Journal of Universal Computer Science, vol. 21, no. 5 (2015), 679-711
submitted: 15/12/14, accepted: 10/3/15, appeared: 1/5/15 © J.UCS

Seeking Open Educational Resources to Compose Massive Open Online Courses in Engineering Education An Approach Based on Linked Open Data

Nelson Piedra
(Universidad Técnica Particular de Loja, Loja, Ecuador
nopiedra@utpl.edu.ec)

Janneth Chicaiza, Jorge López
(Departamento de Ciencias de la Computación y Electrónica
Universidad Técnica Particular de Loja, Loja, Ecuador
[jachicaiza, jalopez2]@utpl.edu.ec)

Edmundo Tovar
(Universidad Politécnica de Madrid, Madrid, Spain
etovar@fi.upm.es)

Abstract: The OER movement has tended to define "openness" in terms of access to use and reuse educational materials, and to address the geographical and financial barriers among students, teachers and self-learners with open access to high quality digital educational resources. MOOCs are the continuation of this trend of openness, innovation, and use of technology to provide learning opportunities for large numbers of learners. In the last years, the amount of Open Educational Resources on the Web has increased dramatically, especially thanks to initiatives like OpenCourseWare and other Open Educational Resources movements. The potential of this vast amount of resources is enormous. In this paper an architecture based on Semantic Web technologies and the Linked Data guidelines to support the inclusion of open materials in massive online courses is presented. Linked Data is considered as one of the most effective alternatives for creating global shared information spaces, it has become an interesting approach for discovering and enriching open educational resources data, as well as achieving semantic interoperability and re-use between multiple Open Educational Resources repositories. The notion of Linked Data refers to a set of best practices for publishing, sharing and interconnecting data in RDF format. Educational repositories managers are, in fact, realizing the potential of using Linked Data for describing, discovering, linking and publishing educational data on the Web. This work shows a data architecture based on semantic web technologies that support the discovery and inclusion of open educational materials in massive online courses in engineering education. The authors focus on a type of openness: open of contents as regards re-use and re-mix, i.e. freedom to reuse the material, to combine it with other materials, to adapt and to share it further under an open license.

Keywords: OER, OCW, MOOC, Linked Data, Linked OER Data, Integration, Reuse, Openness
Categories: L.1.1, L.1.3, L.1.4, L.3.0, L.3.2

1 Introduction

Different projects have emerged from the Open Access philosophy and have contributed to facilitate online learning. UNESCO believes that universal access to high quality education is the key to the building of peace, sustainable social and economic development, and intercultural dialogue. The International Covenant on Economic, Social and Cultural Rights (Article 13.1), recognizes “the right of everyone to education” [ICESC, 1976]. The Education for All is a worldwide UNESCO programme that involves more than 150 countries, and a large number of non-governmental organizations. World Education Forum 2000 in Dakar, Senegal, resulted in the Millennium Declaration and the 2000 Dakar Framework for Action, which made global commitments to provide quality basic education for all children, youth and adults [UNESCO, 2000]; and, the 2003 World Summit on the Information Society, Declaration of Principles, committing “to build a people-centred, inclusive and development-oriented Information Society where everyone can create, access, utilize and share information and knowledge”.

In this context, in 2001, the Massachusetts Institute of Technology published the first courses open, the MIT OpenCourseWare [MIT, 2001]. The term “Open Educational Resources” (OER) was first adopted at UNESCO’s 2002 Forum on the Impact of Open Courseware for Higher Education in Developing Countries funded by The William and Flora Hewlett Foundation [UNESCO, 2002a]. In this year, UNESCO defined Open Educational Resources as "technology-enabled, open provision of educational resources for consultation, use and adaptation by a community of users for non-commercial purposes". OER are part of the Open Education movement, and teachers, students, self-learners, and learning institutions are driving its development. OER are typically made freely available over the Web or the Internet. Their principal use is by teachers and educational institutions support course development, but they can also be used directly by students and self-learners. OER include learning objects such as lecture material, references and readings, simulations, experiments and demonstrations, as well as syllabi, curricula and teachers' guides [UNESCO, 2002b].

The purpose of the Open Educational Resources movement is to provide open access to high quality digital educational resources. There is broad participation by universities, private organizations, and others. Projects include the Internet Archive (see http://internetarchive.org), Project Gutenberg (see http://gutenberg.org), Wikipedia (see http://wikipedia.com), Creative Commons (see http://creativecommons.org), Sun Microsystems Global Education Learning Community (see https://edu-gelc.dev.java.net/nonav/index.html) and, as is the focus of this article, the OpenCourseWare Consortium (see http://ocwconsortium.org). The list of participating organizations grows every year as the principles of openness spread. Currently, the arrival of Massive Open Online Courses (MOOCs) and the growth of open and online education – Open Educational Resources (OER), Open Course Ware (OCW) - are increasingly focusing on self-learners as the primary target group.

In the last years, the amount of Open Educational Resources on the Web has increased dramatically, especially thanks to initiatives like OCW and other OER movements. The potential of this vast amount of resources is enormous but in most cases it is very difficult and cumbersome for users (teachers, students and self-learners) to visualize, explore and use these resources, especially for lay-users without experience on searching technologies [Tovar and Piedra, 2014].

The Semantic Web is a collaborative movement led by the World Wide Web Consortium (W3C). Tim Berners Lee, the creator of the Web, coined the term [Berners-Lee, Hendler and Lassila, 2001]. The Semantic Web promotes common data formats for publishing content on the World Wide Web, by encouraging the inclusion of semantic content in Web pages. The objective is to convert the current Web, dominated by unstructured and semi-structured documents, into a "Web of Linked Data."

The purpose of this paper is to present an approach based on Semantic Web technologies [Bizer, Cyganiak and Heath, 2007] to support the inclusion of open materials in massive online courses. The authors focus on a type of openness: opening of contents which allows alteration i.e. freedom to reuse the material, to combine it with other materials, to adapt it, and to share it further under an open license [Hilton et al. 2010; Hodgkinson-Williams and Gray, 2009]. This approach will enable people to discover and access open educational resources that are extracted from open distributed repositories. Our principal OER providers are OCW institutions. In this context, we opted to apply the principles of Linked Data [Heath and Bizer, 2011; Berners-Lee, Hendler and Lassila, 2001] to integrate, interoperate and mash up data from distributed and heterogeneous repositories of open educational materials. The purpose is to significantly improve discovery, accessibility, visibility, and to promote reuse of open educational content in massive courses. [Piedra et al., 2014a]

2 Improve the reuse of Open Educational Resources by mean of Semantic Technologies

2.1 "Openness" in terms of access to use and re-use educational materials

The OER movement has tended to define "openness" in terms of access to use and reuse educational materials, and to address the geographical and financial barriers among students, teachers and self-learners [Petrides et al. 2008]. MOOCs are the continuation of this trend of openness, innovation, and use of technology to provide learning opportunities for large numbers of learners. MOOCs, promote unprecedented massive access to the world-class education that has so far been available only for a select group of few students. MOOC initiatives emphasize free access and interactive features rather than static content, the dominant message is of the quantity of access rather than the openness of educational resources for use, re-use, adaptation or repurpose. [Piedra et al. 2013]

The Open Definition sets out principles that define "openness" in relation to data and content. It makes precise the meaning of "open" in the terms "open data" and "open content", and thereby ensures reuse, interoperability and integration between different initiatives of open educational. It can be summed up in the statement that: "A piece of data or content is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and/or share-alike" [OKFN, 2012].

The above Open Definition gives full details on the requirements for 'open' data and content in the Context of Open Educational Initiatives. Open data are the building components of open knowledge. Open knowledge is what open educational resources and data can be freely used, reused, remixed, adapted and redistributed by anyone.
Therefore, the content and data must be available over the Internet as a whole and at no more than a reasonable reproduction cost; and the data must also be available in a convenient and modifiable form. Moreover, the content and data must be provided under terms that permit reuse and redistribution including the intermixing with other OER collections. The data must be machine-readable.

In the context of MOOCs, "Open" refers to providing a learning opportunity to a wide number of participants around the globe regardless of their geographic location, age, income, ideology, and level of education, without any entry requirements, or course fees to access high quality education. In the context of MOOCs, openness does not refer to providing open educational materials.

In this regard, an Open Educational initiative should be considered open if it is made public in a way that observance with the guides: completeness, primary, timeliness, ease of access to digital resources and metadata, metadata documented, metadata in Standard and machine readability Formats, universal Participation, formats non-proprietaries, ensures interoperability between different collections of OER using open licenses both a human-readable description and computer-readable metadata, and persistence.

2.2 Learning supported in Open Educational Resources

Different social and economic connotations have produced the Open Access movement; specifically in the learning field the following benefits can be highlighted: i) it facilitates people to access and use different learning resources, ii) it promotes the universal right to education, mainly trying to reach communities and individuals who, for economic resources or even time reasons, cannot access to a traditional education system, iii) it improves the key skills that individuals need to participate in the knowledge society (such as self-direction and information management), and iv) it promotes the openness, sharing and creation of knowledge. Specifically, the content and Open Educational Resources have the potential to improve substantially the quality of life of learners worldwide [Caswell et al., 2008].

One of the first educational resources that were put into open was the MIT OpenCourseWare (OCW) offered by Massachusetts Technological Institute in 2001. One of the missions under which this project was conceived was to provide free, virtual and non-commercial access to MIT courses to educators, students and self learners around the world [MIT, 2001]. The next aim was to create a flexible movement based on an efficient model that other universities could emulate at the moment of publishing their own educational materials, generating synergies and spaces of collaboration.

One year later, the Open Educational Resources (OER) term was introduced in order to include a wide range of learning objects and free applications, from whole course, open access journals, to lecture material, references and readings, simulations, experiments and demonstrations, as well as syllabi, curricula and teachers' guides [UNESCO, 2002b]. OERs are typically freely available over the Web. Their principal use is by teachers, students and self-learners.

