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

Conference Paper — Accepted Manuscript (Postprint)

EduArc: Federated Infrastructures for Digital Educational Resources

Suggested Citation: Saleh, Ahmed; Abdel-Qader, Mohammad; Tochtermann, Klaus (2022) : EduArc: Federated Infrastructures for Digital Educational Resources, In: Tseng, Yuen-Hsien et al. (Ed.): From Born-Physical to Born-Virtual: Augmenting Intelligence in Digital Libraries, 24th International Conference on Asian Digital Libraries, ICADL 2022, Hanoi, Vietnam, November 30 – December 2, 2022, Proceedings, Springer, Cham, pp. 501-508, https://doi.org/10.1007/978-3-031-21756-2_42

This Version is available at:
http://hdl.handle.net/11108/636

EduArc: Federated Infrastructures for Digital Educational Resources

Ahmed Saleh, Mohammad Abdel-Qader, and Klaus Tochtermann

ZBW – Leibniz Information Centre for Economics, Kiel/Hamburg, Germany
{a.saleh,m.abdel-qader,k.tochtermann}@zbw.eu

Abstract. The goals of the EduArc project are to explore the conditions for successful dissemination of open educational resources (OER) at universities and the provision of a federated infrastructure for digital educational resources. The project aims to develop a design concept (EduArc platform) for distributed learning infrastructures that provide federated access to digital educational resources. The EduArc platform connects decentralized systems via open standards and interfaces and is open for the integration of content from future providers and users.

Keywords: Digital Libraries · Search Engine · Learning Management System.

1 Introduction

The project addresses the question of what infrastructural requirements are necessary to build a sustainable platform for digital educational resources for universities. Marine et al. published a study on the state of digital educational resources (OER) in the context of digital transformation of various higher education systems [1]. The results of the study were taken into consideration while developing EduArc infrastructure. In order to successfully build such a platform, the functional requirements and design concepts were defined through use case studies. Afterwards, a common data model (CDM) for learning resources was developed based on the Learning Object Metadata standard (LOM) [2]. The common data model enables the metadata of open educational resources from different repositories to be mapped to the EduArc standard and stored in a central search index (EduArc index). Based on this common data model, a concept for an overall system architecture of the EduArc platform for connecting distributed learning resources was designed. A search engine, that searches for OERs based on the stored metadata, was developed. In addition, we developed various subject indexing models to analyze the indexed documents and generate new metadata. The generated metadata contains additional information about the content of a document (e.g., business, computer science, etc.). Finally, the repositories used in the EduArc project were documented in the common project wiki. A guide for linking repositories to EduArc has been made publicly available. This guide allows universities to link their repositories to the EduArc infrastructure. More information can be found in the reproducibility section.

2 Case studies and Requirements Specification

A user-centered approach was used. The approach focuses on the needs of the users and includes evaluations at different stages of the development process with appropriate feedback in the design or development process. The following research questions were defined:
RQ1: How do educators currently work with resources?
RQ2: How can EduArc be integrated into teachers' workflows?
RQ3: What are the functional requirements of EduArc from the teachers' perspective?
Case studies were developed to answer these three sequential research questions.

2.1 Interviews: Data Collection

Data collection and data analysis were conducted in an iterative process, resulting in a robust requirements specification. One iteration consisted of multiple data collections. Data collection was based on interviews. This technique had the advantage of focusing directly on the case study objectives and allowing for perceived causal relationships [3]. An interview guide was created to ensure a logical structure to the research questions. The interview questions were open-ended in order to elicit teachers' ways of working and motivations during the interviews without prescribing answers. Since the learning resource is the central element and starting point for the technical implementation of EduArc, the functionalities of the learning resource were considered as use cases for the creation of the guide. Specifically, three use cases were considered: Creating a resource, Searching a resource, and Viewing a resource.

2.2 Data Analysis

Qualitative data analysis is based on Lamnek's approach[4], which includes four phases: (1) transcription, (2) individual analysis, (3) generalizing analysis, (4) control phase. The first phase involved transcription of the interviews using the audio recording. In the second phase, irrelevant passages were deleted and relevant passages were highlighted to reduce the density of details in a single transcript. The analysis of each interview was done by highlighting all passages that are answers to the interview questions, grouping the highlighted passages into categories, reviewing the relationships between the highlighted passages, and finally comparing the results with the transcript. In EduArc, the user stories represented the categories. In the third phase, all similarities and differences between each analysis were noted. The final phase ensured that all relevant information was included. To this end, the results of the third phase were verified to be consistent with the transcripts, and errors were corrected as necessary. In addition, unclear passages were removed. Prior to conducting the interviews, a set of initial wireframes was developed to give the interviewees the idea of two alternative implementation approaches. The wireframes were used to design an initial graphical user interface. The approach was based on the user interface of the Hamburg Open Online University (HOOU) repository[1]. During the interview, both approaches were demonstrated to the participants along the use cases.

