A Framework for improving the effectiveness of the Openness in OER Repositories and Open Educational Datasets

FROM OER TO OPEN OER DATA

Nelson Piedra, Autor2, Autor3
(Universidad Técnica Particular de Loja
nopiedra@utpl.edu.ec)

Autor4
(Institución2
[]@example.org)

Abstract—One of the two fundamental concepts related to OER is "the ability to freely adapt and re-use existing pieces of knowledge". Discovering and re-use of educational resources by both individuals and organizations may have significant creative and economic benefit for educational environment. To facilitate discovering, combining, reusing, integration or adaptation of OER, a key condition is to improve the metadata interoperability between different collections of open material. In this paper, a set of criteria for assessing the effectiveness of the openness in OER Repositories based on Semantic Linked Open Data guidelines is presented. Based on the principles for opening of Open Data, an OER 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-proprieties, ensures interoperability between different collections of OER using open licenses both a human-readable description and computer-readable metadata, and persistence. Also, this study advocates 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.

I. INTRODUCTION

It is widely recognized that re-use of educational resources by both individuals and organizations may have significant creative and economic benefit for educational environment [7]. In this way, one major barrier on the way to a sustainable open educational environment based on open educational resources is an appropriate global - scale interoperable and integrated approach ecosystem.

A global -scale interoperable and integrated approach consists of a legal framework that promotes free and open access to educational materials, a focus of quality assurance in shared academic resources, and a technological vision that through the Web allows automatic processing of information associated with OER. Although, the OER itself has been described as early as 2001, detailed reusable insight into the relationship between OER offers and OER re-utilization has remained scarce.

Linked data is mainly about publishing structured data in RDF using URIs rather than focusing on the ontological level or inference [2][3][4]. In this work, the authors demonstrate that the use of linked open data on OER repositories provides the framework for their evolution into a more interoperable and integrated system for sharing [5], connecting and discovering resources, data and metadata of OER initiatives. Moreover, OER resource metadata can be enriched using datasets hosted by the Linked Open Data cloud.

II. OPEN DATA AND OPEN CONTENT IN THE CONTEXT OF OPEN EDUCATIONAL INITIATIVES

The Open Definition1 sets out principles that define “openness” in relation to data and content. It

1 Open Definition. See more at: http://opendefinition.org
makes precise the meaning of “open” in the terms “open data” and “open content”, and thereby ensures interoperability and integration between different initiatives of open educational resources. 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” [6].

David Wiley [9] indicates that the open content is content that is licensed in a manner that provides users with the right to make more kinds of uses than those normally permitted under the law. The primary permissions or usage rights open content is concerned with are expressed in the "5Rs Framework:"

- **Retain** - the right to make, own, and control copies of the content (e.g., download, duplicate, store, and manage)
- **Reuse** - the right to use the content in a wide range of ways (e.g., in a class, in a study group, on a website, in a video)
- **Revise** - the right to adapt, adjust, modify, or alter the content itself (e.g., translate the content into another language, adapt images)
- **Remix** - the right to combine the original or revised content with other open content to create something new (e.g., incorporate the content into a mashup)
- **Redistribute** - the right to share copies of the original content, your revisions, or your remixes with others (e.g., give a copy of the content to a friend)

These permissions are granted to users free of charge.

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.

### III. USE AND REUSABILITY IN OER DEVELOPMENT

We agree that all human beings are endowed with a capacity to learn, improve, and progress. Educational opportunity is the mechanism by which we fulfill that capacity. When educational materials can be electronically copied, reused, adapted and transferred around the world at almost no cost, we have a greater ethical obligation than ever before to increase the reach of opportunity [4].

#### A. The reusability of OER

Open educational digital materials and repositories occur in a rich diversity of types and representations. They are linked to the specific application packages that were used to create or manage them. They are easily misidentified. They are generally poorly described or annotated; they often have insufficient metadata attached to them to avoid their gradual susceptibility to loss of value. OER needs to be not only gathered and stored, but also made useful and visible, a process that takes substantial human work, even if heavy automation can aid in the process.

The reusability of OER is about maintaining the semantic meaning of the educational material and its context, about maintaining its provenance and quality, about retaining its interrelatedness, and about securing information about the context of its creation, use, reuse and adaptation, and the right semantic representation and data used to describe the OER.

#### B. Enhance the reusability of OER

One of the two fundamental concepts related to OER is "the ability to freely adapt and re-use existing pieces of knowledge" [1]

Reuse is to make use of a resource for other aim, usually for a purpose unintended by the original creator. Thus creators of OER should consider the degree to which they want their OER to be open, and license the resource accordingly. In addition for licensing there are technical aspects that make OER suitable for a new use or purpose, easier to discover, adapt and remix, and consequently affect the level of openness of an OER. This implies the right to adapt, adjusts, modify, or alter the content itself.

The openness of a digital resource is to work in the open as much as possible. It’s not enough to just have a policy or principle of action of openness adopted or proposed by a government, consortium, organization or individual. OER initiatives seek to be as understandable, reusable, adaptable, and inspectable as possible. Consequently, one of the primary benefits of an OER is that it can be discovered and adapted to the needs of specific situations.
The OER should be designed to be easily adaptable for other users. It should have metadata sufficient for discoverability. 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, re-mix, adapt and re-distribute.

C. Enhance the discoverability of OER

Most of OER repositories are licensed under Creative Commons Licenses. The use of open licenses can help users to discover materials that they know can use, reuse, adapt and redistribute.