The main purpose of Open Educational Resources (OER) movement is to provide open and free access to high quality digital learning materials. For this reason, its effects or implications on higher education are well known. On the other hand, looking at the interest of people accessing open educational resources, we have identified that a large group of users are not linked to an institution, they are self-taught or they are immersed in a process of non-formal or informal learning. According to a study by the OpenCourseWare Consortium (now, the Open Education Consortium) which results were published on March 2013, over 40% of the users with access to courses OCW type are self-learners and professionals [OCWC, 2013]. Therefore, the OER movement could act as a catalyst to promote the universal right to education. [Caswell, 2008]

A concept that has emerged from the OER movement is Massive Open Online Course (MOOC). Dave Cormier coined this term in 2008. MOOCs represent the next stage in the evolution of open educational resources. Unlike OER model, MOOCs promote training scenarios at large-scale participation and open access via the web. MOOCs are a progression of the kind of open education ideals suggested by open educational resources.

A MOOC is a course designed to be an all-inclusive learning experience, and it is based on a wide blend of traditional tools, such as video lessons, assessment activities and final project combined with Web 2.0 tools [Alario-Hoyos et al., 2014] Also, MOOC is a model for delivering learning content online to any person who wants to take a course, with no limit on attendance. Though the design of and participation in a MOOC may be similar to college or university courses, MOOCs typically do not offer credits awarded to paying students at schools. However, assessment of learning may be done for certification.

Accredited institutions are now accepting MOOCs as well as free courses and experiential learning as partial credit towards a degree. Students do not pay fees to the content provider for basic enrolment in the course, nor do they receive credit from the content-providing institution. Social networking, interactive services, and automated grading or peer assessment are provided by the platform provider, as is a nominal certificate for the completion of assignments.

2.3 Issues of the openness and reusability of OERs

Open Educational Resources (OER) provide a strategic opportunity to improve the quality of education as well as facilitate policy dialogue, knowledge sharing and capacity building. [UNESCO, 2012] The open educational resources movement is growing in the higher education environment. And although teachers contribute and share their knowledge and experience [Caswell et al., 2008], some key aspects of the opening of knowledge related to the use, reuse, adaptation and discovery are still unresolved.

From the point of view of the openness, in a previous work [Piedra et al., 2014b] we have discussed about the true meaning of Open of MOOCs: they are free as in gratis (without paying), but they are not open in the sense of being reusable of openly accessible, i.e. students are forced to sign up and get access to the course; so many researches highlight that it would be much more useful to have complete access at all times and reuse elements in other courses. On the other hand, MOOCs may be considered open in the sense of “free to try”; they are not offered under an open license. Any use of the content or services for academic credit-bearing purpose is restricted and requires payment to the MOOC provider.

From the point of view of the designing, making a MOOC from scratch is not an easy task because it involves several issues of logistical, technological, pedagogical and financial nature that educators must face [Alario-Hoyos et al., 2014].

In this context, it makes sense to reuse the thousands of Open Educational Resources that are available on the Web: materials from OCW (lessons, quizzes, syllabus, readings), multimedia or presentations contained in different OER repositories. OER reusability means that the content is relevant to the specific needs of a user, which is technologically accessible and that it is sufficiently open for use, re-use [Abeywardena, 2012], re-mix, adapt and re-distribute.

Before OER can be reused, the openness of content can be measured in terms of the rights a user of the content is granted. One of the primary benefits of an OER is that it can be discovered [White and Manton, 2011] and adapted to the needs of specific situations.

The OER should be designed to be easily adaptable for other users. It should have metadata enough to discover and process. However, the reality is that there are difficulties on mixing, comparing, classifying and interoperating metadata resources. And as mentioned in [Frango-Silveira et al., 2005] in order to provide a certain degree of external reusability of learning resources among repositories, some interoperability issues must be resolved.

Up to this date, most OER data are collected in heterogeneous and distributed repositories, such as OER Commons[1], OCW initiatives[2], Merlot[3], Serendipity[4], OpenCulture[5], and other OER repositories, where data is annotated using different metadata mechanisms (e.g. IEEE LOM[6], ADL SCORM[7], custom metadata schemas), and retrieved by ad-hoc mechanisms, individual Web APIs/Services or other mechanisms (e.g. OAI-PMH[8]); however, these technologies are limited because the data cannot be dereferenced.

Heterogeneity is also evident in the data or descriptions of OER. One of the reasons why OER could stay hidden and therefore to be underutilized is that each institution and producer of this kind of resources, labels them using tags or informal and heterogeneous knowledge schemes. This problem was identified in [OCWC, 2013], where respondents noted that one way to improve the OCW is to make a "major better categorization of courses according to subject areas".

---

1 Open Educational Resources Commons: http://www.oercommons.org

2 Open Education Consortium: http://www.oeconsortium.org/ and UNIVERSIA: http://ocw.universia.net/es/

3 Merlot: http://www.merlot.org/merlot/index

4 Serendipity: http://serendipity.utpl.edu.ec

5 Open Culture: http://www.openculture.com

6 IEEE Learning Object Metadata (LOM): http://ltsc.ieee.org/wg12/

7 Advanced Distributed Learning (ADL) Sharable Content Object Reference Model (SCORM): http://www.adlnet.gov/capabilities/scorm

8 Open Archives Initiative – Protocol for Metadata Harvesting: http://www.openarchives.org/pmh/

In previous works, the authors present the Linked OpenCourseWare Data project (LOCWD) [Piedra et al., 2014a], which published metadata of courses coming from different open educational datasets. Among metadata that was extracted are: title, description, author, institution, tags and subjects. So far there are over 4000 indexed courses and 700.000 OERs associated to 626 unique category names or knowledge fields, many names correspond to similar areas written in different ways or different languages and also correspond to different levels of detail. The semantic lack in the relations between areas and subjects make it difficult to find associations between topics and to list recommendations about courses for self-learners.

As a solution to the problems described, our proposal is to combine the description of OERs with Linked Data approach in order to improve the integration of repositories and materiales. This would lead to a new generation of OERs (described in machine-readable formats), that would facilitate automatic processing tasks. In this work, the proposal is to find and merge OERs into a great variety of learning programs, i.e. made-custom courses according to a MOOC profile is presented in Section 3. Below in the next section, related works in this area are presented.

2.4 Main approaches to enhance the reusability of learning material

The need to reduce the workload for educators during the creation of learning material and the need to provide personalized learning paths according to styles and preferences of learners have been addressed in some works.

The Wiley's concept of learning object, presented in [Koper, 2003], highlights three characteristics related to its reusability: online availability, reproducibility and addressability. In this sense, the OERs extracted from OCWs meet these features and can be reused in different contexts. Koper (2003) explores some underlying issues in the reuse of learning resources and presents these within the context of a teacher and an instructional designer who wish to reuse resources within their own practice. Moreover, Kellar et al. (2004) highlights the importance of the fact that a reuse environment includes a component that models the user’s profile and thus supports the dynamic composition of objects into personalized content.

From the technological point of view, the system based on Adaptative Hypermedia (AH) and the intelligent tutoring systems (ITS) have been used to create courses based on existing educational material. However, as noted in [Brusilovsky and Nijhavan, 2002] the approaches based on these technologies "are building around a close corpus material. Collecting and preparing this material to use in adaptive systems is an expensive process. Thus these systems can’t directly benefit from existing repositories of learning material.

One of the reuse environments of educational material is provided by the ARIADNE Foundation [Klerkx et al., 2010] The courseware re-use framework of ARIADNE allows a course author to search for the learning objects in repositories of educational material and include them in their courses [Brusilovsky and Nijhavan, 2002]. Despite the large amount of resources indexed by ARIADNE, updated content could not be found and services of recommendation based on particular needs have not been located.

Brusilovsky and Nijhavan (2002) suggest a courseware-reuse approach named KnowledgeTree. It is a framework for adaptive e-learning based on distributed re-usable learning activities, the most recent version has already been used in several courses at the University of Pittsburgh. The framework allows the presence of multiple portals, activity servers, and user-modeling servers. However, the service of resource discovery has not been addressed in the current version of KnowlegdeTree; that is one of the core components of our proposal.

Kellar et al. (2004) describe an architecture that supports the dynamic composition of Web based lessons based on a database of learning objects tagged according to the IMS Metadata. A prototype has been developed specifically for an e-learning environment (users of health informatics learning modules). The scalability of this proposal can be discussed.

Other proposals found in the reuse of educational material have been defined for closed corpus, i.e. repositories of learning objects previously that have been described and classified by metadata schemas as LOM or IMS.

Instead, the OERs are published on the Web in an open and extensible repository where each person can share resources; therefore, an open architecture that captures new resources and classify them according to different criteria is required to provide more relevant resources when developing MOOCs.

Regarding web technologies and linked data, there is an increasing use of topic classification and annotation of digital resources. The Semantic Web approach is about adding formal structures and semantics (metadata and knowledge) to Web content for easy access, management, discovery and integration, to make the resources machine-understandable. Some proposals such as [Cano et al., 2013; Husby and Barbosa, 2012] are based on the use of repositories of linked open data to determine the topics that describe social content as micropost or blogs. These and other studies have found similar findings, "DBpedia resources are a good starting point to define keyword meanings due to the fact that a huge part of the knowledge base is related to classes in the DBpedia Ontology." [Cano et al., 2013]

This work explores how to reuse, integrate and interoperate isolated OER repositories using Semantic Web Technologies.

3 The Web as source of data and resources

The last stage of the Web is the Semantic Web. According to the W3C, the Semantic Web is a Web of Data. Data can be in different formats, languages, styles and structures. This approach of the Semantic Web is aligned with the original vision that Berners-Lee had the Web in the late 1980s in which the meaning of information plays a key role and the information is stored in a global database, distributed and linked data through the web. Unlike the current Web of linked documents, the Semantic Web is a Web of linked data that can be used for describing data models, concepts and data properties. Furthermore, connect, query and recombine data from the Web, as if they were simply part of a global database. These advances can be a way to support interoperability, accessibility and reusability of all types of data.

Linked Data is the way that the Semantic Web has to link data that are distributed on the Web, so that they are referenced in the same way they do the links of the web pages

Semantic Web technologies and, more precisely, Linked Open Data (LOD) are changing the way information is stored, published and exploited. The term "Linked Data" refers to a set of best practices for publishing and connecting structured data on the Web [Heath and Bizer, 2011; Berners-Lee, Hendler and Lassila, 2001]. Linked data is mainly about publishing structured data in RDF using URIs rather than focusing on the ontological level or inference. OER provided with Linked Data (Linked Open Educational Resources Data) supports the process of discovery, reuse, integration and interoperability of open educational materials.

