CONSUMING AND PRODUCING LINKED OPEN DATA: THE CASE OF OPENCOURSEWARE

NELSON PIEDRA

Universidad Técnica Particular de Loja, Ecuador nopiedra@utpl.edu.ec

EDMUNDO TOVAR

Universidad Politécnica de Madrid, Spain etovar@fi.upm.es

RICARDO COLOMO PALACIOS *

Universidad Carlos III de Madrid, Spain ricardo.colomo@uc3m.es

JORGE LOPEZ, JANNETH CHICAIZA

Universidad Técnica Particular de Loja, Ecuador {jalopez2, jachicaiza} @utpl.edu.ec

Purpose

The aim of this paper is to present an initiative of application of the Linked Data principles to promote data interoperability between heterogeneous OpenCourseWare (OCW) repositories and to enhance the search and discovery of contents created and shared by the universities.

Design/methodology/approach

This paper is a case study of how Linked Data technologies can be applied for the enhancement of open learning contents.

Findings

Results presented under the umbrella of OpenCourseWare Consortium (OCWC) and OCW-Universia consortium, as the integration and access to content from different repositories OCW and the development of a query method to access these data, reveal that Linked Data would offer a solution to filter and select semantically those open educational contents, and automatically are linked to the Linked Open Data Cloud.

Originality/value

The new OCW integration with Linked Data adds new features to the initial framework including improved query mechanisms and interoperability.

1. Introduction

OpenCourseWare (OCW) is one of the most successful emerging initiatives in the way to Open Content. The great impact of the initiative has led many international and prestigious universities to join the movement; creating their own projects in coordination with OCW-Massachusetts Institute of Technology.

There is an enormous amount of academic information and digital resources produced in universities with OCW Initiatives. The internal organization, structure and technological infrastructure of an OCW project are diverse, and respond to the vision of each university. There is not a standardized way to implement OCW initiatives, however some guidelines and recommendations on the characteristics of repositories and the structure of an OCW project can be found. Materials in OCW repositories are not usually described by metadata. Therefore, one of the greatest challenges that OCW initiatives, and that in particular, the OCWC and OCW-Universia* have to face is the heterogeneity of existing OCW repositories. Heterogeneity leads to problems of interoperability and accessibility of open content among institutions and within them. The lack of interoperability shows some disadvantages in the discovery, reuse, remix and adaptation of OCWs. OCW Community must find a way to exchange quick and easy access to Open Educational Resources (OER).

The use of Linked Data approach on OCW repositories provides the framework for evolving into a more interoperable and integrated system to sharing, connecting and discovering data and metadata of OCW initiatives.

In this paper, authors apply Linked Data design approaches to describe and retrieve information that is semantically related to open educational resources accessible via the OCWC and OCW-Universia websites. Once Linked Data and OCW model are introduced (sections 2 and 3 respectively), the main problem that authors face is described in section 4.1 and the deployment of the solution is detailed in section 4.2. Description of the way to elaborate queries to extract Linked data is presented in the section 5. Finally, conclusions are summarized in the section 6.

2. Linked Data

^{*} OCW-Universia promotes and disseminates the OCW concept in Ibero-America.

Semantic Web technologies and, more precisely, Linked Data are changing the way information is stored and exploited. The term "Linked Data" refers to a set of best practices for publishing and connecting structured data on the Web (Bizer, Heath, & Berners-Lee, 2009). In summary, the Linked Data Design Issues, outlined by Tim Berners-Lee back in 2006, provide guidelines on how to use standardized Web technologies to set data-level links between data from different sources (Bizer, 2009). These Linked Data Design Issues are:

- Use URIs as names for things
- Use HTTP URIs so that people can look up those names
- When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
- Include links to other URIs, so that they can discover more things 4.

In contrast to the full-fledged Semantic Web vision, linked data is mainly about publishing structured data in RDF using URIs rather than focusing on the ontological level or inference (Hausenblas, 2009). This new and simplified approach minimizes entry barriers and fosters a softer adoption. The Open Data movement has seen successful pressure on governments to expose and open up data sets—in many cases this is being done using a Linked Data approach (Bechhofer et al., 2012). Linked Data—particularly data available using open licenses—has an important role to play in information systems and could be a key feature for the Web of Data (Passant et al., 2009). However, and in spite of its huge potential, an important weakness is that it has not been established completely a formal reference that integrates the necessary infrastructure in terms of components (Colomo-Palacios et al., 2012).

The use of Linked Data is widely documented e.g. (Álvarez et al., 2012; García-Crespo et al., 2010; Joerg et al., 2012), but in education is in particular remarkable. For instance (Ruiz-Calleja et al., 2012) propose a tool to collect data from third-party sources, align it to a vocabulary understandable by educators and finally publish it to be consumed by educational applications. In educational video settings, (Yu et al., 2012) present a video annotation and browser platform that enables the annotation of video resources using vocabularies defined in the Linked Data cloud along with the browsing of semantically linked educational video resources with enhanced Web information from different online resources. MetaMorphosis+ by (Dovrolis et al., 2012), designed to be applied in the medical domain, is able to publish richly annotated educational resources that are further semantically enriched and exposed in the Linked Open Data cloud. The work of (Fernandez, d' Aquin, & Motta, 2011) presents an effort that is focused on selecting, extracting, structuring and interlinking information of video lectures produced by 27 different educational institutions.

Considerable work has been devoted to increase the interoperability between Learning Object Repositories that rely on different metadata schemas e.g. IEEE LOM. To achieve this goal, for example, the work of (Ebner et al.,2009) presents an application profile of the IEEE-LOM and a Web-based tool for description of learning resources on organic agriculture, in the context of the Organic Edunet initiative[†]. However, learning object metadata is typically not linked across repositories and not is possible navigate or interoperate between different data sources available on the Web. This problem is addressed through Linked Data by (Sicilia et al., 2011) that describes how linked data has been integrated to Organic Edunet to navigate learning resources.

A recent survey has been put forward by (Dietze et al., 2013) in relation to the application of Linked Data to solve interoperability issues in the field of e-learning. In this paper are presented initiatives whose aim has been to share or allow access through services to educational web data. Then, authors present a general-purpose approach for the integration of existing data on open public educational repositories (accessible via Web services and available APIs), and later expose metadata as linked data. Their proposal has not been implemented.

Our work differs from the above, e.g. (Ebner et al.,2009; Dietze et al., 2013) because we propose a vocabulary to describe OER, resources that have particular characteristics that differentiate it from a traditional learning object, or have features that in the context of Open Content are important, as structure, origin (consortium, educational institution, repository), intellectual property rights, among others.