Different studies have highlighted the difficulty, finding OERs and how this affects their use. In [8], some of the causes that affect the location of OERs are identified: technical issues around search engines and repositories, practical searching skills and the volume of available resources in different subject areas. In [8], authors identified 3 factors that influence the decision to reuse of digital material: improve quality, meet a teaching need and peer suggestion.

While discoverability is probably the major barrier to reuse, tutors are still expecting to find useful materials online and are prepared to spend time searching for them [8].

IV. Framework Proposed.

Based on the principles for opening of Open Government Data[2], an OER 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.

A. Completeness

The repositories, resources and datasets released by OER Initiatives should be as complete as possible, reflecting the entirety of what is shared about a particular subject. All raw information from a repository or a specific OER should be released to the public. Metadata that defines and explains the raw data should be included as well, along with practices and explanations for how derived data was calculated. Doing so will permit users to understand the scope of information available and examine each OER at the greatest possible level of detail.

B. Primary

Repositories released by the OER initiatives should be primary source digital resources and data, with the highest possible level of granularity and detailed way that is practicable. This includes the original information that describe the OER, details on how the digital resource was created and how it can be used, modified or adapted for others to build their own materials.

C. Timeliness

Both, digital resource released by an OER Initiative, and their metadata should be available to the teachers, students, and self-learning’s in a timely fashion.

If OER creator adds the OER dataset to a catalog, such as the Data Hub[3], creator should make sure that indicate the license under which the dataset is available within that catalog. This gives people searching the catalog a quick and easy way of seeing that they will be able to reuse the dataset.

D. Ease of access to digital resources and metadata.

The repositories and resources released by OER Initiatives should be as accessible as possible in convenient, modifiable, and open formats that can be retrieved, re-used, re-mixed, adapted, downloaded, indexed, and searched. An aspect of this is "findability," which is the ability to easily locate and download content (digital resources and metadata).

Barriers to automated access include making data accessible only via submitted forms or systems that require browser-oriented technologies. By contrast, providing an interface for users to download all of the information stored in a database at once (known as "bulk" access) and the means to make specific calls for data through an Application Programming Interface (API) or if possible a Sparql EndPoint that make

---


[3] Datahub is the data management platform from the Open Knowledge Foundation, based on the CKAN data management system.http://datahub.io/
data much more readily accessible.

E. Metadata documented

Because the purpose of OER is reuse, user needs that resources and metadata be as well documented and standardized as possible. What those terms mean depends very much upon the data and potential uses. This challenge is often framed in terms of properly identifying what to collect, or perhaps as a challenge in filtering the great mass of content from which one must carefully select.

F. Metadata in Standard and Machine readability Formats

Formats of open-OER-data should be machine-readable (i.e., data are reasonably structured to allow automated processing). Metadata should be in standard formats to ease processing, for pattern recognition, mining, integration, interoperability, simulation, longitudinal studies, and so forth.

From the perspective of semantic web technologies, OER initiatives should publish a web service of type SPARQL endpoint. A SPARQL endpoint enables users (human or other) to query a knowledge base via the SPARQL language. Results are typically returned in one or more machine-processable formats. Therefore, a SPARQL endpoint is mostly conceived as a machine-friendly interface towards a knowledge base. Both the formulation of the queries and the human-readable presentation of the results should typically be implemented by the calling software, and not be done manually by human users.

G. Universal Participation

Everyone (teachers, students, self-learners), must be able to use, reuse and redistribute — there should be no discrimination against fields of endeavor or against persons or groups.

H. Formats non-proprietaries

Open OER Data do not discriminate against any person or group of persons and should be made available to the widest range of users for the widest range of purposes, often by providing the data in multiple formats for consumption. To the extent permitted by law, these formats should be non-proprietary, publicly available, and no restrictions should be placed upon their use.

I. Ensures interoperability between different collections of OER using open licenses both a human-readable description and computer-readable metadata.

Open licensing stipulates very few restrictions on what anyone can do with the content or data that is being licensed. An open license allows others to do things like:

- Share or republish the content or data
- Derive new content or data from yours
- If the license allows, make money by selling products that use this content or data

Creators can choose to make your content or data available under one of three levels of license:

1. A public domain license has no restrictions at all (technically, these indicate that creator waive your rights to the content or data)
2. An attribution license just says that users must give attribution to you
3. An attribution & share-alike license says that users must give attribution and share any derived work under the same license

OER creator should indicate the license for content or data that make available using both a human-readable description and computer-readable metadata.

The human-readable descriptions and marks that creator should use are spelled out on the Creative Commons and Open Data Commons websites:

- Creative Commons license chooser
- Open Data Commons licenses

According to the above Open Definition, there are only two kinds of restrictions that an open license can place:

- That users must give attribution to the source of the content or data
- That users must publish any derived content or data under the same license (this is called share-alike)

J. Persistence

When OER is made available on the Web, it is important for the integrity of the Web, and the society based upon it, that the digital resources, metadata and specially the URIs used to reference information be

---

4 Creative Commons Choose: http://creativecommons.org/choose/
5 Open Data Commons, Licenses: http://opendatacommons.org/licenses/
used well into the future, and that the information persist as identified.

The OER initiative will ensure that persistent resources continue to be available throughout the life of the organization⁶. Where a persistent resource is modified, a change history will be archived though the archive will not necessarily be available publicly.

The intent is reduce the failure of links due to uncoordinated management or inadequate commitment to information persistence, and to provide a stable reference base of information about Open Educational Resources as a service to the community.

V. CONCLUSION

In this paper, authors have put forward different criteria for improving the openness of OER collections available in the Web. Linked Open 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.

REFERENCES


⁶ See also W3C URI Persistence http://www.w3.org/Consortium/Persistence.html