Abdel-Qader, Mohammad; Saleh, Ahmed; Tochtermann, Klaus

Conference Paper

On the Experience of Federating Open Educational Repositories using the Learning Object Metadata Standard

Suggested Citation: Abdel-Qader, Mohammad; Saleh, Ahmed; Tochtermann, Klaus (2021) : On the Experience of Federating Open Educational Repositories using the Learning Object Metadata Standard, In: Gómez Chova, L. et. al. (Ed.): Proceedings of the 13th International Conference on Education and New Learning Technologies (EDULEARN 2021), 5th-6th of July, 2021, ISBN 978-84-09-31267-2, IATED Academy, Valencia, pp. 4819-4825, http://dx.doi.org/10.21125/edulearn.2021.0998

This Version is available at:
http://hdl.handle.net/11108/497

ON THE EXPERIENCE OF FEDERATING OPEN EDUCATIONAL REPOSITORIES USING THE LEARNING OBJECT METADATA STANDARD

M. Abdel-Qader, A. Saleh, K. Tochtermann

ZBW -- Leibniz Information Centre for Economics (GERMANY)

Abstract

Within the context of this paper, we understand Open Educational Resources (OERs) as freely available educational materials. In order to store OERs metadata in a structured form, modeling standards should be utilized. However, OER providers often do not employ such a standard. This leads to limitations in the development of federated systems that link multiple OER repositories with each other. Such federated systems, in turn, facilitate the search for OER across repository boundaries to help educators (teachers, students, etc.) to search for OER with different origins. In this paper, we describe our methodology for connecting different OER repositories based upon the Learning Object Metadata standard (LOM). For this purpose, we have carefully selected 29 OER repositories that are openly available on the web and hosted by German institutions. Some of the key outcomes of our research are: (i) only 17% of the analyzed repositories employ LOM as a standard to model their data, (ii) only 12% of the metadata from the repositories will get lost when federating these repositories using the LOM standard, (iii) the lost metadata has no impact on the quality of the retrieved OERs, as it is insignificant information about the OERs, such as the number of views and likes. Based on this experience, we recommend that OER repositories use the LOM standard to describe OERs, as this has several advantages, such as facilitating the federation of repositories and increasing the accessibility of OERs.

Keywords: Open Educational Resources, LOM, Open Educational Repositories, Federating Open Educational Repositories.

1 INTRODUCTION

Open Educational Resources (OERs) are all the educational materials that are published under an open access license [1]. OERs include images, videos, text, or any other formats that can be freely used. OER providers offer these resources in different formats. The data describing OERs (e.g. content, author, contributor etc.) is represented and stored as so-called metadata. Usually, metadata is stored in a structured format, referred to as schema or standard.

Due to the availability of different OER standards, OER providers employ these standards to represent the metadata of their OERs. However, many providers do not employ a structured standard for their OER metadata or the standard is not widely accepted in the community and thus often unknown. A crucial requirement for standards is the applicability to different languages and the possibility to represent different object types. For example, some OER repositories are using the Learning Object Metadata (LOM) standard [2], while others use the standard of the Learning Resource Metadata Initiative (LRMI) [3] or the Metadata Object Description Schema (MODS) [4]. LOM is an international standard that aims to facilitate the search for and use of eLearning objects for learners, teachers, and automated software processes. Many OER providers employed the LOM standard, such as the OER repository of the German State Bavaria including OER by higher education institutes (openVHB) [5] and memucho [6]. The use of different standards can lead to problems in merging these different OER repositories into one central repository. However, connecting different OER repositories into one repository has many advantages, such as facilitating the federation of repositories and increasing the accessibility of OERs.

In this paper, we describe a methodology for connecting different OER repositories using the LOM standard. We have carefully selected 29 OER repositories that are openly available on the web and

hosted by German institutions. We study these repositories by analyzing the metadata standard that the providers use to model their data. In the beginning, we investigate which standard the OER providers use to model their metadata. Subsequently, we analyze the repositories that do not use the LOM standard in order to count the fields that will be lost when connecting these repositories using LOM. Finally, we illustrate the process of connecting OER repositories using the LOM standard.

Our results show that more than half of the analyzed repositories use well-known standards to represent their metadata. Furthermore, only 12% of the fields will be lost when we map the metadata using the LOM standard. However, our analysis shows that the lost data does not have an impact on the quality and clarity of the OER metadata. Therefore, connecting the OER repositories using the LOM standard will facilitate searching for OERs. Furthermore, it will make OERs more reachable to the different learning parties, such as students, teachers, and others.