An additional contribution of our work is the extraction of data from unstructured or semi-structured and heterogeneous sources (most OCW courses are published as HTML pages, are on different platforms and can vary in their structure). Works of Linked Data publishing, as (Dietze et al., 2013; Fernandez, d 'Aquin, & Motta, 2011) are based on information extracted through APIs or extracted from a single site.

Finally, the paper of (Tovar et al., 2012), based on previous works e.g. (Chicaiza et al., 2010), describes the connection of open educational resources by means of social and semantic web technologies.

3. Open Educational Resources and the OCW model

The internet has increasingly become the dominant medium for making resources available online in a digital format, in order to be accessed, used and reused by interested audiences (Tzikopoulos et al., 2012). As a

^{† &}lt;a href="http://organic/edunet.eu">http://organic/edunet.eu Organic.Edunet is a federation of learning repositories in the domain of learning resources on organic agriculture.

consequence of this, technology-based education or technology applied to education is a key issue in Knowledge Society. E-learning systems have not only become the backbone in distance education but also are also at the heart of traditional university teaching (Rodriguez et al. 2006). In academia, many institutions use their own learning content management systems (LCMS) and digital repositories to offer digital educational resources as part of their curricula. However, and although LCMS have reached a certain level of maturity (Alier et al., 2012), such tools suffer from problems, such as the lack of openness, resistance to change, failure to take into account the user, lack of integration with informal context and so on (García-Peñalvo et al., 2011).

Many of these institutions have joined the worldwide open access to knowledge movement and are working towards releasing these academic materials as open educational resources (OERs). OER are resources, tools, and materials used to support education that may be freely accessed, reused, modified, adapted and shared by anyone. OER refers to the open provision of educational resources, enabled by information and communication technologies, for consultation, use and adaptation by the community of users for non-commercial purposes (García-Peñalvo et al,, 2010). The most commonly used open licence for OERs is Creative Commons. Open licensing clearly states the permissions and restrictions governing the OERs so that potential users are no longer under obligation to apply for permission from the resource copyright holder. Open licences simplify the copyright procedures for OER users. In this way, the use, reuse, publication and repurposing of OERs in different contexts becomes a simple, lawful and safe matter. From the academic viewpoint, OERs are pedagogically structured, teaching resources of educational value.

Online education breaks the geographical barriers and allows for different modalities of self-learning or peer-learning that complements more traditional instructor-led settings; but the possibilities of online learning are even more far-reaching thanks to the movement toward "open educational resources," which promotes the use of open licenses to share educational resources on the Web (García-Barriocanal et al., 2012). One of these initiatives is OpenCourseWare, the focus of this paper.

OCW model has proved to be one of the most successful emerging initiatives in the global promotion of open and universal access to knowledge in higher education. This model has been globally replicated in higher education institutions by the OCWC and other partner organizations that are working together to build a body of open educational contents, using this sharing model. All OCWC courses are available for adoption and adaptation by faculty and students around the world.

The OCW contents are searched using a keyword-based search engine Users can run a basic keyword search or use an advanced search option according to three metadata: title, author or keywords. Like all search engines, however, OCW-Universia also has some weaknesses and improvement opportunities. In this respect, the weaknesses can be said to revolve around interoperability, accessibility, and visibility.

The quality and quantity of resource descriptors is another search engine shortcoming. The number of keywords associated with each resource is low, generally one or two words. This means that many irrelevant results may pass through the filter, whereas other important ones may be rejected. Additionally, the site offers general information, and does not offer personalization mechanisms capable of retrieving specific information items.

4. Using Linked Data on OCW repositories

4.1. The problem; OCW repositories as independent closed data silos

OCW data is locked away in independent data silos, making it much less useful than it could be. It's difficult to develop tools for consume data from multiple silos. Searching OCW/OER across multiple silos means invoking each one's user interface, and receiving the results in separate groups. The presence of data silos makes accessing data and interoperability between repositories harder in several ways.

Browsing open educational materials suffers from the same problems as searching, made worse because each silo has its own organizational structure. Some silos have no way to link to a particular item, and so hinder the free flow of information. The presence of OCW silos impedes the interoperability, discovery, synthesis, and flow of knowledge. As a result teachers, students and self-learners spend too much time looking for resources — or they spend too little and make decisions based on incomplete information. Linked data have the potential of create bridges between OCW data silos.

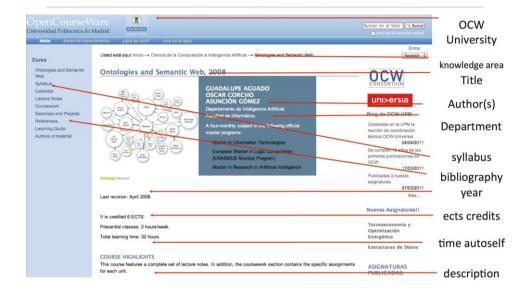


Fig. 1. Some data available from the website of an OCW.

In the context of Linked Data, authors opted to apply the design issues of Linked Data to integrate, interoperate and mash up data from distributed and heterogeneous OCW-repositories. Data available from OCW websites are: title, knowledge area, open educational materials, open licenses, repositories, higher education institutions, content creators, fields of knowledge, learning guides, licenses, accessibility conformance, and others (see Figure 1).

4.2. The solution implemented: a general framework for exposing data in OCW Domain

Authors propose an approach using Linked Data on OCW repositories and establishing the following framework for evolving into a more interoperable and integrated system to sharing, connecting and discovering data and metadata of OCW initiatives.

Linked Data technologies can also help to integrate the work of disperse institutions producing diverse linked data.

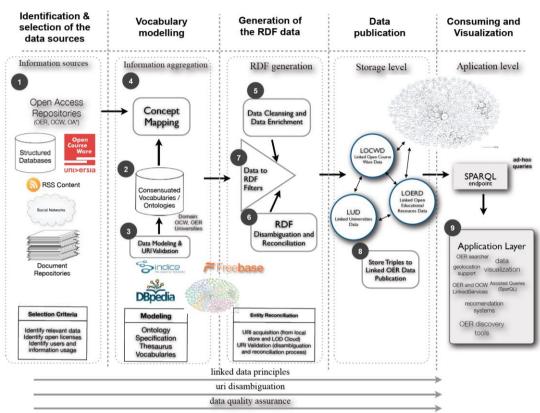


Fig. 2. General Framework for Publishing Open Educational Contents as Linked Data

This framework developed with the aim of integrating heterogeneous OCW repositories and publishing their metadata as Linked Data is divided into five phases (Figure 2):

- (i) Identify and select heterogeneous data sources to determine the scope of the content.
- (ii) Model vocabularies for OCW domain;
- (iii) Data Extraction
- (iv) Generate RDF data,
- (v) Publish linked data, and
- (vi) Consume and display linked data.