The W3C's Semantic Web provides a common framework namely Resource Description Framework (RDF) for describing resources on the Web. With RDF, automated software can store, exchange, and use machine-readable information distributed throughout the Web, in turn enabling users to deal with the information with greater efficiency and certainty; also, RDF data can be shared and reused through application, enterprise, and community boundaries.

RDF is based upon the idea of making statements about resources (in particular web resources) in the form of subject-predicate-object expressions. These expressions are known as triples in RDF terminology. The subject denotes the resource, and the predicate denotes features or aspects of the resource and expresses a relationship between the subject and the object. Uniform Resource Identifiers (URIs) are used to identify these resources. RDF Schema (RDFS) is to represent the web resource and SPARQL (Standard Protocol for RDF Query language) is to extract information from RDF graphs for machine understandable representation.

The Linked Data Design Issues, outlined by Tim Berners-Lee back in [Berners-Lee, 2006], provides guidelines on how to use standardized Web technologies to set data-level links between data from different sources [Heath and Bizer, 2011]. Linked data is an opportunity to mitigate complexity in OER reuse. These Linked Data Design Issues, in OER context, are:

1. Use URIs as names for things, which can be unambiguously identified (e.g. OERs, courses, MOOCs, OER creators, OCW providers or knowledge areas).
2. Use HTTP URIs so that people can look up those names. With the aid of URIs, the corresponding OER data and relevant interlinked data can be dereferenced.
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL) to describe linked OER data, which are machine-readable and repurposed to serve the proposed architecture to enhance integration with reused and interoperated OER data.
4. Include links to other URIs, so that they can discover more entities. Linked Data—particularly data available using open licenses—has an important role to play on information systems and could be a key feature for Open Education based on OER data on the Web of Data.

Linked Data suggests that the value and usefulness of data increases to the extent that these are related to other data. For this reason, Linked Data uses the Web to create different types of links between data from different sources. With this vision of the Web, data and relationships have fundamental roles. The availability of sources based on the principles of Linked Data enable new opportunities for exploiting knowledge representation techniques, information extraction and integration and multi-agent environments, among others.

In [Piedra et al. 2014a], authors apply the Linked Data Design Issues to explore, visualize and use information that is semantically related to open educational resources that are accessible via the OCW Consortium. Linked data has the potential of creating bridges between OCW data silos. The authors demonstrate that OCW resource metadata can be enriched using datasets hosted by the Linked Open Data cloud. Additionally, the Linked OER and OCW Data environment enabled us to discover and reuse open educational materials.

In the architecture that we describe below, we use linked data, sources of social knowledge and services from Web 2.0 to find OERs that can be incorporated in the design of MOOCs. In the two following sections, we mention the most important sources from the perspective of the Semantic Web and the Social Web.

3.1 Data and knowledge on the Web

The most important part of the Linked Open Data is DBPedia[9]. DBpedia is a crowd-sourced community effort to extract structured information from Wikipedia and make this information available on the Web. The Linking Open Data (LOD) Cloud[10] currently provides access to hundreds of datasets in various areas such as Media, Geography, Publications, Government, and Life Sciences. As a consequence of Linking Open Data community project, datasets in a wide range of domains are now semantically described and connected to each other.

Moreover, sources of social knowledge are being used to enrich with data different Web entities as people, organizations, products and knowledge categories. In some works the use of DBPedia is addressed to annotate, to enrich and to classify content. In this way, the DBPedia ontology enables a broad coverage of entities in the world, and allows entities to bear multiple overlapping types; it includes RDF data derived from Wikipedia; each resource is harvested from a Wikipedia article (whose multilingual content is maintained by thousands of volunteers-editors). In addition DBPedia resources are linked to other linked data sources and ontologies such as Geonames, YAGO, OpenCyc, and WordNet, providing more semantic information in the form of relations such as typeOf and sameAs. [Cano et al., 2013; Muñoz-García et al., 2011]

3.2 Resources highlighted by people

In these days much of the information available on the Web is published on social media, represented through social networks such as Facebook, Twitter, naming only the most prominent. Each of the media and social networks has its own scheme of operation and different working characteristics, ranging from the length of text that can be used, the use of different forms to identify topics until reaching the reciprocity of relationship between the participants. For example Twitter is a social network where millions of daily messages called Tweets are exchanged, within the message can be used labels, called hashtags, to identify the subject of the message, in addition the message may also include links to other resources that expand the original content, or showing interesting information, and the relationships between users are represented as non reciprocal relationships named “Following”.

---

[9] DBPedia: http://dbpedia.org/About

[10] LOD Cloud. “Linking Open Data cloud diagram, by Richard Cyganiak and Anja Jentzsch. http://lod-cloud.net/”

In the same context, the information exchanged in social networks can be used for learning, so that Open Educational Resources can be obtained from Twitter, represented as links; find experts that in Twitter are represented as popular users; virtual communities in Twitter user lists; events are described through “hashtags” on Twitter.

One of the main challenges is the extraction of information posted on social networks, but that can be overcome with the use of various technologies, such as linked data that allows retrieving resources and link with other external sources, graphs databases that help represent the working scheme of a social network and social network analysis (SNA) as a technique to discover relevant information that goes beyond the individual properties.

4 Architecture of OER-reuse

The vision of Semantic Web is the idea of having data on the Web described and linked in a way that it can be used by machines not just for display purposes, but for automation, interoperability, integration and reuse of data through various applications and contexts. It provides a promising platform for Open Educational Initiatives. In this section, we describe our proposal for the integration and reuse of OER: architecture based on Linked Data technologies.

The Figure 1 shows an information-filtering framework used to identify a set of N items (Open Educational Resources) that will be of interest to certain massive course designers. The framework uses a Knowledge Organization System (SKOS) for filtering of items. SKOS[11] is a popular ontology to organize knowledge. In 2009, SKOS reached Recommendation status at W3C. In Semantic Web ecosystem, SKOS (Simple knowledge organization system) [W3C, 2009] is used for representing mapping relationships among systems. It provides a standard way to represent knowledge organisation systems using RDF to describe concept schemes such as thesauri, classification schemes, subject heading lists, taxonomies and other types of controlled vocabulary, thus guaranteeing interoperability among applications. In the Datahub site, 149 datasets was found for “format-skos”[12]. In SKOS, the elements of a thesaurus are represented by means of concepts among which there are established hierarchic relations.

The user-based and model-based collaborative filtering approaches are the most successful technology for building recommender systems. The basic assumption in these algorithms is that there are sufficient historical data for measuring similarity between items or users. However, this assumption does not hold in various application domains such as open educational environments where new resources or courses are introduced or are custom made. Each resource is unique and there are very few duplicate items. In this domain, the probability of the same exact two resources offered together is close to zero.

In this context, we discuss the challenges of providing a resource list to reuse or remix in the OER domain where no sufficient historical data exists for measuring

---

11 http://www.w3.org/TR/skos-reference/

12 http://datahub.io/dataset?q=format-skos

similarity between resources or users. We present this approach, which overcomes the limitations of the existing ones.

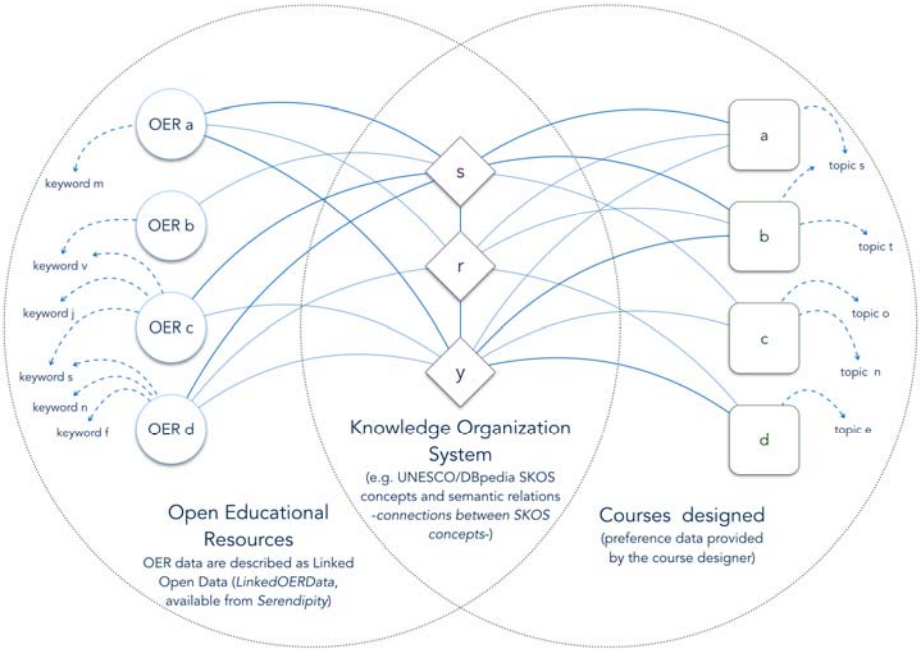

Figure 1: SKOS-based recommendation system

The main objective of this work is to propose a linked OER data architecture (Figure 2), able to adapt, reuse and re-mix OERs (stored in LOCWD repository and others extracted from a social network) in the MOOC context. The architecture is composed by seven services which have been designed to carry out this task in a collaboratively way. This linked data architecture enables us to ask questions and solve open educational problems across a heterogeneous and distributed information landscape extending beyond the traditional boundaries of each OER contributors.

Our approach is based on identifying distinctive features with the help of MOOC preferences and resources needs data. As with all recommender systems, the main goal is to help users to find information or resources and match information that is important about needs with information that is important about resources. Figure 2 summarizes the architecture in a general model of OER recommendation for MOOC Designers. Accordingly, the process can be broken down to the following steps:

1. OER collecting.
2. OER Metadata quality assurance.
3. Generation and Publication of Linked OER Data.
4. Contextualization, classification and enrichment of OER.
5. Seeker of resources (selector of items from OER universe based on SPARQL).
6. Getting course preference data and attributes.

7. Resources collecting, transformation and graph loading from social network.
8. OER discovering via social network analysis.
9. OER Filtering

4.1 Component 1: OER collecting

Goal: Identify and select OER repositories, then extract metadata and educational resources with Open Licenses
Description: The initial step is to identify and select the OER repositories that are available in the Web. There is a large amount of unstructured data of an OCW resource available on the Web, but only in a human-readable representation, HTML.