The remainder of this paper is structured as follows. We review related work in Section 2. A description of the LOM standard and its features, and the federation approach are described in Section 3. We present our results in Section 4 and discuss them in Section 5, before we conclude.

2 RELATED WORK

A good overview of the complexities of federating data infrastructures is provided in [7]. Latif et al. identified a list of challenges during their research on Open Science (OS) projects. The authors illustrate three challenges for federating data infrastructures: harvesting, common metadata models, and metadata mapping tools. Our federation approach presented in Section 3 takes into account many of the challenges addressed in this paper. However, because of the focus on OER, we would like to give more weight to publications from this area here.

Creating and publishing OERs has become increasingly important these days, particularly during the Corona pandemic. As a consequence, many projects for managing OERs have been initiated and provide good solutions. Furthermore, several studies analyzed the quality of the OERs and their repositories and developed platforms for recommending OERs to educators.

Shelton et al. [8] discussed the system of an open-source project that contains a full-text search of OERs (Folksemantic), OER recommendations, and personalized recommendations. The authors described the implementation and evaluation of their system. Using user test protocols and questionnaires, the system was evaluated by teachers and OER providers. Ambite et al. [9] described an index to store the open educational resources for data science. The developer of such an index used Schema.org as a standard for modeling the metadata that will be stored in the index. The authors connected their OER to the web of linked data by referencing the resources to entities in DBpedia, DBLP, and ORCID.

Chicaiza et al. [10] proposed a framework for recommending OERs. The authors used the features of the Semantic Web to search for online resources. The proposed framework takes into account the user's profile when searching for OER. Santos-Hermosa et al. [11] provided a study showing the state of the international OER repositories. The authors conducted a series of educational indicators. The goal of these indicators is to check if the repositories of OERs meet two perspectives, the reuse of OERs and the education context. The authors found that most of the OER repositories are designed only of OER and cannot include research content. Furthermore, they found that a few repositories meet the educational aspects.

Atenas and Havemann [12] review the literature of the Open Educational Resources and their Repositories (ROERs). The authors identified the themes and the quality indicators for designing a ROER. The quality indicators they used to evaluate the effectiveness of ROER are featured resources, user evaluation tools, peer review, authorship of the resources, keywords of the resources, use of standardized metadata, the multilingualism of the repositories, the inclusion of social media tools, specification of the creative commons license, and availability of the source code or original files. The authors analyzed 122 journal articles, conference papers, and books. The authors' main aim from their study is to establish a method to evaluate the repositories of OERs.

3 FEDERATION APPROACH

This section briefly describes the LOM standard and its main properties (Section 3.1). We then describe our methodology for connecting different OER repositories using the LOM standard (Section 3.2).

3.1 The LOM Standard

LOM [2] is a standard for modeling data. It is used for representing educational resources and materials. The modeled data is stored in a digital form. The LOM standard aims to improve the reusability of educational resources between the different educational parties, such as students and teachers. The LOM standard consists of elements to describe the educational resources. These elements are called fields. The fields can be classified into different levels. The main first level consists of nine elements. Each element consists of sub-elements that contain the data. Furthermore, the sub-elements may contain another level of sub-elements. The data that will be modeled using the LOM standard follow the specifications of the values and the data types that can be stored. For example, some elements must follow the rules for storing the dates and times, while some other roles specify how to store the fields that have the language information. Furthermore, some fields in LOM allow the users to enter a free text without any roles or specifications.

3.2 Methodology

Any OER data source provider could use any of the available standards to model their metadata, as metadata modeling is not solely dependent on existing standards. Furthermore, the OER providers could model the data using their own model that does not rely on any available standards. As a consequence of connecting different OER repositories, we need to analyze each OER data source to specify the modeling scheme they use. To achieve this goal, we applied the following steps to connect different OER resources using the LOM standard:

1. Analyze each OER repository to specify the representation model they use.
2. Excludes the OER repositories that already use LOM.
3. In the case of OER repositories that do not use LOM and use other existing standards, we decide how to map their current standard to LOM. In this step, we determine how much data we will lose since not all the fields can be mapped into LOM.
4. The last group is the OER repositories that do not use any of the available standards. This group will be the hardest one to map since each source can use its own vocabulary to represent the metadata.
5. All data sources require a harvester to collect the metadata. The design of the harvesters depends on the standard used in the OER resource. We need to build a focused harvester for each standard used. In addition, a mapper is also needed for each standard.