These are detailed in the next subsections, 4.2.1 to 4.2.5.

4.2.1. Identify and select data sources

The OCW environment is characterized by large amounts of unstructured and semi-structured data; although the collected data from OCW repositories may have certain structure accepted by the community, but not all data have a similar or compatible structure and meaning and open education materials are shared as Information Silos or "Walled Gardens" (see Figure 3).

With respect to the OCW initiative content management systems[‡], their organization and internal structure as content repositories are heterogeneous as they conform to the particular vision of each higher education institution. In terms of similarities, each OCW site is generally accessible via the Internet and their educational resources are published using open licenses.

Although there are agreements and minimum recommendations on the characteristics of the repositories and on the structure of an OCW, there is currently no standardized way of implementing OCW initiatives. However, these agreements have been sufficient to promote the speedy adoption of OCW initiatives, although they have posed problems for the implementation of tools to automatically process OCW data.

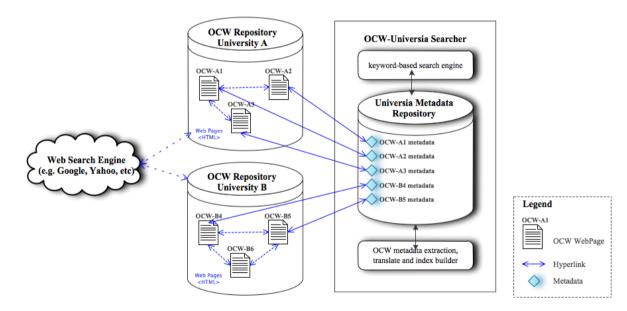


Fig. 3. Diagram of OCW-Universia search engine

[‡] Preferred OCW-CMS: eduCommons, MIT OCW system, Drupal, and Moodle

4.2.2. Vocabulary Modeling

The purpose of this stage is the creation of a vocabulary, of named classes and properties using W3C's RDF technology, for open educational resources with the aim of describing the specific types and classes of resources in OCW domain. This vocabulary was called Linked OpenCourseWare Data (LOCWD) and is available at http://purl.org/locwd/schema/.

We started with the identification of ontologies that could be used to represent the OER data. Although there is no standard OCW structure, authors have observed that the courses created by OCW-Universia and OCWC members tend to conform to a general structure, as shown in Figure 4.

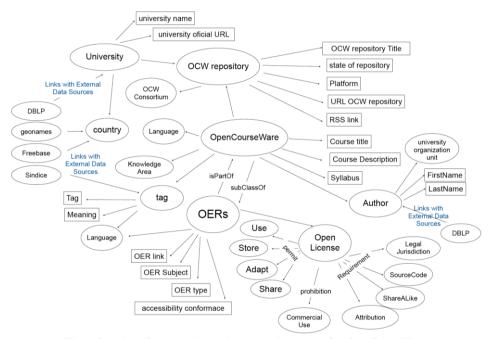


Fig. 4. Overview of conceptual map about general structure of an OpenCourseWare

LOCWD is a RDF 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 LOCWD. With LOCWD, the OER/OCW initiatives can retain some control over their information of materials and courses in a non-proprietary format. LOCWD defines twelve main classes:

- locwd:OER describes an Open Educational Resource.
- locwd:OCW describes an OCW.
- locwd:OERRepository represents a repository that stores OERs.

- locwd:OCWRepository represents a repository that stores OCW courses.
- locwd:EducationalOrganization represents the educational institution that publish open educational content.
- locwd:Person the concept represents both creators and users of OER/OCW.
 - locwd:KnowledgeArea represents a branch of knowledge.
- locwd:RichedTag describes the way to express the meaning of a tag associated to an OER/OCW.
 - locwd:OpenLicense describes an open license.
- locwd:Permission describes the action of officially allowing someone to do a particular thing with an OER/OCW.
- locwd:Requirement describes a thing that is compulsory; a necessary condition.
- locwd:Prohibition describes the action of forbidding something, specially by a law or legal regulation.

LOCWD reuses a set of RDF 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 OCW domain with Datasets in the LOD cloud.

Classes and properties comprise the core of LOCWD. They describe characteristics of OER and OCW that are independent of technology; as such they can be used to describe basic information about open educational materials (see Table 1). LOCWD introduces the following classes and properties. A machine-friendly version is also available in http://purl.org/locwd/schema on RDF/XML format.

LOCWD Classes: EducationalOrganization | KnowledgeArea | OCW | OCWRepository | OER | OERRepository | OpenLicense | Permission | Person | Prohibition | Requirement | RichedTag |

LOCWD Properties: academicUnit accConformanceLevel accessibilityCompliance accGuidelinesTitle | accGuidelinesURI | accountable | accScopeCovered | additionalName | agentTagURI | alternativeTitle | assessment | attachedURI | Attribution | attributionName | attributionURL | audience biblio | contentSize | countryCodeType | courseGuide | courseObjectives | creator | dateCreated dateCreated | dateLicensed | dateModified | datePublished | deprecatedOn | description | duration | editor | encodingFormat | familyName | givenName | grading | hasPart | hasVersion | identifier | image | isPartOf | isVersionOf | jurisdictionCodeType | kAcode | kAName | kARelation | kASuperior | knowledgeArea | language | lectureNotes | legalcode | license | ocwConsortium | ocwPlatform | ocwRepoIsAlive ocwRepoRSS | ocwRepoTitle | ocwRepoURL | orgCity | orgDescription | orgImage | orgLogo | orgName | orgPostalCode | orgState | orgURL | permits | personIsPartOf | playerType | pre-requisites | prohibits | provider | publisher | rawKeyword | relationBtwOER | requires | richKeyword | rightsHolder | rightsYear | storeIn | syllabus | tagText | title | typeOf | uploadDate | url | versionLicense | versionOER |

In addition to various characteristics of OER, LOCWD defines classes for Educational Organizations, Knowledge Areas, OCW Repositories and Open Licenses (Legal jurisdictions, Permits, Prohibitions and Requirements).