Most OCW web sites do not have APIs for data consumption. So, the only other alternative for automatically reconstitute the underlying data from an OCW web site is to use web-scraping techniques. [Piedra et al. 2014a]

Examples of extracted OER properties include the name of the resource, its creation date, abstract, keywords, information about creator, language, open license information, format, MIME type, expected study duration, expected level of difficulty, and so on. On the other hand, content metadata corresponds to the properties of the knowledge and skills designed, such as learning objectives, learning pathways, and examinations.

4.2 Component 2: Generation and Publication of Linked OER Data

Goal: Development and delivery of open educational resources data as Linked Data.
Description: Linked Data design principles are increasingly employed to publish and consume heterogeneous datasets in a distributed way. Data is still locked up in applications. The technical problem with today's most common information architectures is that content, metadata and schema information are not separated well from application logics and presentation layer. Data cannot be re-used as easily as it should be.

Using Linked Data design issues, developers can query Linked Data from multiple sources at once and merge it without the need for a single common schema that all data shares. Linked data technologies can also help to integrate the work of disperse institutions producing diverse linked data.

The following is an outline for producing Linked Data in OER context:

1. Visioning project scope.
2. Identify data providers and select heterogeneous repositories.
3. Model vocabularies for OER domain.
4. Data extraction from OER repositories.
5. Data cleaning.
6. Generate OER data as Linked Data.
7. Publish linked data.
8. Consume and display linked data.

Vocabularies and ontologies provide the mechanism to organize the Web information in a structured way. The web contents can be understood by the computer as well as by human beings. Piedra et al. (2014a) described Linked Open Educational

Resources Data vocabulary[13], for open educational resources with the aim to describe the specific types and classes of resources in OER and OCW domain. LOERD reuses a set of RDF(S) vocabularies. Each vocabulary includes a set of terms and classes that are common to a particular knowledge domain. The aim of these vocabularies is to connect the described OER domain with Datasets in the LOD cloud.

LOCWD is a RDF(S) vocabulary devoted to linking OERs, open licenses, OCW repositories, and other academic information using the Web. Different kinds of applications can use or ignore different parts of LOERD. With LOERD, the OER/OCW initiatives can retain some control over their information of materials and courses in a non-proprietary format.

4.3 Component 3: OER metadata quality assurance

Goal: Assure metadata quality.
Description: Metadata enables users to find the OER they require, therefore it is an important component of this approach. There will always be some aspects of the metadata that are inaccurate, inconsistent or out of date, even in OER providers which have extensive quality assurance procedures in place and have invested heavily in the creation of good quality metadata for OER.

OERs are highly heterogeneous, and different metadata schemes appear to reflect attributes assigned by different OER communities. Accordingly, this component is designed to operate in an environment of considerable and probably increasing diversity OER providers. The quality assurance in this component is guided through these questions:

- Which metadata attributes (e.g. license, title, description, type, etc.) are present in each OER provider?
- Who created the data (machines, humans, both)?
- What percentages of the total number of OERs have each metadata?
- How consistent is the metadata within those attributes?
- What patterns can be detected?
- When was the metadata created, updated or deprecated?

As part of this component have been executed the following tasks:

- Detect controlled vocabulary values and attribute those values to our vocabulary.
- Detect and fix common metadata errors.
- Deprecate values that provide no information value.
- Add new attributes to the OER, such as the addition of topics

4.4 Component 4: Contextualization, classification and enrichment of OER

Goal: support semi-automatic classification of Open Educational Resources, taking advantage from linked data available in the Web through systems made by people who can converge to a formal knowledge organization system.

13 About Linked OER Data Vocabulary: http://purl.org/locwd/schema#

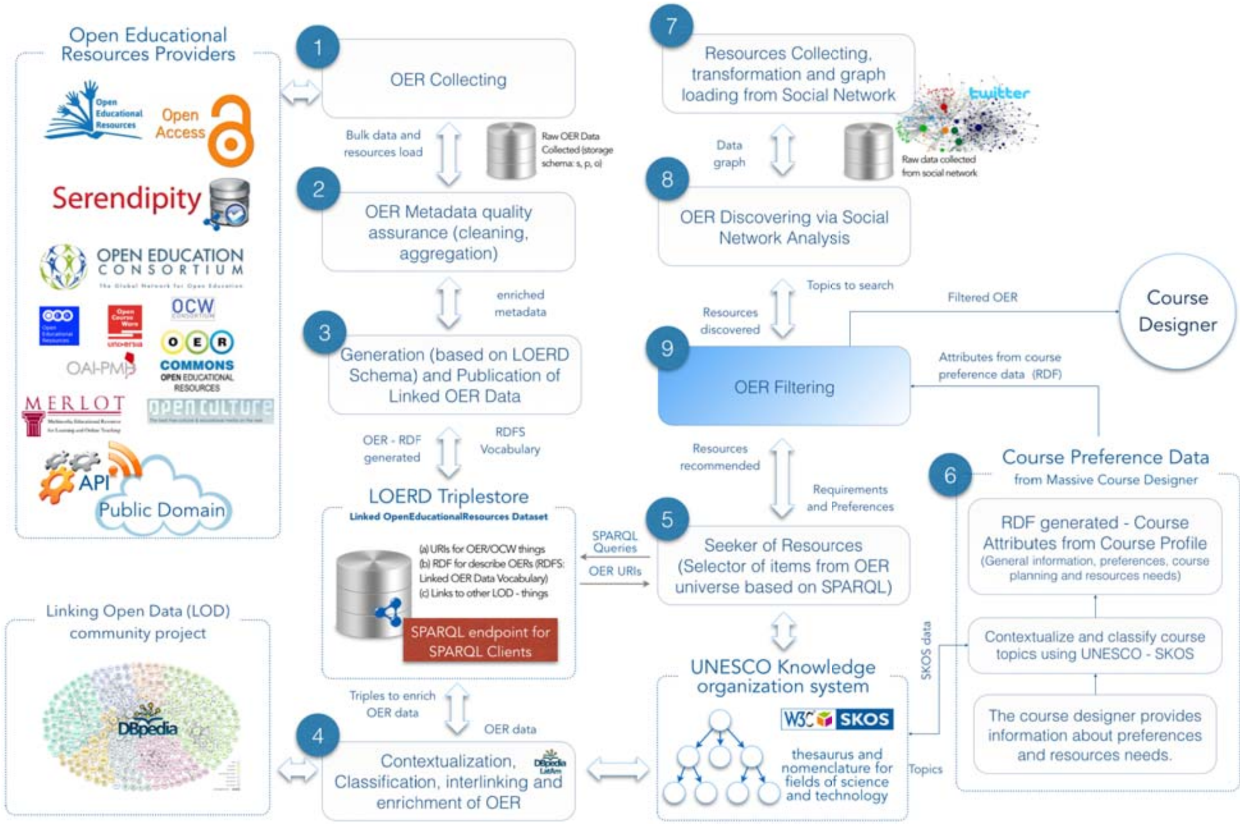

Figure 2: Architecture for Seeker of Educational Materials from Linked Open Educational Resources Data (LOERD) Dataset

Description: Identifying the main topics and concepts associated with an OER is a task common to many educational applications including classification, retrieval and recommendation or educational materials. Classification systems of knowledge, such as thesaurus, have traditionally been used to improve the organization and retrieval of documents. It provides a standard way to organize knowledge organisation systems using RDF to describe concept schemes and ensure interoperability among applications. In the context of this proposal, SKOS is used for representing mapping relationships among courses and OERs. The properties *skos:broader* and *skos:narrower* are used to assert a direct hierarchical link between concepts from a formal knowledge organization system.

In this phase, OERs must be described, classified and characterized so that those could be suggested to the course designers. In the Figure 3, we explain the workflow proposed to classify OERs that takes advantage of data, relations and structure of directed graph of social knowledge sources.

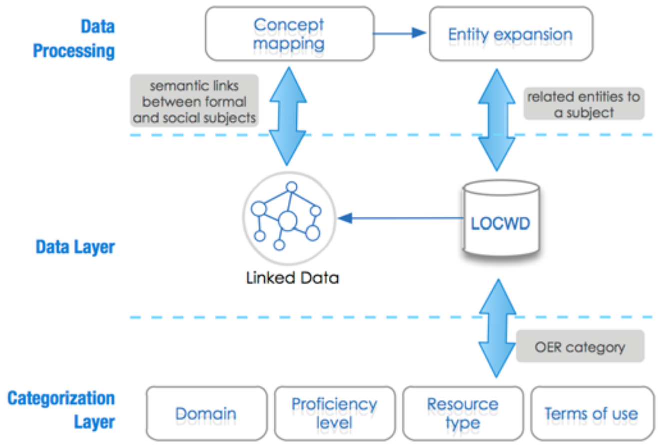

Figure 3: General workflow for contextualizes, classify and categorize OERs

1. Data Processing. At this stage as seen in Figure 4, two tasks are performed in an offline way: i) mapping between controlled-formal classification system and open-social organization system, and ii) expansion of main concepts through categories existing in social data sources.
2. Concept or topical mapping: To achieve the semantic interoperability amongst different collections or schemes, it is necessary to create links between equivalents items. During the creation of links between entities it must face key challenges as: name variation, entity ambiguity or absence of data. For the first challenge, a previous task of pre-processing text is needed; to solve ambiguity issues, it will be necessary to obtain or to provide additional information. As for the third aspect, it can happen due to the fact that some sub-disciplines of the chosen thesaurus have lost force; therefore, it does not produce major impact, if it is not possible to find the equivalent resource.