The process of connecting OERs using the LOM standard is divided into the following steps:

1. Web harvesting: This is the process of collecting (crawling) data from websites [13]. The crawlers can look for any information, and its role is to collect websites. Web crawlers are also known as Web spiders or Web robots.
    - General harvesting: The general harvester is a crawler that can be used for collecting data from any given URL of repositories [13]. From its specification, the general harvester's main advantage is that you construct the harvester once and then use it for any repository. The main disadvantage is that the amount of harvester data using the general harvester is limited compared to the focused harvesters explained below.
    - Focused harvesting: The focused harvester is designed to crawl data on a specific topic [14]. Furthermore, it can be designed to harvest data from a specific website since each website has a different structure. The main advantage of the focused harvester is that it can collect as much data as possible. The focused harvester goes through the website and collects the items needed to represent the OER. The disadvantage of the focused harvesters is that they only work for a specific repository. Thus, we need a harvester for each repository we want to harvest.

2. Mapping: The harvested data from websites are modeled using the known standards or the free ones. The providers of OERs can name the fields that represent the metadata in any name. Therefore, the mapping (translating) process is used to map the field names from the form they were harvested to the new field names. In our case, the new fields are those specified by LOM.
    - General mapping: As in general harvesting, general mapping is the process of translating the harvested data into the LOM standard. Here, we need one mapper to perform such a process since the harvested data have the same field name when using the general harvesters.
    - Focused mapping: Using focused harvesters, we need a harvester for each repository, and each one has a different structure and field names. Therefore, we need a mapper for each focused harvester we build.
3. Storing the results: The processed data that are resulted from the mapping process are stored in a format called the JavaScript Object Notation (JSON). JSON is a structured format to store data that is human-readable and can be parsed by computers [15]. This will allow us to easily store, process, and exchange the data. The metadata in the JSON files can be stored in an index. This index can be used as a database for the information of the OERs. Then, using this database, a platform can be developed to present the metadata stored. The development of a search engine using these metadata will allow users to access these OERs.

The process of harvesting and mapping different OERs using the LOM standard is shown in Fig. 1.