Table 1. Main concepts for Specification of LOCWD Vocabulary

CLASS	rdfs:label	rdfs:comment	owl:equivalentClass
OER	Open Educational Resource OpenCourseWare	"OER are teaching, learning, and research resources that reside in the public domain or have been released under an intellectual property license that permits their free use or re-purposing by others. OER include full courses, course materials, modules, textbooks, streaming videos, tests, software, and any other "An OCW is a free and open digital publication of high	Schema:CreativeWork aiiso:Course, bibo:Course,
rdfs:subclassOf :OER	'	quality educational materials, organized as courses.	Schema:CreativeWork
EducationalOrganization rdfs:subclassOf foaf:Organization	Educational Organization	"A college, university, or other third-level educational institution that publish their course materials for free online, under the OpenCourseWare (OCW) initiative. An Institution is the upper most level of an academic institution."	aiiso:Institution, dbpedia- owl:Organisation, dbpedia- owl:Agent, dbpedia- owl:University, schema:CollegeOrUniversi ty, dbpedia- owl:EducationalInstitution, schema:Organization, schema:EducationalOrgan ization, umbelrc:University
Person	Person	"The Person class represents humans beings.This concept represents both creators (the people	foaf:Person, schema:Person
rdfs:subclassOf foaf:Person		responsible for creating the intellectual content of OER or for adapting open educational content) and users of OER/OCW. Examples of subclasses of person are creators, students, teachers and self-learners."	
KnowledgeArea	Knowledge Area	"Knowledge areas are areas of specialization or branch of knowledge to which open course or open educational resource belongs."	aiiso:KnowledgeGrouping
OERRepository	Repository of Open Educational Resources	"A repository that contains Open Educational Resources (OER). An OER requires its creators to deposit resources to be deposited in an Open Access repository."	Schema:DataCatalog, freebase:Institutional repository, yago- res:Institutional repository, yago:Repositories

OCWRepository	Repository of	"A repository that contains OCW courses.	Schema:DataCatalog,
	OpenCourseWare		freebase:Institutional
rdfs:subclassOf			repository, yago-
:OERRepository			res:Institutional repository,
			yago:Repositories
OpenLicense	Open License	"A statement about the intellectual property rights (IPR)	cc:License,
		held in or over a open resource, a legal document	odrl:Agreement,
rdfs:subclassOf		giving official permission to do something with a	dct:RightsStatement
dct:LicenseDocument		resource, or a statement about access, use, reuse,	
		remix and adaptation rights. () OER are made	
		available under a Creative Commons, GNU, FOSS	
		licenses	
Permission	Permission	"Permission, an action that generally allows more liberal	cc:Permission
		to use, reuse, redistribution, re-mix and adaptation with	
		attribution of the origin of the OER for academic use	
		and is otherwise in compliance with the open license	
		(e.g. permits commercial derivates actions)."	
Requirement	Requirement	"Requirement, an action that may be requested of user	cc:Requirement
		in compliance with the open license. () Requirements	
		depend on each open license types, e.g. attribution and	
		notice requirements vary widely among licenses."	
Prohibition	"Prohibition"	"Prohibition, something that a license may be asked not	cc:Prohibition
		to do."	
RichedTag	"Riched Tag"	"A tag is a non-hierarchical keyword or term assigned to	
		a OCW/OER. This kind of metadata helps describe an	
		resource and allows it to be found again by browsing or	
		searching. In LOCWD vocabulary, the meaning of a tag	
		is expresses to through of a tag and an attached URI."	

An extract of data model is presented below:

A. locwd:OER Class

Description: OER are teaching, learning, and research resources that reside in the public domain or have been released under an intellectual property license that permits their free use or re-purposing by others. Open educational resources include full courses, course materials, modules, textbooks, streaming videos, tests, software, and any other tools, materials, or techniques used to support access to knowledge. In addition to the generic properties of a web resource, such as title, description, URL, language and encoding format.

B.locwd:OCW Class (subClassOf locwd:OER)

Description: An OCW is an open educational resource. The OCW concept is applied to course materials created by academic organization and shared freely with the world via the Internet. An OCW is available for use and adaptation under an open license, such as a Creative Commons license. See properties in table 2.

Table 2. Extract of properties of OpenCourseWare class

Property	R*	rdfs:range	Note
Prerequisites	Y	xsd:string	Sequence of academic requirements of each course
Course Objectives	Y	xsd:string	The OCW course expectations are usually expressed as goals and objectives
Syllabus equivalentTo dct:tableOfContents	N	rdfs:resource	An academic schedule outlining studies and summarizing the topics covered by the OCW. Syllabuses are established by academic teams, and be design to monitor the teaching quality of the OCW
Course Guide	N	rdfs:resource	Modules or units, suggested learning time, basic learning materials and readings, exercises and problems, projects, notes, coursework
Lecture Notes	Y	rdfs:resource	Additional documents recommended by the authors and to be studied as part of the course contents
Bibliographic Citation	Y	dct:bibliographicCita tion	A Bibliographic Citation
Assessment	Y	rdfs:resource	Descriptions of exercises, self-assessments and projects to be completed to guarantee the OCW learning progress
Store in	N	OCWRepository	Online container where OCW is deposited or stored.
Grading	N	xsd:string	Relate the OCW to the description about its grading

R* -> Repeatable? [Yes/No]

The **locwd:OpenLicense** class and properties describe official permission to do something with an OER/OCW.

C. locwd:OpenLicense

<u>Description</u>: A legal document giving official permission to do something with a Resource. OER are made available under a Creative Commons, GNU, FOSS licenses or similar license that generally allows more liberal use, reuse, redistribution, re-mix and adaptation than a traditional copyrighted work. See an extract of properties in table 3.

Table 3. Extract of Properties of OpenLicense class.

Label/Property	R*	rdfs:range	EquivalentTo	Note
Permits	Y	locwd:Permission	cc:permits	A open license permits a particular use of
			dc:rights, dct:rights,	the OER
			rdf:relation	
Requires	Y	locwd:Requireme	cc:requires,	A open license requires certain actions of
		nt	de:rights, det:rights	the user when enjoying the permissions
				given by Permits.
Prohibits	Y	locwd:Prohibition	cc.prohibits,	A open license prohibits a particular use of
			dc:rights, dct:rights	the OER.

Jurisdiction	Code	N	udfrs:legalJurisdic	cc:jurisdiction	A open license may have a jurisdiction code
Type			tionCodeType		type. A jurisdiction code type associates the
					open license with a particular legal
					jurisdiction
Country	Code	N	udfrs:countryCode	cc:jurisdiction	A open license may have a jurisdiction code
Type			Type		type. A country code type associates the
					open license with a particular country
Deprecated C)n	N	xsd:date	cc:deprecatedOn	A open license may be deprecated, indicates
					the open license has been deprecated on the
					given date.