2.3 Usage Scenarios, Personas and Requirements Specification

Based on the interview responses, user requirements were defined in terms of user stories. As explained earlier, three use cases were considered: creating, searching, and viewing resources. However, when answering RQ1 and RQ2, it became clear that resource reuse was a critical issue for teachers. For this reason, another use case was added: reusing a resource. Below is a summary of the main scenarios along the use cases, along with some of the wireframes.

Create a Resource As shown in Figure 1, the teacher clicks browse and has the option to select files in different formats (1). The teacher clicks on a file, which is added to the list of selected resources (2). The file type is displayed (mp4) and the resource type (video) is determined automatically. If the resource type is not supported, the user has the option to change it before uploading by clicking the Change Document Type link[3]. If the teacher selects a folder that contains, for example, a video and a slide set, both files will be included in the list. Alternatively, the teacher can simply drag and drop files or folders. Then the teacher clicks Next and a window opens asking the teacher to enter more information (3). The teacher assigns his or her resource to a topic and determines the target audience and access. Before specifying access, the teacher reads an explanation of the relevant CC licenses. For references, the teacher searches EduArc and adds two resources as references (Figure 1 and Slides 4). Additionally, the teacher can upload a bibliography, e.g. a BibTex file. Finally, the teacher briefly describes the content of the resource and clicks Add (4). The files are saved in the EduArc hub and their metadata is added to the index.

[1] https://www.hoou.de/

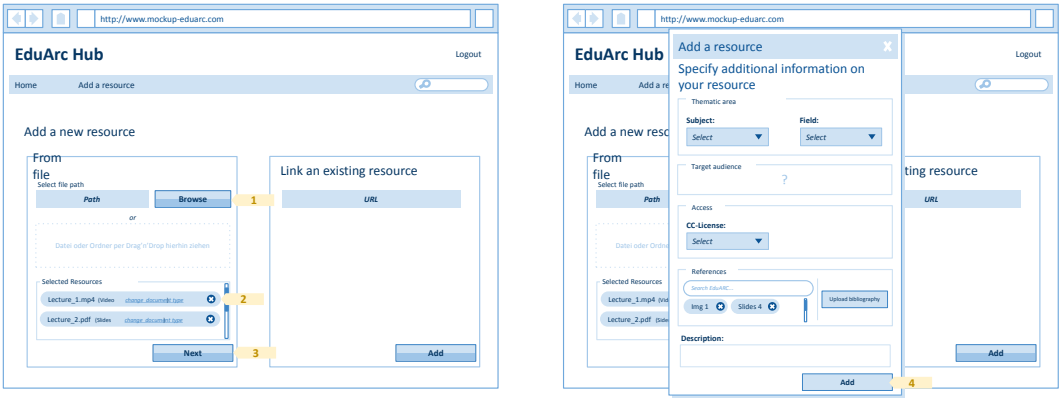

Fig. 1. Create a resource

Search a Resource The teacher has entered the search term "bitcoin" and the search view opens with two search results. The teacher tries to assess which resources are relevant and considers the given information: one resource about Block Chain from 2018 with the resource type slides and a short description; another resource about crypto currency from 2010 with the resource type video and a short description. By default, the results are sorted by relevance, but the teacher can also sort them in reverse chronological order (date). Then the teacher can narrow the search results by setting the creation date to 2015. The teacher sets the access to the license type to CCv4.

View a Resource The teacher clicks on the resource Block Chain after searching with the keyword bitcoin. The teacher can view various information such as the title, creator, topic, scope, summary, and some references and keywords.

Reusing a resource The teacher wants to save the Block Chain resource locally to use it again later.

3 Metadata for Learning Resources

A common data model for learning resources was developed based on the LOM standard. LOM was chosen because it clearly focuses on learning resources. In addition, LOM forms the basis for metadata representation in several German OER repositories for higher education, such as the Central Repository for Open Educational Resources in Baden-Württemberg (ZOERR) [2]. The common data model (CDM) made it possible to map the metadata of OERs from different repositories to this project standard and to store them in a central search index (EduArc index).

EduArc's CDM consists of 20 fields. It distinguishes between two levels of commitment:

- Mandatory properties that must be present, and
- recommended properties that can be optionally present.

The EduArc CDM incorporates the FAIR principles [5], which originated with Wilkinson [6]. The four FAIR principles describe guiding principles for discoverability, accessibility, interoperability, and reuse of digital resources.

[2] https://www.oerbw.de/

4 System Architecture

The development of the software and system architecture includes the central services located at EduArc and the existing systems and infrastructures for storing learning resources. Based on EduArc's CDM, a concept for an overall system architecture of the EduArc platform was designed to connect distributed learning resources. The overall EduArc infrastructure pipeline is shown in Figure 2.