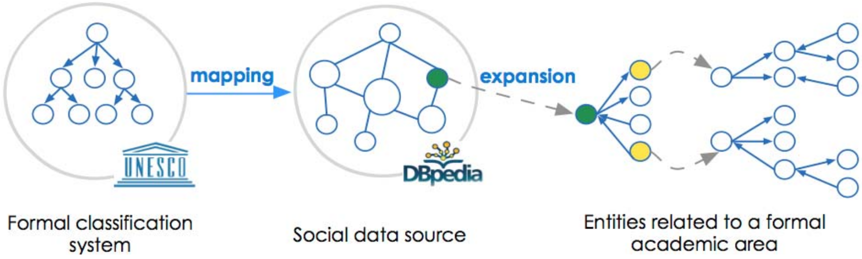

Figure 4: Tasks of data processing: mapping and expansion

In this work, we chose the system of 6-digit of the nomenclature UNESCO. Since 2013, there is an implementation of this thesaurus according to SKOS, which groups the different categories of the nomenclature in a single scheme of concepts (*skos:ConceptScheme*). On the other hand, the selected target scheme is DBPedia; an approach of multiple sources to enrich, disambiguate and get better results can be adopted [Cano et al. 2013]. From possible associations found, semantic relations were created, which are stored in a semantic repository. In the future these equivalences will be used to categorize OERs.

3. Entity expansion. Next, to find entities related (e.g. topics or knowledge subjects) to each of the sub-disciplines of the formal system, the spreading activation method is applied. The spreading of the activation occurs from a node to their adjacent nodes. This spreading activation algorithm looks for semantically related candidate concepts, which reinforce each other. The process of activation and spreading will be performed, at the most, until the third level.

Category	**Range of possible values**
Knowledge area	Each field of the chosen thesaurus (e.g. UNESCO Nomenclature)
Proficiency level	Introductory, Intermediate, Advanced
Type of resource	Assignments, Resource, Coursework, Lecture Note, Syllabus, Learning Guide.
Rights for use and reuse	According to the specification of Creative Commons Licenses[14]: Share — copy and redistribute the material in any medium or format. Adapt — remix, transform, and build upon the material. For any purpose, even commercially.

Table 1: Categories for OER organization

4. A weight has been associated to each semantic relationship. This weight is used to sort the retrieved nodes. To finish this task, descriptive

14 http://creativecommons.org/licenses/by/4.0/

attributes of every entity (e.g. label, abstract and comment in several languages) are retrieved. The data obtained in this stage has been used to annotate and connect OERs; also, the extracted information is stored in the knowledge base.

5. Categorization Layer. At this point, the goal is to organize the OERs according to at least 4 categories: i) knowledge area according to a formal classification system, ii) proficiency level that allows achieve, iii) type of resource, and iv) legal rights for use and reuse. In Table 1, it can be seen a set of possible values that can take each of these categories.

4.5 Component 5. Seeker of resources

Goal: Merge the functionalities of recommendation seeker and Course Profile.

Description: The architecture uses an approach based on SPARQL to express preferences and resources needs by rating OERs. The system focuses on SPARQL query-based algorithms for matching OER based on MOOC preferences and weighting the interest of MOOC designer with similar taste to produce a recommendation for the resources seeker.

This module is designed to access two data sources: The first one reline on LOCWD data, which provides RDF data extracted from OCW and OER websites. The second one uses the LOD Project, particularly from DBPEDIA, which provides RDF data extracted from the info-boxes of Wikipedia pages in a structured way.

4.6 Component 6. Getting course preference data and attributes

Goal: Define a course profile that serves as a view of filter onto the whole universe. The course designer provides information about preferences and resources needs.

Description: The idea is evolving into a more interoperable and integrated system to share, connecting and discovering data and metadata of MOOC profiles. Users don't know precisely what they can find on OER site, or what to search for. Self-learners are trying to discover relationships or trends between MOOC profile and OER data.

In this stage, we focus on semantic or conceptual annotation of courses, which consists in attaching “semantic labels” to a resource or parts of a course using semantic features provided by a formal knowledge organization called ontology or semantic vocabulary. The use of ontologies in the annotation process of items has a particular benefit: the queries for information retrieval and annotated resources share the same vocabulary.

A semi-automatic semantic or conceptual annotation of resources (OERs and/or Courses) is proposed to bridging the gap between the ontology formal language and user’s natural language terms or keywords. DBpedia is used as a controlled vocabulary for annotates OERs.

In Figure 5, the metadata of MOOC profile is showed. In this work, authors believe that the key property to find relevant OERs is skos:subject . The relationship subject, defined by SKOS vocabulary, binds a resource to each of the concepts that describe it.

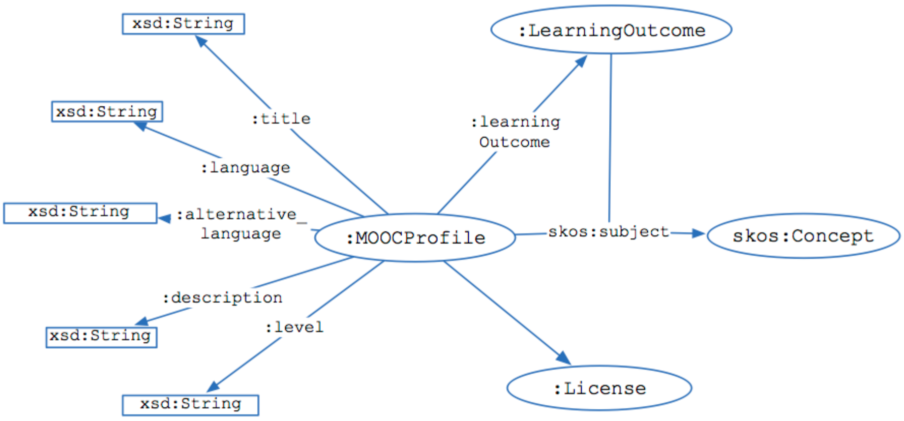

Figure 5: MOOC Profile

To enrich a MOOC profile, the course should be annotated with SKOS concepts, so semantic relations between SKOS concepts can be used to obtain related topics. In this work, SKOS properties are used to infer hierarchical links, to annotate resources and access direct or indirect hierarchical links between course and concepts (see Figure 6.

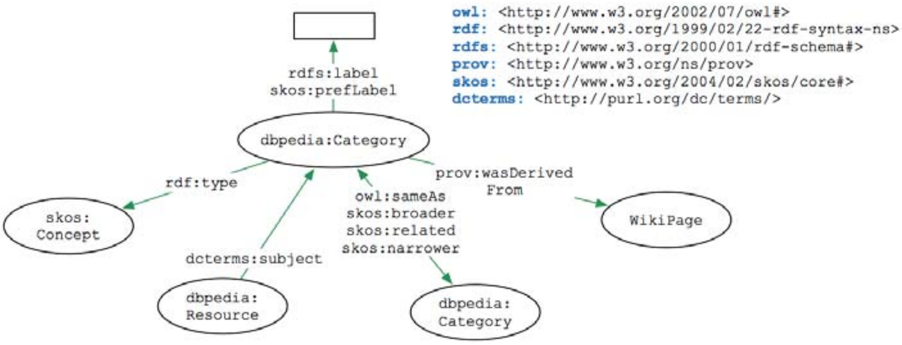

Figure 6: Key resources used for OER annotation and enrichment from DBPedia

Several relations link resources to their concepts, such as “owl:sameAs”, “skos:broader” (a hierarchical link between two concepts which indicates that one is in some way more general -broader- than the other -narrower-), “skos:narrower”, “skos:related” (an associative link between two concepts indicates that the two are inherently "related", but that one is not in any way more general than the other) is used to assert an associative link between two SKOS concepts, and “dcterms:subject” to describe the topic of the resource (typically, the subject will be represented using keywords, key phrases, or classification codes. The properties skos:broader and skos:narrower are used to assert a direct hierarchical link between two resources using SKOS concepts.

In the architecture proposed, it is necessary to make use of both direct and indirect hierarchical links between concepts, to improve search recall through query expansion. For this purpose, the properties skos:broaderTransitive and skos:narrowerTransitive are provided.

4.7 Component 7. Resources collecting, transformation and graph loading from Social Networks

Goal: Obtain potential OERs from social media networks such as Twitter. **Description**: The amount of digital resources is variable in time. Tasks executed is presented in Figure 7 and are described below:

1. TAW Crawler: This component recover information through of search using the API that Twitter offers in an open, free of charge way but with certain policies, and limits both the number of daily requests and the number of results. Once registered the search criteria that can be hashtags, words or exact phrases, the crawler begins to gather information and determine, based on the amount of information collected, to get the optimal execution time of each query to comply with the policies imposed by Twitter and to try to capture recent tweets. The information collected on this point is called raw data and no further processing as described.
2. Data normalization: With the data collected from Twitter, processes of harvesting and structuring are executed, in order that the information is ready for discovery tasks. The information is classified as concepts: tweets, users, hashtags, mentions, re-tweets and URLs, as well as their relationships: a user posts a tweet, a tweet containing hashtags and URLs; a tweet has mentions and re-tweets.
3. Extension of shortened URLs, on Twitter is common for all URLs published are shortened by any of the services that exist today (t.co , bitly.com , goo.gl , etc.). The expanding process, takes the URLs, and through of the protocol HTTP executes requests (get) and receive response codes, as well as other data such as MIME type referenced by the URL. This process is important because it allows: i) determine the status (valid or not) and the type of the resource (MIME type), of the URL; ii) disambiguate URLs, because the same URL can be shortened by different services and each one producing different shortened URLs iii) clean the URLs of the tags that are used in marketing campaigns, for example UTM tags are widely used with Google Analytics through the sharing buttons that can also cause problems of ambiguity.
4. URL enhanced: Once the URLs valid are identified, is necessary to enrich the information available to this moment (state and MIME type). This enrichment is done with those resources whose data type MIME is html or xhtml and try to extract the following information: title of web page (HTML tag <title>), and at the level of meta-tags: description, keywords and language <meta name=”description” content=...>, <meta name=”keywords” content=...> and <meta name=”language” content=...>

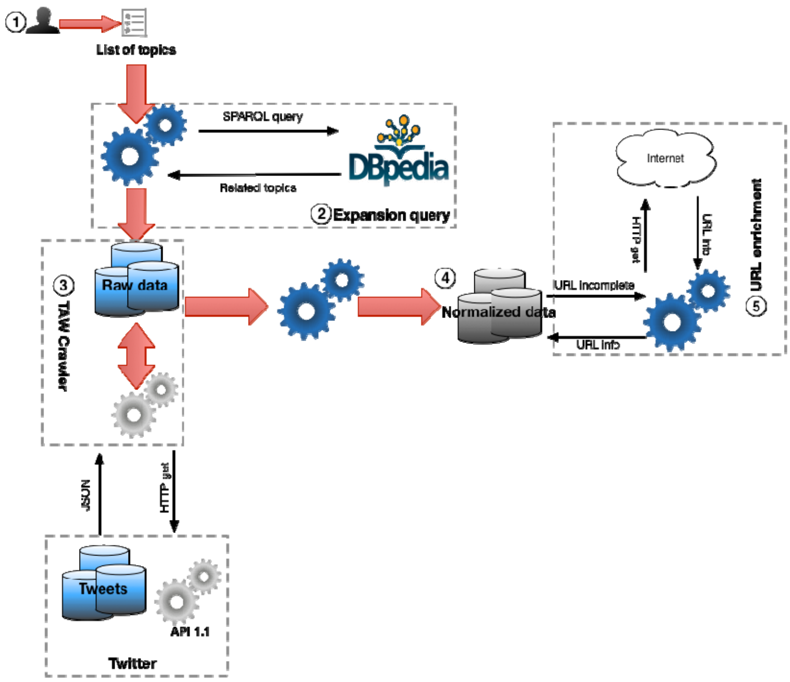

Figure 7: Workflow for the OER discovery based-on SNA

This information was used for two purposes, firstly to improve the presentation of results and use it to find resources that meet the search criteria.