R* -> Repeatable? [Yes/No]

The possibility of content adaptation is a key permission in the context of OERs: those using the open educational materials can customize them the way they want. Adapt refers to the process of taking OER developed for one context and adapting/localizing them for other contexts. The possible values for properties: locwd:permits, locwd:requires, locwd:prohibits can be seen in Table 4.

Table 4. Possible values for predicates of Open Licenses

LOCWD:AdaptPermit A Possible PERMISSION Granted

Label: Adapt or localize permit | Type: <locwd: Permission> | equivalent Class: <cc: Derivate Works>

Definition: Granted permit to adapt and re-mix (mix different OER) educational materials. Adapt refers to the process of taking OER developed for one context and adapting/localizing them for other contexts (e.g. geographical, language, pedagogical, political, or technical contexts). Adapt is the ability to personalize materials to meet specific teaching and learning needs.

LOCWD:SharePermit A Possible PERMISSION Granted

Label: Share Permit | Type: <locwd:Permission> | equivalentClass: <cc:Distribution>

Definition: Granted permit to share, distribute, copy, display and perform the educational resources openly.

LOCWD:StorePermit A Possible PERMISSION Granted

Label: Store Permit | Type: <locwd:Permission> | equivalentClass: <cc:Distribution>

Definition: Granted permit to store the educational resources

LOCWD:UsePermit A Possible PERMISSION Granted

Label: Use permit | **Type:** <locwd:Permission>

Definition: Granted permit to use and re-use the OER.

LOCWD:ShareAlike A Possible REQUIREMENT to enjoy the permission given

Label: Share a like | Type: locwd:Requirement | equivalentClass: <cc:ShareALike>

Definition: Open licenses require share alike when redistributing derivate OER, using the same open license.

LOCWD:Attribution A Possible REQUIREMENT to enjoy the permission given

Label: Attribution | Type: <locwd:Requirement> | equivalentClass: <cc:Attribution>

Definition: Open licenses require attribution (or credit) to the author of a educational resource.

<rd>s:Comment> OER creators let others share, store, distribute, copy, distribute, display and perform your

licensed work – and derivative other OER based upon it – but only if the credits are attributed.

LOCWD:SourceCode A Possible REQUIREMENT to enjoy the permission given

Label: Source Code | Type: <locwd:Requirement> | equivalentClass: <cc:SourceCode>

Definition: Open licenses require source code when redistributing this OER (this requirement is used withour of type software)

LOCWD:CommercialUse>A Possible PROHIBITION to Particular Use

Label: Commercial Use | Type: <locwd:Prohibition> | equivalentClass: <cc:CommercialUse>

Definition: Using the OER for commercial purposes.

New terms may be added at any time and consequently this specification is an evolving work. Future versions may incorporate multilingual translations of the term definitions.

4.2.3. Data Extraction

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. However, web scraping, or the automated harvesting of data on the Internet, is a huge problem. In theory, Web scraping gives access to any data. According to our experience, Web Scraping is tricky and time-consuming to write a reliable and complete scraper because of the complexity and changeability of each OCW user interface.

Information has been extracted with ad-hoc Python scripts. In the present work authors selected and extracted information from 80 heterogeneous OCW repositories from OCWC and OCW-Universia members, sifting through a total of 7,239 OCW courses and 47053 OERs approx. Data scraping were used to extracts data from OCW platforms that was later structured and stored in a database. Scraping eliminated the need for having to do the retrieval manually. These courses have 6340 creators, including organizations and academic personnel linked to one of the 80 higher education institutions from 28 countries. They describe 657 branches of knowledge (and 27588 keywords, which are used to categorize OCW courses. The scope is showed in Table 5.

Data extraction was conducted in two phases: a.) General Data extraction from the OCW-Universia; the available metadata are course title, course description, knowledge area, course website address, authors, higher education institution, and keywords. b.) Specific Data extraction from each URL where the OCW is originally stored. The OCW data obtained from the different OCW repositories were cleaned, disambiguated and formalized for later processing according to Linked Data Design Issues and using LOCWD Vocabulary.

Table 5. The OCW published by OCWC and OCW-Universia (December 2012)

Category	Description and number of OCW courses in parentheses
Region	Africa (398) Asia (673) Europe (2953) Latin America (330) North America (2885)
Countries	Argentina (40) Brazil (37) Canada (23) Chile (71) Colombia (31) Costa Rica (12) Ecuador (12) France (3)
	India (11) Indonesia (28) Iran (7) Israel (25) Japan (364) Korea (160) Malaysia (48) Mexico (75) Netherlands
	(15) Peru (37) Poland (86) Republic of China (15) Russia (12) South Africa (197) Spain (2327) Taiwan (15)
	United Kingdom (510) United States (2862) Venezuela (15) other (201)
Language	Catalan (93) Chamorro (15) Chinese (15) Dutch (7) English (4084) French (3) Hebrew (32) Indonesian (2)
	Japanese (182) Korean (160) Polish (86) Portuguese (37) Russian (12) Spanish (2511)
OCW Platform	Dspace (201) e-learning (86) eduCommons (1627) Plone (72) HTML (79) Microsoft Content Management
	Server (2149) Moodle (628) other (2397)
Universities	African Virtual University (201) AGH University of Science and Technology (86) Capilano University (23)
	Fundação Getulio Vargas - FGV Online (37) Fundación Universitaria San Pablo CEU (4) Gunadarma
	University (26) Hanyang University (61) Hokkaido University (64) Indian Institute of Management Bangalore
	(11) International Excellence (10) Johns Hopkins Bloomberg School of Public Health (118) Kaplan University
	(9) Korea University (99) Kyoto Seika University (20) Kyoto University (153) Kyushu University (10)
	Massachusetts Institute of Technology (2149) Moscow Architectural Institute (12) National Chiao Tung
	University (15) New Jersey Institute of Technology (20) ParisTech (3) Pontificia Universidad Católica de
	Chile (10) Pontificia Universidad Católica de Valparaíso (21) Shahid Beheshti University (7) Taipei Medical
	University (15) Tecnológico de Monterrey (11) The Open University (510) The Open University of Israel (25)
	Tokyo Institute of Technology (10) TU Delft (15) Tufts University (50) United Nations University (17)
	Universidad Anáhuac (32) Universidad Carlos III de Madrid (206) Universidad Central de Venezuela (15)
	Universidad de Alicante (162) Universidad de Cádiz (67) Universidad de Cantabria (134) Universidad de Chile
	(40) Universidad de Granada (22) Universidad de Huelva (30) Universidad de Monterrey (32) Universidad de
	Murcia (138) Universidad de Navarra (22) Universidad de Oviedo (77) Universidad de Salamanca (94)
	Universidad de Sevilla (243) Universidad de Zaragoza (31) Universidad del País Vasco (72) Universidad del
	Valle (12) Universidad Estatal a Distancia (12) Universidad Icesi (9) Universidad Industrial de Santander (10)
	Universidad Internacional de Andalucía (31) Universidad Nacional de Córdoba (40) Universidad Nacional de
	Educación a Distancia (44) Universidad Nacional de Ingeniería (26) Universidad Nacional San Augustín (11)
	Universidad Politécnica de Cartagena (47) Universidad Politécnica de Madrid (134) Universidad Politécnica de
	Valencia (118) Universidad Técnica Particular de Loja (12) Universitat Autònoma de Barcelona (12)
	Universitat de les Illes Balears (13) Universitat de Lleida (32) Universitat de València (64) Universitat Jaume I
	(54) Universitat Oberta de Catalunya (161) Universitat Politècnica de Catalunya (305) Universiti Teknologi
	Malaysia (48) University of California, Irvine (326) University of Cape Town (188) University of
	Massachusetts Boston (45) University of Michigan (86) University of Notre Dame (48) University of Sumatera
	Utara (2) University of the Western Cape (9) University of Tsukuba (90) Weber State University (7) Wheelock
	College (4)
Total OCW	7239 OCW courses (March 2013)