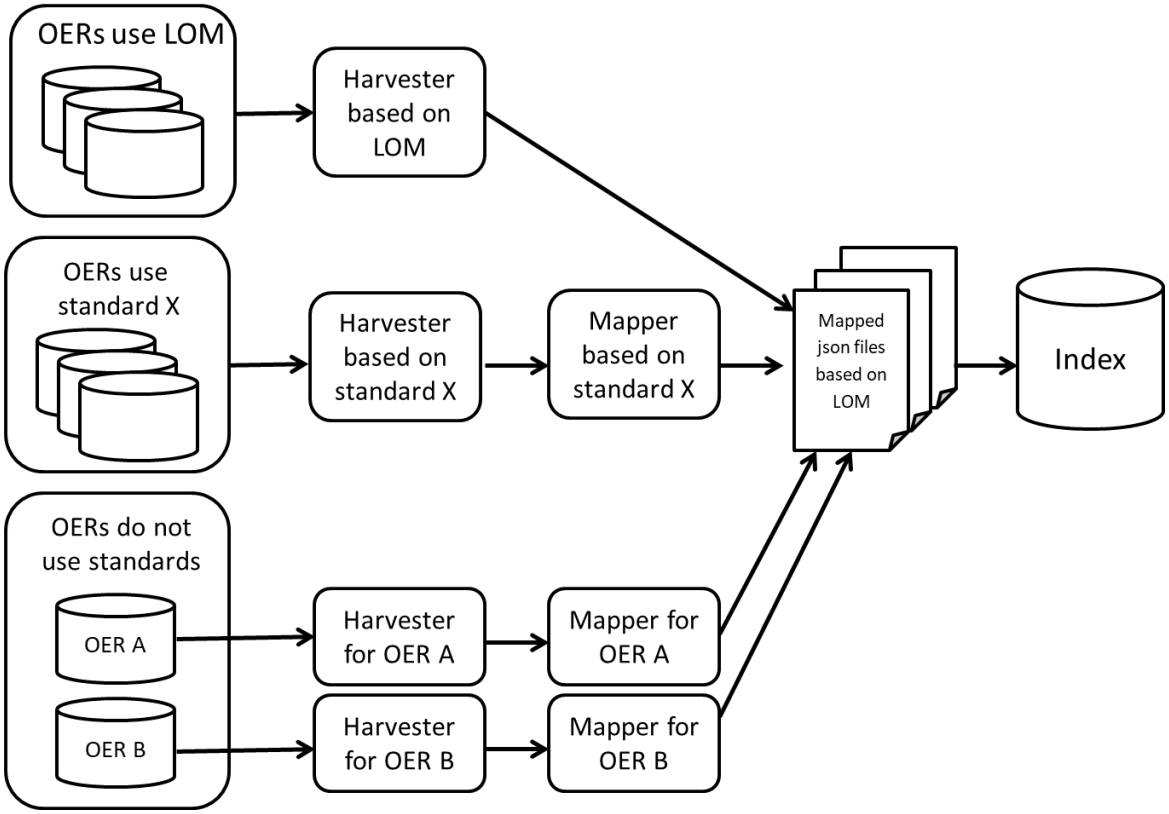

Figure 1. The different scenarios of the process of harvesting and mapping the OERs metadata into the LOM standard starting from harvesting the metadata until storing the results into an index.

4 RESULTS

In Table 1, we list the names of the 29 OER repositories with some information, such as the standard they use and whether they provide an Application Programming Interface (API) or not. The APIs allow applications to communicate with each other.

Table 1. Overview of the name of the 29 OER repositories, the standard they use, and if the repository provide API.

Repository name	Standard	API?
Chemnitz	Free/Unknown	No
Darmstadt	MODS	No
digLL-Hessen	LOM	No
DuE Publico2	MODS/MyCoRe	Yes
Elixier	LRMI	No
Heureka.NRW	Free	No
Hochschule für Musik	Fee	No
HOOU	LRMI	No
Koln	Free/Unknown	No
Konstanz University	LOM	No
Kunst Labore	Free	No
lecture2go	Free	No
Mainz	Free/Unknown	No
mediathek.hhu.de	LRMI	No
Memucho	LOM	No
Niedersachsen	Free/Unknown	No
OER UDE	DC	Yes
open vhb	LOM	No
OpenLearnWare	Free/Unknown	No
openRUB	Free/OCW	No
PeDOCS	DC	Yes
PubServer Braunschweig	MODS	Yes
Serlo	Free	No
Smart VHB	Free	No
The Open University	Moodle	No
UniWeimar	DC	Yes
VCRP	Free	No
ZOERR	LOM	No
ZUM	Free	No

The analysis of 29 OER repositories shows that more than half of the repositories use known standards, as shown in Table 2. These repositories are openly available on the web and hosted by

German institutions. Furthermore, 17% of the analyzed repositories use LOM as a standard to model their metadata.

Table 2. The percentage of repositories that use LOM, other standards, or do not use any standard to model their metadata.

Standard	Count	Percentage
LOM	5	17%
Other standards	11	38%
None	13	45%

Our analysis shows that only five repositories (17%) provide an API that allows them to access and save their metadata. In order to connect the different data sources using the LOM standard, we first need to harvest the metadata of each repository. Second, we map the harvested metadata into the LOM standard. The APIs make the process of harvesting the metadata easier. However, since only 17% of the repositories provide an API, the harvesting process becomes more difficult and time-consuming due to the different structures of each repository's websites.

Only 12% of the harvested data will get lost when we map them into the LOM standard. This percentage seems to be high, but most of the lost data is not essential for defining and describing the OER. Examples of lost data are the number of views and likes. Thus, in terms of discoverability, connecting the OER repositories using the LOM standard does not have a significant impact on the quality of the “normalize” metadata of OERs.

We need to divide these repositories into two groups to connect the different repositories that use different standards. The first group contains the repositories that provide APIs. This will make harvesting the metadata more efficient. The second group is the one that does not provide APIs. Here, we need to build a harvester that collects data. This will lead to the loss of some hidden data that are not shown in the portal.

Subsequently, the harvested data will be divided into two groups. The first one is the group for the repositories that use the LOM standard to model their metadata. We do not need a mapper for this group to translate the data from one standard to another one since the harvested data is already mapped into LOM. The second group contains the remaining repositories. This includes the ones that use other standards and the ones that do not use standards at all. For this group, we need to build a mapper for each standard and a mapper for each repository that does not use standards.

5 DISCUSSION

While some repositories still use their own model, known standards for representing the metadata of OERs are widely used by OER repositories. More than half of the analyzed repositories are using known standards. We expect this percentage to increase due to widespread technology and the clear trend towards open and distance learning. Furthermore, the COVID-19 pandemic further promoted the creation and publication of OERs.