4.8 Component 8. OER discovering via social network analysis

Goal: Discover OER applying techniques of social network analysis to find influential users whose publications are considered interesting by a group of users that interact in a knowledge domain.

Description: This component is a combination of technologies of Web of Linked Data and social network analysis (SNA); the first with the goal of making an expansion of the search criteria, while the second was used to find relevant information through the relationships between users, who write posts about the subjects, which are of our interest and we need OERs. Tasks executed is presented in Figure 7 and are described below:

1. Topic list: The process starts with the need to find useful OERs that have been posted on Twitter. This need should be expressed as a list of topics that proceed from course attributes.
2. Expansion query: With the topics identified, the next step is to extend with the aim of finding information of each of them, do not only literal way, but to extend through related topics, stepping beyond of the syntax of words to the semantic thereof. DBPedia was used for the contextual

enrichment of a Tweet's entities by providing information that can help to disambiguate the role of a given entity in a particular context. This enrichment is based on a developed technique for deriving semantic meta-graphs from different sources.

This process is beneficial as it allows collect much information that in turn allows getting more resources. To accomplish this goal a set of SPARQL queries was built, one for each topic, with the goal of seeking the concepts related to each topic, these queries run against the DBPedia allowing find related topics through the predicates related the topic identified as an object within an RDF triple in DBpedia.

Once identified the topics along with their related topics, the following steps can be summarized in the collection and exploitation of information to discover OERs.

4.9 Component 9. OER Filtering

Goal: For our purpose, a preference is an individual mental state concerning a subset of items from the universe of alternatives. Users can use the architecture because a single taxonomic order or a single folksonomy is not suitable or sufficient for explorer OER resources.

Description: The architecture proposed attempts to recommend OERs that are similar to educational resources planned by the MOOC designer and others records of social activity, such as OER Discovering via Social Network Analysis and OCW Syllabus.

1. OER discovery using social network analysis: After running the previous component, a set of processes is executed to find related hashtags, influential users, and finally find OERs to be recommended. Processes that run here have as data sources tweets with valid URLs and are:

   Hashtags graph: The nodes of this graph (undirected) are hashtags published in a tweet and a link between two nodes indicates a co-occurrence of hashtags in a tweet. In each of the complete graphs, the Betweenness metric that allows us to find the most influential users in both retweets and mentions networks is applied.

   The end result is a network that displays the topics associated to the topic of search and that gives user an updated overview of the topic. The metric to find related topics is Betweenness, which shows those hashtags that are links to other information that the user can use to focus their interests.

   Network of users: two networks are used here, the first is a directed network with nodes that are members involved in a relationship retweet, the source node is the author of a tweet and the target user is the name of the author who he makes a retweet. The second network is similar to the above but with the difference that the destination node is the name of the user mentioned.
2. OERs Retrieval. The results obtained are a set of recommendations on users (identified as experts), virtual communities (lists of Twitter users), OER (tweets with links) and related events, according to the learning needs described as tags. The message also may include links to other

resources that expand the original content or showing interesting information. Much of the information exchanged in social networks can be used for learning.

5 Validation: construction of a MOOC using OER

In this section, a proof of concept for the proposed architecture is described. The process has been driving from the syllabus of a real MOOC, "Intro to Java Programming, Building Programs with Classes & Objects" offered by Udacity (see syllabus in https://www.udacity.com/course/cs046).

5.1 Description of the MOOC Profile

Purpose: Building a Java course from open educational materials as an alternative to an existing MOOC.
Title of course: "introduction to Java Programming"
Level: Beginner (undergraduate)
Main language: English
Alternative language: Spanish
Knowledge Area: Computer Science
Topics: Introduction to computers, programming languages, algorithms, and the java programming environment, data types, variables, operators classes and objects, graphics, data types, decisions structures, loops and Iterations, arrays, arraylists and simple array algorithms, methods, inheritance.
Existing courses:

- Udacity: Intro to Java Programming, Building Programs with Classes & Objects. Type: MOOC. Built by San José State University. Instructor Cay Horstmann. This course has over 175,000 students enrolled. https://www.udacity.com/course/cs046.
- OCW-MIT: Introduction to Programming in Java. Course Number: 6.092. Type: OpenCourseWare. Built by MIT. Instructor(s): Evan Jones, Adam Marcus, Eugene Wu. http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-092-introduction-to-programming-in-java-january-iap-2010/
- METU-OCW: Object Oriented Programming with Java I. Instructor: Altan KOÇYİĞİT. http://ocw.metu.edu.tr/course/view.php?id=23

5.2 Creation and enrichment of the MOOC Profile

The aim of the proof of concept is to design an introductory course about Java, whose content is based on the reuse of OER. Table 2 shows the course profile in RDF.

Course as RDF resource	**Property**	**Value**
:JavaCourse	rdf:type	:MOOCProfile
:JavaCourse	:title	Java Fundamentals
:JavaCourse	:description	This course will cover the main concepts about Java. Students will learn the fundamentals of Java. The focus is on developing high quality, working software that solves real problems.
:JavaCourse	:language	English
:JavaCourse	:alternative_language	Spanish
:JavaCourse	:level	Basic
:JavaCourse	:requirements	The course is designed for students with some programming experience
:javaCourse	:learningOutcome	Learners will be guided through the fundamentals of object-oriented programming on the Java platform.
:JavaCourse	:relatedConcept	"Java"
:JavaCourse	:relatedConcept	"Programming"
:JavaCourse	:Syllabus	Introduction to computers, programming languages, algorithms, and the java programming environment, data types, variables, operators classes and objects, graphics, data types, decisions structures, loops and Iterations, arrays, arraylists and simple array algorithms, methods, inheritance.

Table 2: Some properties for describe a MOOC profile for Java Course in RDF triplets

Figure 8 shows the semantic graph that describes the main properties of the course profile design.

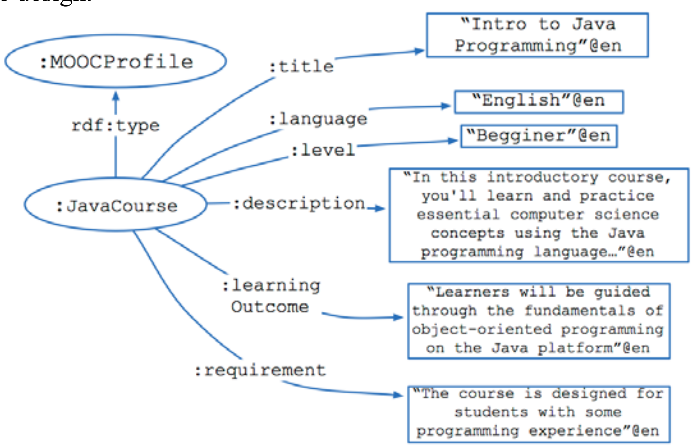

Figure 8: Semantic Graph for describe course profile

The course consists of nine lessons or units that are presented in Figure 9. Each lesson was mapped to a knowledge topic. Each of the topics is represented as a SKOS concept (see Figure 9).

Each topic of the course is linked to a set of subtopics and related concepts. The course designer must identify the most appropriated concepts, as well as the context for each of these topics.

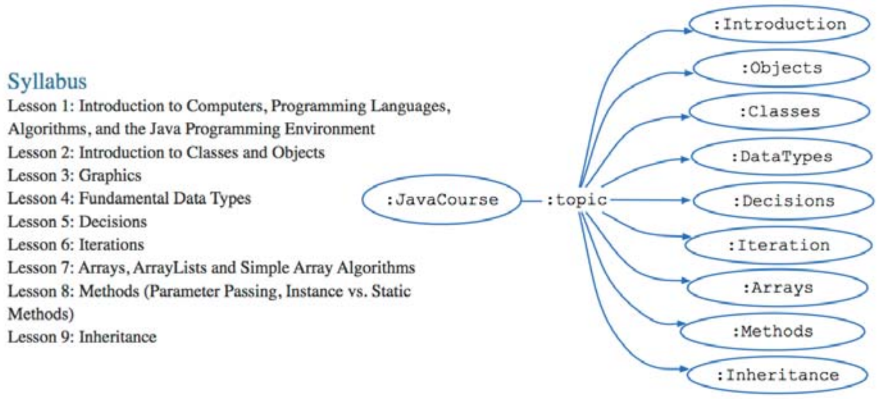

Figure 9: Main topics related to the course

For example, for the topic "inheritance", the course designer has selected two elements as context: the topic "Object-oriented_programming" and the topic "Java language"; and describes the level of relationship through associating a weight to each of the relationships (value between 0 and 1).

```
:JavaCourse :topic  :Inheritance .
:Inheritance       rdf:type       skos:Concept .
:Inheritance       rdf:label      "Inheritance"@en .
:Inheritance       skos:related   :Interface .
:Inheritance       skos:related   :Polymorphism .
:Inheritance       skos:related   :Cast .
:Inheritance       skos:related   :Hierarchy .
:Inheritance       skos:related   :SubclassInJava .

:Inheritance       :contextNode   :Context_Inheritance_OOP .
:Inheritance_OOP :context :Object-oriented_programming.
:Inheritance_OOP :weight  : "0.8"^^xsd:float.
:Object-oriented_programming      rdf:type       skos:Concept .

:Inheritance       :contextNode   : Context_Inheritance_JL .
:Inheritance_JL :context  :Java_language.
:Inheritance_JL :weight   : "0.8"^^xsd:float.
:Java_language     rdf:type       skos:Concept .
```