4.2.1. RDF Generation

The next step was to consolidate the metadata in a linked data schema for the purpose of simplifying future processing. Then, the selected data were converted to Linked Data using the LOCWD vocabulary and a data conversion program written in Jena.

RDF has been generated with ad-hoc Jena scripts that are executed in two processing steps, integrating with ease the generation of RDF and tasks such as scraping of the OCW sites that stores OCW Data.

The first step generates data about OCW repositories and Education Institutions, which is static, and thus which is executed only a few times. The second step generates the information about OCW courses on a regular basis, keeping the data updated. The information is obtained by extract data from each OCW site. When new OCW courses are added or updating, these are processed.

The resources described in Linked Data/RDF are stored in Virtuoso[§], making them available for humans and computers. At this point, each resource was identified by a URI with a dereferencing option, and thus display the results retrieved as Linked Data.

Finally, we have used Silk Framework to link data** (region, country, university, city, language) of the OCW to other datasets in LOD; however, in this work too were made directly SPARQL queries. The power of data lies in its links.

Querying DBPedia, authors obtained additional information about universities such as name in different languages, label, comment, latitude, and longitude. From Geonames were extracted data about locations like continent, country, capital, city or state.

The correspondence between institutions of OCW data repository and DBpedia was achieved through the names and the URL of an institution, and in the case of locations, the name was enough to find data. A little manual work had to be done to find missing data when there were no correlation. Based on DBPedia URIs and using the property sameAs were obtained equivalent links from FreeBase; this enabled to consolidate a greater amount of information. For example the university acceptance rate could be used later to combine with other criteria to help determine the quality and / or relevance of courses.

For the 97.5 percent of universities and organizations that produce OCWs and appearing in our dataset, were found information in DBPedia, and for countries get the URI of all. In total, 196 links were established with DBpedia, 102 with Freebase.

4.2.2. Consume and display linked data.

The domain of OCW resources was described as Linked OpenCourseWare Data because, Linked Data holds the potential to move OCW collections out of their silos and therefore opening the data:

- To leverage the knowledge capital represented by our OCW repositories
- To enrich our information landscape, to improve visibility
- To improve ease of discovery open academic resources

[§] http://virtuoso.openlinksw.com

^{**} We created links to three external datasets: DBPedia, Freebase and Geonames. From these repositories were linked information of interest related to universities and locations.

- To improve ease of consumption and reuse of OCW
- To reduce redundancy in search of OCW
- To promote innovation and added value to Open Educational

Because the linked data on the Web are accessible via services or public query libraries, OCW data can be processed, reused, combined, integrated, and used for several purposes. Thus, benefits of this approach have been defined through the elaboration of the following study case:

Study Case Title: OCW Community must explore a way to find exchange and easily access to open educational materials.

The Problem: OCW data is locked away in independent data silos, making it much less useful than it could be. The presence of OCW silos impedes the synthesis, share, delivery and flow of knowledge. It's difficult to develop tools for consume data from multiple silos. Teachers, students and self-learners spend too much time looking for resources – or they spend too little and make decisions based on incomplete information. Searching OCW/OER across multiple silos means invoking each one's user interface and receiving the results in separate and disconnected groups. The academic community may have difficulty reusing, re-mixing and adapt of OCW resources.

The solution: The application of LOCWD Vocabulary and Linked Data principles promote the data interoperability between heterogeneous OpenCourseWare (OCW) repositories and enhance the search and discovery of contents created and shared by the universities. The Semantic Web provides a framework and the technologies needed to describe and access resources and data from the Web. The machines could do a better job integrating, mixing, reusing and consolidating data if its meaning is explicitly defined by formal languages.

Key Benefits of LOCWD;

- For teachers, self-learners and students the benefits include
 - discovery, promote Improve interoperability, provenance and facilitate automated processing by increased flexibility to changes in presentation, adaptation, use, reuse, remix, and reduced ambiguity. The community can harness knowledge and resources from Open Content initiatives for self-learning over time
 - Less effort to find resources and don't miss relevant OCW resources.
 - Linked data available on the Web can be harnessed to know more about the authors of educational material, the institutions that provide resources, teachers who are

interested in incorporating learning objects into their courses, and information of context as locations, language and culture.

- For the higher education institutions the benefits include
 - Provides services to: evolve, interoperate, share, connect and discovery of data and metadata of the OCW initiatives.
 - Perform learning analytics for purposes of understanding and optimizing learning and the environments in which it occurs
- For specialized agency, like UNESCO
 - Know the state of higher education and OER initiatives via large-scale analysis of structure data.
 - o Increase use and dissemination of open educational resources
 - Improve awareness of the existence of such resources in society
- For developers community
 - Building prototypes, demos and innovative tools that exploit, use, integrate or analyze open educational data. Web, desktop and mobile applications
 - Opportunity to create and develop new services and features that maximize the performance of search engines (e.g. Using the semantics for better interpretation of data by machines, hence find more relevant results to a user.)
 - On the Web of Data in addition to the power of global reference and the semantics of RDF data enable developed search engines or recommender systems to find non-explicit associations between courses and their content, reuse and integrate data coming from different sources, resolve problems of ambiguity or identity of things, and find more relevant results for end users.

