to creating them
can be constructed by applying filters an educational resource (title, tags to different types of Creative Commons Licenses with jurisdiction instantiation dealing with CC domain was considered using in Ecuador26.

ontological model has been implemented a process is required and classes are created according to this structure. When the ontological model has been implemented to move instances from the new classes to existing classes.

B. OER-CC Ontology Instantiation
To instantiate the OER-CC ontology, we considered some resources generated in the Computer Science School (CSS) of the Technical University of Loja (Universidad Tecnica Particular de Loja23, UTPL-Ecuador).

UTPL are using social software for knowledge management, learning and OER development based on a co-production model. From 2008, the School has opened completely all the training courses of Engineering free of charge; this has enabled the exchange of experiences internally and with other academic units.

A manual process has been developed to populate the OER-CC ontology with the CSS’ educational resources, because Protégé plugins import schema structure and data from structured sources (unlike what is required in this work since we already have the model). Concretely, those resources are videos and keynotes authored by teachers and students basically. In addition, they are available on youtube24 and slideshare25 of the School’s channels also. Furthermore, instantiation dealing with CC domain was considered using different types of Creative Commons Licenses with jurisdiction in Ecuador26.

C. Information Retrieval
As introduced in section 3.2 we used SPARQL language to retrieve knowledge represented at our OER-CC ontology. Thus, we developed some queries to retrieve information, e.g., properties of OERs and CC licenses.

What is more, our initial results show that our OER-CC ontology has been deployed successfully, i.e., using metadata from initial instantiation of educational objects.

Some queries were designed to extract information from OER-CC ontology, geared mainly to determine its use in tasks of recovery, accessibility and OERs re-mix.

Table I and Table II list some competency questions that can be answered by the ontology. SPARQL panel of Protégé was used to run queries, the next step is to develop or integrate applications (software agents or web services) that implement the above activities.

As shown in Table I, one of the advantages of using OER-CC ontology and the IEEE-LOM standard is that we can complete information for each resource metadata in various languages and thus facilitates multilingual search.

<table>
<thead>
<tr>
<th>Table I. Competency Questions for OERs Retrieval</th>
</tr>
</thead>
<tbody>
<tr>
<td>More specific searches can be constructed by applying filters (assembling different types of combinations) on different OERs properties.</td>
</tr>
<tr>
<td>Multilingual search. If multilingual descriptions of metadata were specified</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Competency questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>¿What is general information of an educational resource (title, description, keywords, and other metadata)?</td>
</tr>
<tr>
<td>What educational resources are labeled with a particular word? Such questions are especially common in a co-production model in which multiple annotators can add, in a simple way, tags to each resource. The tags could be used to improve search results as proposed in [27].</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table II. Competency Questions for Accessibility and Remix</th>
</tr>
</thead>
<tbody>
<tr>
<td>We can exploit the information on licensing, contribution and attribution to facilitate accessibility and OERs re-mix</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Competency questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>What are the resources created by a particular contributor and what type of contribution have done?</td>
</tr>
<tr>
<td>What are the direct contributions of a university in the OERs production and what are indirect contributions through their teachers / students?</td>
</tr>
<tr>
<td>Who should be given attribution and under what conditions we can make use of OERs?</td>
</tr>
<tr>
<td>What are all contributors (individual and organizational / institutional) of a specific OER?</td>
</tr>
<tr>
<td>What kind of licenses we use to can generate educational resources in a country and under what conditions these can be remixed?</td>
</tr>
<tr>
<td>How can access (URI or URL) to an OER and how it can display (depending on the type of resource)?</td>
</tr>
<tr>
<td>What educational use and in what context we can use an OER?</td>
</tr>
</tbody>
</table>

Next, we introduce a implemented query (to answer the question: what are the direct contributions of the UTPL in the OERs production and what are indirect contributions through their teachers/students, see Figure 4) within OER-CC ontology.

These queries results confirm that a short-term improvement in information retrieval task can be achieved considering educational resources generated at the Computer Science School of the UTPL.

21 http://protegewiki.stanford.edu/index.php/XML_Tab
22 http://protegewiki.stanford.edu/index.php/Excel_Import
23 www.utpl.edu.ec
24 http://www.youtube.com/eccutpl
25 http://www.slideshare.net/eccutpl
26 http://creativecommons.ec/
CONCLUSION

In this work, we have given initial evidences that machines could understand and interpret semantic meaning of educational contents.

As a main contribution, we have implemented OER-CC ontology to model knowledge dealing with OER and CC domains respectively. Therefore, we have been able to infer knowledge using OER-CC ontology, i.e., thorough instantiation and classification of educational objects respectively. About work in progress, research team continues working to improve OER-CC ontology deployment. Currently, we are involved with more methods to promote accessibility according to OER user requirements; e.g., the content of an educational resource could be ranked within the OER-CC ontology according to student / lecturer profile respectively.

Furthermore, this work has enabled us to implement ontologies from concept maps (process used to generate CC AL09-PID-36, UTPL and scholarship provided by the Secretaría Nacional de Ciencia y Tecnología (SENACYT) of Ecuador.

Our thanks to eMadrid Consortium. Authors are members of GICAC-UPM that is part of eMadrid consortium.

ACKNOWLEDGMENT

Support for this work was provided by UPM under project AL09-PID-36, UTPL and scholarship provided by the Secretaría Nacional de Ciencia y Tecnología (SENACYT) of Ecuador.

REFERENCES


[31] DC The Dublin Core Metadata Initiative http://dublincore.org/