The next step is to associate external URIs to the topics in the course profile through the process of annotation. These annotations are represented using RDF. The initial course profile is enriched with data extracted from Wikipedia (webpages URIs describing each topic) and DBPedia resources. In Table 5, the annotation made in

RDF shows the topic "Inheritance" linked to Wikipedia and DBpedia resources using the predicate skos:related.

skos:Concept	Wikipedia page title	DBPedia Resource
:Inheritance	Inheritance (object-oriented programming)	db:Inheritance_(object-oriented_programming)
:Inheritance	Generalization	db:Generalization
:Inheritance	Object-oriented design	db:Object-oriented_design
:Inheritance	Object-oriented programming	db:Object-oriented_programming
:Inheritance	Polymorphism (computer science)	db:Polymorphism_(computer_science)
:Inheritance	Parametric polymorphism	db:Parametric_polymorphism
:Inheritance	Bounded quantification	db:Bounded_quantification
:Inheritance	Method (computer programming)	db:Method_(computer_programming)
:Inheritance	Type conversion	db:Type_conversion
:Inheritance	Comparison of Java and C++	db:Comparison_of_Java_and_C%2B%2B
:Inheritance	Comparison of C Sharp and Java	db:Comparison_of_C_Sharp_and_Java
:Inheritance	Object (computer science)	db:Object_(computer_science)
:Inheritance	Hierarchy	db:Hierarchy
:Inheritance	Class-based programming	db:Class-based_programming

Table 3: List of semantic resources related to Inheritance keyword

5.3 Selection and collection of open educational materials

Scope: 15 associate consortia, as well as 212 higher education institutions and 57 organizational members compose OCWC; all courses are available for adoption and adaptation by faculty and students around the world.
When publishing an OER Repository, the providers are sharing digital educational resources with an attached open license that allowing others to reuse, adapt and share their work. Ideally, this resource when combined with others OERs provides great value.

On this paper, we focus on finding OER published by open licenses and useful to MOOC or Open Course Ware production. Authors selected and extracted information from 80 heterogeneous OCW repositories from OCWC and OCW-Universia members [Piedra et al., 2014a], sifting through a total of 30,000 OCW courses and 90.000 OERs approx. Data scraping was used to extract data from OCW platforms that was later structured and stored in a database. Scraping eliminated the need for having to do the retrieval manually.

In the context of this paper, authors opted to apply the design issues of Linked Data to integrate, interoperate and mash up OER data from MOOC Designer requirements. In Piedra et al. (2014) the components: 1) OER collecting and 2) Generation and Publication of Linked OER Data were presented. From OER data published as linked data, the fourth component of the framework (contextualization, classification, and enrichment of OER) was implemented based on [Chicaiza et al. 2014].

5.4 OERs Annotation and categorization

The annotation process not only provides access to a large amount of structured data sources but also enables machines and software agents to automatically analyse this semantic knowledge.

In this work, we put forward the categorization of OER according to the disciplines of the Computer Science field defined by the UNESCO nomenclature. Figure 8 shows an example of semantic annotation of OERs.

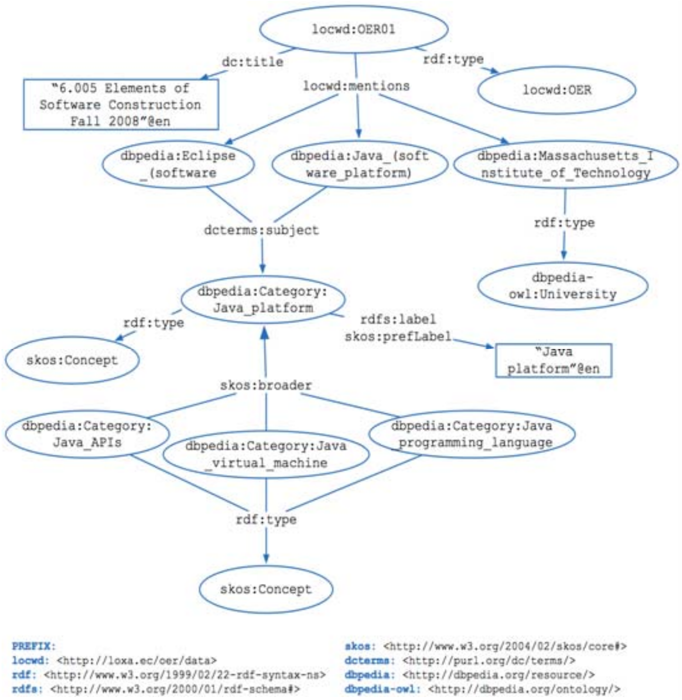

Figure 10: An example of annotation of OER from DBPedia

As we discussed, the architecture not only uses LOCWD dataset, but also uses information from Linked Open Data project. This allows exploiting the LOD community benefits. Table 2 enlists the resources retrieved from DBPedia related to "programming and Java" subject. The query returned 899 related concepts.

SPARQL queries are used to semantically annotate OER materials. SPARQL is a query language designed to gather data from multiple sources for anything that asks a question.

Subject - URI of Category	Related subjects
http://dbpedia.org/resource/Category:Java_platform	231
http://dbpedia.org/resource/Category:Object-oriented_programming_languages	162
http://dbpedia.org/resource/Category:Java_libraries	86
http://dbpedia.org/resource/Category:Concurrent_programming_languages	83
http://dbpedia.org/resource/Category:Java_specification_requests	77
http://dbpedia.org/resource/Category:Articles_with_example_Java_code	55
http://dbpedia.org/resource/Category:Class-based_programming_languages	48

Table 4: Topics and subjects related to Programming and Java, from DBpedia

5.5 Integration and reuse of OER in MOOCs Context

When the representation of OERs is available in machine-readable format and the MOOC profile has been defined, it is time to apply the procedures described by components 5-9.

In this work, the authors integrate the best features of recovery and filtering of Web information, with the aim of finding the most suitable resources that can be recommended to the course designer. Specifically, we used the Lucene library to index and retrieve open educational resources related to each of the topics defined by the syllabus of the course and we used SPARQL queries to filter the most relevant resources. As we discussed, the architecture not only uses LOCWD dataset, but also uses information from Linked Open Data project. This allows exploiting the LOD community benefits.

Topics related with Java	Number of OER Filtered	Provider of OER	Type of OER
Introduction Data Types Decisions Iteration Arrays	176	Massachusetts Institute of Technology (165) Korea University (10) University of Massachusetts Boston (1)	Lecture Notes (45) Assignments (61) Readings (33) Syllabus (9) Projects (5)
Objects Classes Methods Inheritance	281	Massachusetts Institute of Technology (276) Korea University (5)	Lecture Notes (145) Readings (67) Assignments (48) Syllabus (4) Demonstration – video (2)

Table 5: Summary of OER found in LOCWD data source for Topics related with "Introduction to Java programming"

The framework provides transparent access to RDF data sources for OER stored in OCW repositories. In the architecture, the recommendation seeker is based on SPARQL. With Sparql –it is possible filter OER using multiple category or taxonomy terms at the same time, and combine text searches, category term filtering, and other search criteria. Then, it may ask for an OER recommendation based on attributes and annotation from the course profile designed. Table 6, summarizes OER found in LOCWD data source

Table 7 shows some recommended OERs for introductory course to Java. The topic is "Introduction to Java programming", the aim is study the necessary elements that allow you to build simple Java programs, the estimated time is one week; the subtopics used for filter OER are: Programming Introduction. Java program structure Program flow. Arithmetic Operators. Primitive data types. Recommended OERs are extracted from OCW courses, this measure provides a dimension of quality in filtered resources.

OER Title	**Kind of resource**	**URL of OER recommended**	**Some metadata**
6.170 Laboratory in Software Engineering. Java Style Guide	Learning material	http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-092-java-preparation-for-6-170-january-iap-2006/study-materials/java_style.pdf	:Language English :Provider MIT :File java_style.pdf
Types, Variables, Operators	Lecture	http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-092-introduction-to-programming-in-java-january-iap-2010/lecture-notes/MIT6_092IAP10_lec01.pdf	:Language English :Provider MIT :File MIT6_092IAP10_lec01.pdf
Fundamentals of Java	Tutorial Lecture	http://ocw.uc3m.es/ingenieria-informatica/programacion/manuales/java2-U-Navarra.pdf http://ocw.uc3m.es/ingenieria-informatica/programacion/transparencias/tema2.pdf	:Language: Spanish :Provider: UC3M :File java2-U-Navarra.pdf :File tema2.pdf
Problems: Getting Started	Lab	http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-092-java-preparation-for-6-170-january-iap-2006/labs/problems1_4.pdf	Language English :Provider MIT :File problems1_4.pdf
More types, Methods, Conditionals	Lecture	http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-092-introduction-to-programming-in-java-january-iap-2010/lecture-notes/MIT6_092IAP10_lec02.pdf	Language English :Provider MIT :File MIT6_092IAP10_lec02.pdf

Table 6: OER recommended for subtopic: Programming introduction

6 Conclusions

The architecture presented is an approach of components that could provide in order to improve openness in MOOCs design. We believe that the integration of OERs and MOOCs offer a remarkable opportunity for the training of thousand of participants distributed all over the world, allowing free and online access to quality education through a new way of open learning.

A special type of OER is full open courses or OpenCourseWare. An OCW is a free and open digital publication of high quality educational materials, organized as courses. The OCW initiatives combines two things: the traditional openness and outreach and democratizing of education, and the ability of the Web to make vast amounts of information instantly available. OCW projects deliver high quality instructional content to an unlimited number of users at virtually no additional cost beyond the original cost of production or cost of adaptation. An OCW institution provides its core materials. However, this is not enough, the real education requires interaction. We believe that a next generation of interactive OCW must be developed; where their structure and dynamic is being constantly adapted as more experience is gained with their delivery and so it is important to understand their benefits and expectations in a systematic manner.