Existing APIs facilitate the process of harvesting and linking different OER repositories. They allow users to harvest metadata as much as possible and in an appropriate format. However, based on our analysis, only a low percentage (in our case, 17%) of the repositories provide API to access the OERs. We think that APIs could make the linking process of OER repositories more efficient as the harvesting and mapping process requires less effort than creating a focused harvester for the repositories that do not provide APIs.

Based on our analysis, only a low percentage (in our case, 12%) of the harvested metadata will be lost if we mapped them using the LOM standard. The good news is that for discovery purposes, the lost metadata will have no impact on the quality of OER. The main fields that can represent and provide sufficient information about OER are represented in the LOM standard.

Using the LOM standard to connect OER repositories can increase interoperability between them. Furthermore, when connecting such repositories using a standard such as LOM will make it easier to reach their OERs through search.

The LOM standard is used for higher education OER repositories. No such standard can represent all the data from different OERs. Using LOM shows that the amount of lost data is reasonable, and the kind of these lost data is not essential, so that the OER will be inefficient without those lost data. If we can connect different repositories using one standard, we have a significant advantage in finding OERs and enabling them to reach as many users as possible. Furthermore, providing a platform for the connected repositories using our methodology can increase the reachability of the OERs.

6 CONCLUSIONS

Modeling standards used to store OERs metadata in a structured form. However, OER providers often do not employ such a standard. This leads to limitations for searching for OERs and makes them less reachable. Therefore, we think that we need a way to make the OERs more reachable through search. The availability of different modeling standards allows OER providers to model their data into different formats. These formats make the process of connecting the OER repositories a hard task. Thus, connecting such repositories using one standard achieves the goal of making OERs more reachable. Our experiments show that when connecting OER repositories using the LOM standard applicable. The amount of lost data when making this connection is around 12%, and they are affecting the quality of OERs since these lost data are minor ones.

ACKNOWLEDGEMENTS

The project results have been developed by the EDUArc project, which is funded under the grant agreement number 16DHB2128 of the German Ministry for Science and Education.

REFERENCES

[1] J. Hylén, “Open educational resources: Opportunities and challenges,” Proceedings of open education, vol. 4963, 2006.

[2] U. RISK, “Draft standard for learning object metadata,” IEEE standard, vol. 1484, no. 1, 2002.

[3] P. Barker and L. M. Campbell, “LRMI, learning resource metadata on the web,” in Proceedings of the 24th International Conference on World Wide Web, pp. 687–687, 2015.

[4] S. H. McCallum, “An introduction to the metadata object description schema (MODS),” Library hi tech, 2004.

[5] The OER repository in Bayern, Accessed 6 May, 2021. Retrieved from https://open.vhb.org/

[6] Memucho, Accessed 6 May, 2021. Retrieved from https://memucho.de/

[7] A. Latif, F. Limani, and K. Tochtermann, “On the complexities of federating research data infrastructures,” Data Intelligence, vol. 3, no. 1, pp. 79–87, 2021.

[8] B. E. Shelton, J. Duffin, Y. Wang, and J. Ball, “Linking open course wares and open education resources: creating an effective search and recommendation system,” Procedia Computer Science, vol. 1, no. 2, pp. 2865–2870, 2010.

[9] J. L. Ambite, J. Gordon, L. Fierro, G. Burns, and J. Mathew, “Linking educational resources on data science,” in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, pp. 9404–9409, 2019.

[10] J. Chicaiza, N. Piedra, J. Lopez-Vargas, and E. Tovar-Caro, “Recommendation of open educational resources. an approach based on linked open data,” in 2017 IEEE Global Engineering Education Conference (EDUCON), pp. 1316–1321, IEEE, 2017.

[11] G. Santos-Hermosa, N. Ferran-Ferrer, and E. Abadal, “Repositories of open educational resources: An assessment of reuse and educational aspects,” International Review of Research in Open and Distributed Learning: IRRODL, vol. 18, no. 5, pp. 84–120, 2017.

[12] J. Atenas and L. Havemann, “Questions of quality in repositories of open educational resources: a literature review,” Research in Learning Technology, vol. 22, 2014.

[13] C. Olston and M. Najork, Web crawling. Now Publishers Inc, 2010.

[14] J. Johnson, K. Tsioutsiouliklis, and C. L. Giles, “Evolving strategies for focused web crawling,” in Proceedings of the 20th International Conference on Machine Learning (ICML-03), pp. 298–305, 2003.

[15] N. Nurseitov, M. Paulson, R. Reynolds, and C. Izurieta, “Comparison of JSON and XML data interchange formats: a case study,” Caine, vol. 9, pp. 157–162, 2009.