In OCW domain there is special interest in gathering information about higher education institutions, geographical locations, keywords, information on people, scientific publications, and educational and scientific resources. In this phase, authors demonstrated that OCW resource metadata could be enriched using datasets hosted by the LinkedOpenData cloud. Additionally, the linked data environment enabled us to run data queries of distributed and heterogeneous of open educational contents repositories through SparQL-EndPoint, APIs or Web services.

5. Demonstration of Linked Data Oueries in LOCWD

Linked data were obtained through queries written in SPARQL, one of the most widespread query languages used in the scientific community conducting research in the field of the Semantic Web. SPARQL queries retrieved information items from OCW domain based on the subset of data that satisfied particular conditions.

To demonstrate that the framework operates correctly, authors connected RDF data from one data space with another, which authors linked with heterogeneous data sources. Some of the queries run are described below.

Query A: Title for the OCW UPMSW08

```
PREFIX dc: <a href="http://purl.org/dc/elements/1.1/">http://purl.org/dc/elements/1.1/</a>
PREFIX xsd: <a href="http://www.w3.org/2001/XMLSchema#">http://www.w3.org/2001/XMLSchema#</a>
PREFIX locwd: <a href="http://www.example.org/schema/locwd/">http://www.example.org/schema/locwd/</a>
SELECT DISTINCT *
WHERE { <a href="http://www.example.org/data/locwd/id/ocw/UPMSW08">http://www.example.org/data/locwd/id/ocw/UPMSW08</a>> locwd:title ?ocwTitle.
 <\!\!\text{http://www.example.org/data/locwd/id/ocw/UPMSW08}\!\!>\!\!\text{locwd:description.}
 <a href="http://www.example.org/data/locwd/id/ocw/UPMSW08">http://www.example.org/data/locwd/id/ocw/UPMSW08</a>> locwd:Prerequisites ?Prerequisites.
 <a href="http://www.example.org/data/locwd/id/ocw/UPMSW08">http://www.example.org/data/locwd/id/ocw/UPMSW08</a> locwd:description ?ocwDescription. }
Results
```

>> Ontologies and Semantic Web, 2008

>> "The general objective is to provide students with a sound grounding of scientific, methodological and technological fundamentals in Ontological Engineering and the Semantic Web areas....

Query B: Show all countries with less than 3 ocw-initiatives

```
PREFIX rdfs: <a href="http://www.w3.org/2000/01/rdf-schema#">http://www.w3.org/2000/01/rdf-schema#</a>
PREFIX type: <a href="http://dbpedia.org/class/yago/">http://dbpedia.org/class/yago/</a>
PREFIX xsd: <a href="http://www.w3.org/2001/XMLSchema#">http://www.w3.org/2001/XMLSchema#</a>
PREFIX locwd: <a href="http://www.example.org/schema/locwd/">http://www.example.org/schema/locwd/</a>
SELECT ?country_name
WHERE { ?ocw a locwd:OCW; locwd:provider ?organization.
 ?organization locwd:ocwConsortium ?consortium; locwd:orgCountry ?country.
 ?country foaf:name ?country_name.
FILTER REGEX (?consortium, 'i', 'Universia'). }
GROUP BY ?country HAVING (COUNT (distinct ?organization ) < 3) .}
Results
>> Argentina
>> Ecuador
>> Peru
```

Other Queries. In the following authors describe the LOCWD data query execution procedures.

- (a) Query OCW course lecturers-authors located by scientific/academic status and/or OCW course importance (measured in terms of OCW visits) in particular fields of knowledge:
- Input parameters: query text, e.g.: "eLearning", "Semantic Web"
- Selected dataset: OCW author particulars, scientific impact factor (each author's h-index can be queried by linking to ISIWeb/DBLP), data indicating the number of visits and social rating of the OCW course (these data can be gathered from the search tool)
- **Process:** Select authors of OCW courses related to the term to be queried, comparing OCW authors according to h-index (or any other index) and data on the use of their OCW courses (or OCW social rating).
- (b) Query OCW repositories according to geographical input information or information on OCW repository features.
- **Input parameters**: (a) geographical information (country, longitude, latitude, city, others) (b) OCW repository: platform type, language, repository CC license, number of courses, associated higher education institution, etc.
- Selected dataset: OCW repository data, and link to external RDF data sources: DBPEDIA, GEONAMES, FREEBASE, etc.
- **Process**: link the repository properties with external RDF data sources.

6. Discussion and Conclusions

The key idea the OER movement is that open educational content should be maximally shared. In this paper, authors have shown that interoperability is an important aspect of open educational contents integration and reuse.

Linked Data is a matter of organizations publishing their data, but even more, of creating links between datasets in order to build interoperability in a global information space based on links: different databases sharing the same vocabulary (a "hub and spoke" approach to map linked data).

The use of linked data approach on OCW repositories provides the framework for their evolution into a more interoperable and integrated system to sharing, connecting and discovering data and metadata of OCW initiatives. Based on the perspective of Linked Open Data, free open OCW data also fosters interoperability and creates a basis on which the use, reuse, remix, and adaptation of open educational tools or commercial applications can be built more easily.

Liked Data is published not only as passive datasets, but as web services that can be utilized by other application (open source or commercial software) via SPARQL web services. As a result, the research team developed and successfully used a software interface that demonstrates how the Linked Data vision enables a new generation of OERs and OCW that can be semantically described and connected with other data and discoverable sources. In the work (Piedra et al., 2012),

authors, show a three-tier client-server architecture using Android framework that integrates into a common interface the OCW linked data management on mobile devices. Moreover, the research team has built a search engine^{††} based on facets that consume Linked Data from LOCWD.

The knowledge organization systems of all kinds are more than ever needed to serve as hubs between datasets described according to different data formats. The vocabulary LOCWD contributes to this approach.

Finally, as regards future work, authors intend to continue research on Linked Data projects in the field of OER and OCW. This new interoperability style, built on Linked Data Design Issues, will allow offering new services, more adapted to personalized and serendipitous discovery of open educational contents.

Acknowledgments

This work has been partially funded by Madrid Regional Government through the project eMadrid (Ref:S2009/TIC-1650) 'Research and Development for e-learning technologies', Project AL12-PID-09 (UPM) Universidad Politécnica de Madrid, Universidad Técnica Particular de Loja and scholarship provided by the "Secretaría Nacional de Educación Superior, Ciencia y Tecnología" of Ecuador (SENESCYT). Authors thank to OCW Consortium, OCW-Universia and technical personal of TAWSBC on UTPL.