As part of this effort, this work advocates the use of Linked Data technologies as an enabler for the development of the next generation of OER (Linked Open Educational Resources Data), allowing the separation of semantics from syntax, the improvement of discoverability and access and the use of common vocabularies. In addition, the proposed architecture provides to data consumers an opportunity to merge data distributed across different libraries.

Linked Data approach contribute a solution to one of the main problems in Open Educational Resources repositories, which is the large variety of norms, standards and application profiles that preclude efficient discovery and reuse for OER within multiple and distributed repositories.

Important developments in the integration of OER and MOOCs required tools and educational practices to support creation and maintenance of courses. These are in process but not complete, making the challenge considerable for MOOC practitioners. In the meantime, it’s critical the use of linked data approach on OER/OCW repositories to shift towards a more interoperable and integrated context of re-use of data and metadata. The purpose is to significantly improve discovery, accessibility, visibility, and promote reuse of open educational content in open and online courses.

7 Future Works

MOOCs have forced the OER movement to re-evaluate itself in terms of how it fits in to the emergent education landscape. Our future work will focus on a next generation of interactive and social OERs (e.g. S-OCW Social OpenCourseWare) by providing an OER platform where self-learners and students can take an active role in the management of their personal learning environments, through self-organized dashboards and collaborative workspaces. Teachers will have a functionality that allows them to extract OER for their courses based on learning paths. The general purpose is to enhance the professional skills of learners using OERs, ICT-innovative and pedagogically-rich and tailored learning paths based on open educational materials with a specific focus on the development, extension and expansion of professional skills.

A main component is the establishment of an innovative method for evaluating individual users' learning needs to develop professional skills, providing effective resources, and setting up a peer review and sharing community to ensure the quality of the contents. In the following, we suggest research opportunities in relation to

- Develop research on how to improve the OpenCourseWare environment by investigating new open learning models.
- Conduct and share more experimental studies about integration of OER on MOOC scenarios.
- Identify new ways about business models that preserve the learning experience supported by OER initiatives.
- Examine open and collaborative assessment methods that fit better to the MOOC and OER environments characterized by networking, openness, and self-organization.
- Develop open education practices related with accreditation and recognition of OERs and MOOCs deals.

The authors consider important the use of Linked Data technologies as an enabler for the development of the next generation of Open Educational Resources, allowing the separation of semantics from syntax, the improvement of discoverability and access, and the use of common vocabularies. Additionally, the Linked OER and OCW Data environment enabled us to discovery educational resources, develop courses, and show data visualizations

Acknowledgements

This work has been partially funded by scholarship provided by the "Secretaría Nacional de Educación Superior, Ciencia y Tecnología e Innovación" of Ecuador (SENESCYT), and Universidad Técnica Particular de Loja

References

[Abeywardena, 2012] Abeywardena, I. S.: The Re-use and Adaptation of Open Educational: Resources (OER). An Exploration of Technologies Availables. Commonwealth of Learning, (2012).

[Alario-Hoyos et al., 2014] Alario-Hoyos, C., Pérez-Sanagustín, M., Cormier, D., Delgado-Kloos, C.: Proposal for a Conceptual Framework for Educators to Describe and Design MOOCs. Journal of Universal Computer Science, 20, 1, (2014), 6-23.

[Berners-Lee, Hendler and Lassila, 2001] Berners-Lee, T. Hendler, J. & Lassila, O.: The Semantic Web. Scientific American, 284, 5, (2001), 34-42.

[Berners-Lee, 2006] Berners-Lee, T.: Linked Data: Design Issues. W3C. (2006). http://www.w3.org/DesignIssues/LinkedData.html.

[Bizer, 2007] Bizer, C., Cyganiak, R., Heath, T.: How to Publish Linked Data on the Web.

[Brusilovsky and Nijhavan, 2002] Brusilovsky, P., Nijhavan, H.: A Framework for Adaptive E-Learning Based on Distributed Re-usable Learning Activities. Proceedings of World Conference on E-Learning, E-Learn 2002 (2002).

[Cano et al., 2013] Cano, A., Varga, A., Rowe, M., Ciravegna, F., He, Y.: Harnessing Linked Knowledge Source for Topic Classification in Social Media. In the proceedings of the ACM Hypertext and Social Media Conference. Paris, France, (2013).

[Caswell et al., 2008] Caswell, T., Henson, S., Jensen, M., Wiley, D.: Open Content and Open Educational Resources: Enabling universal education. The International Review Of Research In Open and Distance , 9, 1, (2008) 1-11.

[Chicaiza et al., 2014] Chicaiza, J., Piedra, N., López-Vargas J. & Tovar, E.: Domain Categorization Domain Categorization of Open Educational Resources Based on Linked Data Sources. Communications in Computer and Information Science, Springer Verlag, (2014), 15-28

[Frango-Silveira, Nizam and Notargiacomo-Mustaro, 2005] Frango-Silveira, I., Nizam, O., Notargiacomo-Mustaro, P.: Reusability and Interoperability of Adaptive Learning Objects Repositories. XVI Simpósio Brasileiro de Informática na Educação - Workshop em Informática na Educação (SBIE - UFJF) (2005).

[Heath and Bizer, 2011] Heath, T. & Bizer, C.: Linked Data: Evolving the Web into a Global Data Space. Synthesis Lectures on the Semantic Web: Theory and Technology, 1, 1, (2011), 1-136, Morgan & Claypool.

[Hilton et al., 2010] Hilton, J., Wiley, D., Stein, J., Johnson, A.: The four R's of openness and ALMS Analysis: Frameworks for Open Educational Resources. Open Learning: The Journal of Open and Distance Learning, 25,1, (2010), 37-44.

[Hodgkinson-Williams and Gray, 2009] Hodgkinson-Williams, C., Gray, E.: Degrees of Openness: The emergence of Open Educational Resources at the University of Cape Town. International Journal of Education and Development using Information and Communication Technology (IJEDICT), 5,5, (2009), 101-116.

[Husby and Barbosa, 2012] Husby, S. D., Barbosa, D.: Topic Classification of Blog Posts Using Distant Supervision. Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics, (2012), 28 – 36.

[ICESC, 1976]. International Covenant on Economic, Social and Cultural Rights. ICESC. 1976. http://www.ohchr.org/EN/ProfessionalInterest/Pages/cescr.aspx

[Kellar et al., 2004] Kellar, M., Stern, H., Watters, C., Shepherd, M.: An Information Architecture to Support Dynamic Composition of Interactive Lessons and Reuse of Learning Objects. Proceedings of the 37th Hawaii International Conference on System Sciences, (2004).

[Klerkx et al., 2010] Klerkx, J., Vandeputte, B., Parra, G., Santos, J. L., Van Assche, F., Duval, E.: How to Share and Reuse Learning Resources: The ARIADNE Experience. Sustaining TEL: From Innovation to Learning and Practice Lecture Notes in Computer Science, 6383, (2010), 183-196.

[Koper, 2013] Koper, R.: Combining reusable learning resources and services to pedagogical purposeful units of learning. In A. Littlejohn (Ed.), Reusing Online Resources: A Sustainable Approach to eLearning, (2003). 46-59. London: Kogan Page.

[MIT, 2001]. MIT to make nearly all course materials available free on the World Wide Web. 2001. http://web.mit.edu/newsoffice/2001/ocw.html

[Muñoz-García et al., 2011] Muñoz-García, O., García-Silva, A., Corcho, O., de la Higuera, M., and Navarro, C.: Identifying Topics in Social Media Posts using DBpedia, Proceedings of the Networked and Electronic Media Summit (NEM summit 2011), 2011, Torino, Italia

[OCWC, 2013] OpenCourseWare Consortium: OCWC User Feedback Survey Results» Announcements, Retrieved October 2013 from http://www.openedconsortium.org/news/2013/06/ocwc-user-feedback-survey-results/ (2013)

[OKF, 2012)] Open Data Handbook Documentation. Release 1.0.0 Open Knowledge Foundation. http://opendatahandbook.org/pdf/OpenDataHandbook.pdf

[Petrides et al., 2008] Petrides, L., Nguyen, L., Jimes, C., Karaglani, A.: Open educational resources: inquiring into author use and reuse. International Journal of Technology Enhanced Learning, 1, (2008), 98 - 117.

[Piedra et al., 2013] Piedra, N., Tovar, E., Dimovska, A., Chicaiza, J.: OCW-S: enablers for building sustainable Open Education. In Proceeding of IEEE Global Engineering Education Conference (EDUCON), Berlin, March 2013.

[Piedra et al., 2014a] Piedra, N., Tovar, E., Colomo-Palacios, R., López, J., Chicaiza, J.: Consuming and producing linked open data: the case of Opencourseware. Program: electronic library and information systems, 48, 1, (2014),16 - 40.

[Piedra et al., 2014b] Piedra, N., Chicaiza, J., López, J., Tovar-Caro E.: Supporting openness of MOOCs contents through of an OER and OCW framework based on Linked Data technologies, Global Engineering Education Conference, IEEE, Istanbul-Turkey, April 2014, 1112 - 1117.

[Tovar and Piedra, 2014] Tovar, E., Piedra, N. "Guest editorial: Open Educational Resources in Engineering Education: Various perspectives opening the Education of Engineers". IEEE Transactions on Education, 57(4), pp. 213-219, 1 November 2014. DOI: 10.1109/TE.2014.2359257

[UNESCO, 2000]. The Dakar framework for action. Paris: UNESCO. 2000 http://www.unesco.org/education/wef/en-conf/dakframeng.shtm

[UNESCO, 2002a] UNESCO: Forum on the Impact of Open Courseware for Higher Education in Developing Countries. Final report, (2002).

[UNESCO, 2002b] UNESCO Promotes New Initiative for Free Educational Resources on the Internet. 2002. http://www.unesco.org/education/news_en/080702_free _edu_ress.shtml.

[UNESCO, 2012] UNESCO. 2012 Paris OER Declaration. 2012 World Open Educational Resources (OER) Congress. Paris, June 20-22, (2012).

[White and Manton, 2011] White, D., Manton. M.: Open Educational Resources: The value of reuse in higher education. JISC-funded OER Impact Study, University of Oxford, (2011).

[W3C, 2009] W3C: SKOS Simple Knowledge Organization System Reference: W3C Recommendation 18 August 2009, (2009), Retrieved February 2014 from http://www.w3.org/TR/skos-reference/.