References

Alier, M., Mayol, E., Casañ, M. J., Piguillem, J., Merriman, J. W., Conde, M. Á., García-Peñalvo, F. J., et al. (2012). Clustering Projects for eLearning Interoperability. Journal of Universal Computer Science, 18(1), 106-122.

Álvarez, J. M., Labra, J. E., Cifuentes, F., Alor-Hérnandez, G., Sánchez, C., & Guzmán Luna, J. A. (2012). Towards a pan-european e-procurement platform to aggregate, publish and search public procurement notices powered by linked open data: the moldeas approach. International Journal of Software Engineering and Knowledge Engineering, 22(03), 365.

Bechhofer, S., Buchan, I., De Roure, D., Missier, P., Ainsworth, J., Bhagat, J., Couch, P., et al. (2012). Why linked data is not enough for scientists. Future Generation Computer Systems, in press.

Bizer, C. (2009). The Emerging Web of Linked Data. *Intelligent Systems*, *IEEE*, 24(5), 87 –92.

^{††} Serendipity is an faceted search engine (see http://purl.org/ocw/serendipity) As an important feature of Serendipity, Serendipity POIs (Points of Interest), allows users visualize OCW Repositories from an dataset based on locwd (see http://purl.org/ocw/serendipity/pois)

- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked Data The Story So Far. *International Journal on Semantic Web and Information Systems*, 5(3), 1–22.
- Chicaiza, J., López, J., Piedra, N., Martínez, O., & Tovar, E. (2010). Usage of social and semantic web technologies to design a searching architecture for software requirement artefacts. *IET Software*, 4(6), 407.
- Dietze, S., Sanchez-Alonso, S., Ebner, H., Yu, H., Giordano, D., Marenzi, I., Pereira Nunes, B. (2013) Interlinking educational Resources and the Web of Data a Survey of Challenges and Approaches, Emerald Program: electronic Library and Information Systems, Volume 47, Issue 1 (2013).
- Colomo-Palacios, R., Sánchez-Cervantes, J. L., Alor-Hernández, G., & Rodríguez-González, A. (2012). Linked Data: perspectives for IT professionals. *International Journal of Human Capital and Information Technology Professionals*, 3(2), 1–15.
- Dovrolis, N., Stefanut, T., Dietze, S., Yu, H. Q., Valentine, C., & Kaldoudi, E. (2012). Semantic Annotation and Linking of Medical Educational Resources. In Á. Jobbágy, R. Magjarevic, & R. Magjarevic (Eds.), 5th European Conference of the International Federation for Medical and Biological Engineering, IFMBE Proceedings (Vol. 37, pp. 1400–1403). Springer Berlin Heidelberg.
- Fernandez, M., d' Aquin, M., & Motta, E. (2011). Linking Data across Universities: An Integrated Video Lectures Dataset. In L. Aroyo, C. Welty, H. Alani, J. Taylor, A. Bernstein, L. Kagal, N. Noy, et al. (Eds.), *The Semantic Web ISWC 2011*, Lecture Notes in Computer Science (Vol. 7032, pp. 49–64). Springer Berlin / Heidelberg.
- García-Barriocanal, E., Sicilia, M.-A., & Sánchez-Alonso, S. (2012). Providing semantic metadata to online learning resources on sustainable agriculture and farming: combining values and technical knowledge. *Interactive Learning Environments*, *In press*. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/10494820.2011.559170
- García-Crespo, Á., Colomo-Palacios, R., Gómez-Berbís, J. M., & García-Sánchez, F. (2010). SOLAR: Social Link Advanced Recommendation System. *Future Generation Computer Systems*, 26(3), 374–380.
- García-Peñalvo, F. J., Colomo-Palacios, R., & Lytras, M. (2012). Outcomes of International Research Projects on Technology Applied to Education. *Journal of Universal Computer Science*, 18(1), 1–4.
- García-Peñalvo, F. J., Conde, M. Á., Alier, M., & Casany, M. J. (2011). Opening Learning Management Systems to Personal Learning Environments. *Journal of Universal Computer Science*, 17(9), 1222–1240. García-Peñalvo, F. J., Figuerola, C. G. de, & Merlo, J. A. (2010). Open knowledge: challenges and facts. *Online Information Review*, 34(4), 520–539.
- Hausenblas, M. (2009). Exploiting Linked Data to Build Web Applications. *Internet Computing*, *IEEE*, *13*(4), 68 –73.

Joerg, B., Ruiz-Rube, I., Sicilia, M.-A., DvořáK, J., Jeffery, K., Hoellrigl, T., Rasmussen, H. S., et al. (2012). Connecting closed world research information systems through the linked open data web. International Journal of Software Engineering and Knowledge Engineering, 22(03), 345.

Passant, A., Laublet, P., Breslin, J. G., & Decker, S. (2009). A URI is Worth a Thousand Tags. International Journal on Semantic Web and Information Systems, 5(3), 71-94.

Piedra, N., Chicaiza, J., Lopez, J., Tovar E., Martinez-Bonastre, O. (2012) Combining Linked Data and Mobiles Devices to improve access to OCW. Global Engineering Education Conference (EDUCON), 2012 IEEE. 10.1109/EDUCON.2012.6201202

Rodriguez, D., Sicilia, M. A., Cuadrado-Gallego, J. J., & Pfahl, D. (2006). e-Learning in Project Management Using Simulation Models: A Case Study Based on the Replication of an Experiment. Education, IEEE Transactions on, 49(4), 451 -463.

Rodríguez-González

Ruiz-Calleja, A., Vega-Gorgojo, G., Asensio-Pérez, J. I., Bote-Lorenzo, M. L., Gómez-Sánchez, E., & Alario-Hoyos, C. (2012). A linked data approach for the discovery of educational ICT tools in the Web of Data. *Computers & Education*, 59(3), 952–962.

Tovar, E., Piedra, N., Chicaiza, J., Lopez, J., & Martinez-Bonastre, O. (2012). OER Development and Promotion. Outcomes of an International Research Projecton the OpenCourseWare Model. Journal of Unviersal Computer Science, 18(1), 123–141.

Tzikopoulos, A., Manouselis, N., Kastrantas, K., & Costopoulou, C. (2012). An online information system to support blended training of rural SMEs on e-government. Program: electronic library and information systems, 46(1), 123–143.

Yu, H. Q., Pedrinaci, C., Dietze, S., & Domingue, J. (2012). Using Linked Data to Annotate and Search Educational Video Resources for Supporting Distance Learning. IEEE Transactions on Learning Technologies, 5(2), 130-142.