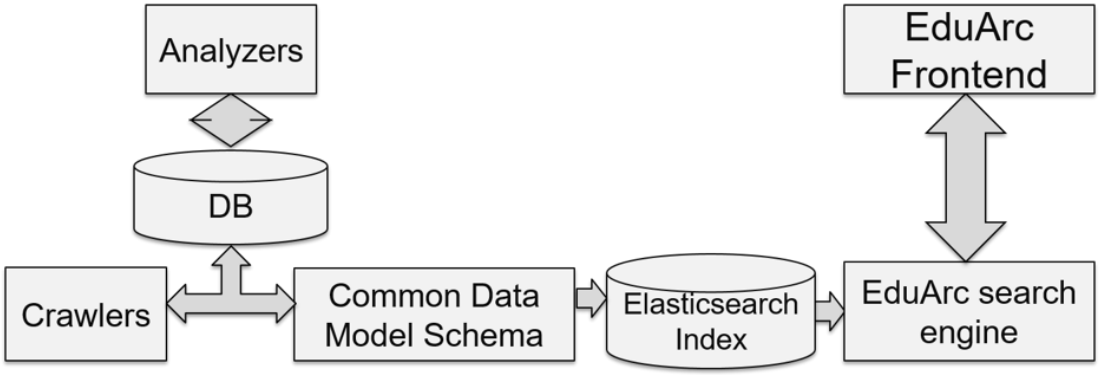

Fig. 2. The infrastructure of the EduArc project

The main components of the pipeline are:

1. Crawler: a computer program that automatically searches documents on the web [7]. Their job is to collect web content. Crawlers can search for any information, and are also called web spiders or web robots.
2. Common data model: the EduArc CDM is described in Section 3.
3. Subject Indexing Model for Metadata Generation: The subjects of the documents in the index are automatically generated using Learning-To-Rank models. Learning-To-Rank consists of a set of supervised ranking models that are trained with a numerical set of vectors to find the top-k relevant documents for a user query. The numerical set of vectors, called feature vectors, is computed based on the content of the documents and/or queries.
4. Search engine frontend: To help users of the EduArc platform find the most relevant results for their search query, a user-friendly frontend was developed, consisting of four main components: search bar, advanced search form, faceted search widgets, and search results form. The faceted search widgets allow users to filter search results based on some common data model fields such as language, author, or date ranges. In addition, users of the platform can download the metadata of the results. Figure 3 shows the search results form of the EduArc platform.

The process of collecting metadata, mapping it and storing it in the index is shown in Figure 4. The figure shows three scenarios. The first scenario is when the OER repositories use the LOM standard to model their metadata. As a result, no mapper is needed for those types of repositories. The second scenario is when the repository uses another known standard rather than LOM. Then, a mapper is required for those repositories based on their standard. The last scenario is when a repository does not use any standard. Therefore, a mapper is needed for each repository of that type.

4.1 Subject Indexing

Various subject indexing models were developed to parse the indexed documents and generate new metadata. The generated metadata contains supplementary information about the content

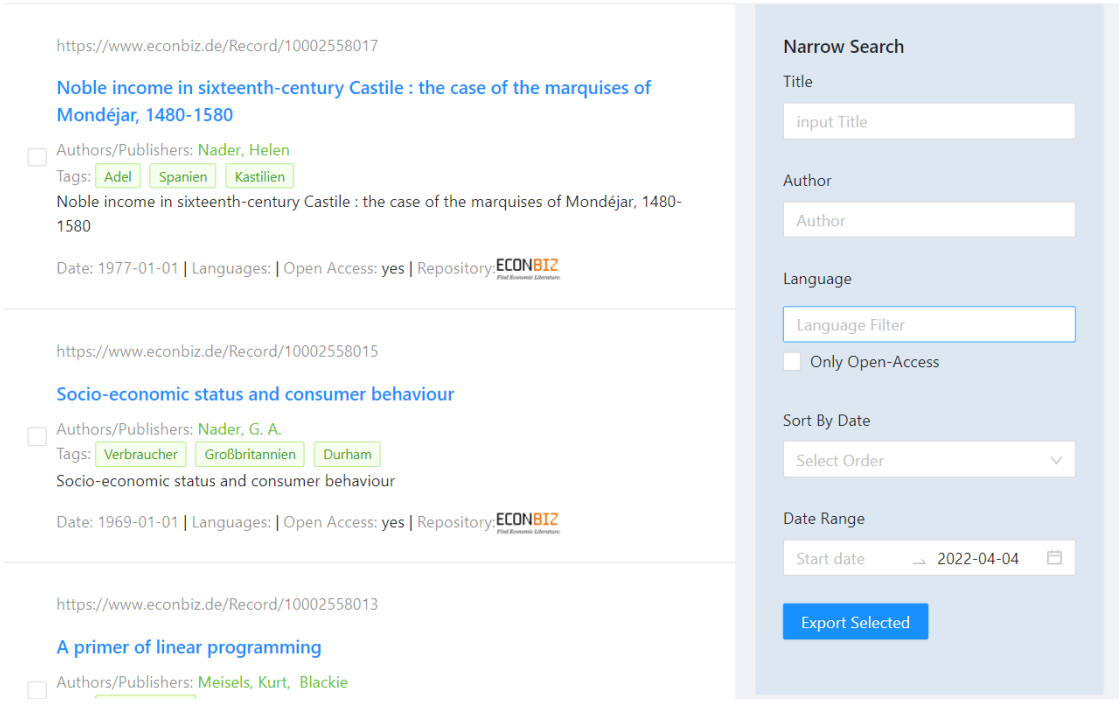

Fig. 3. The frontend of EduArc search

.

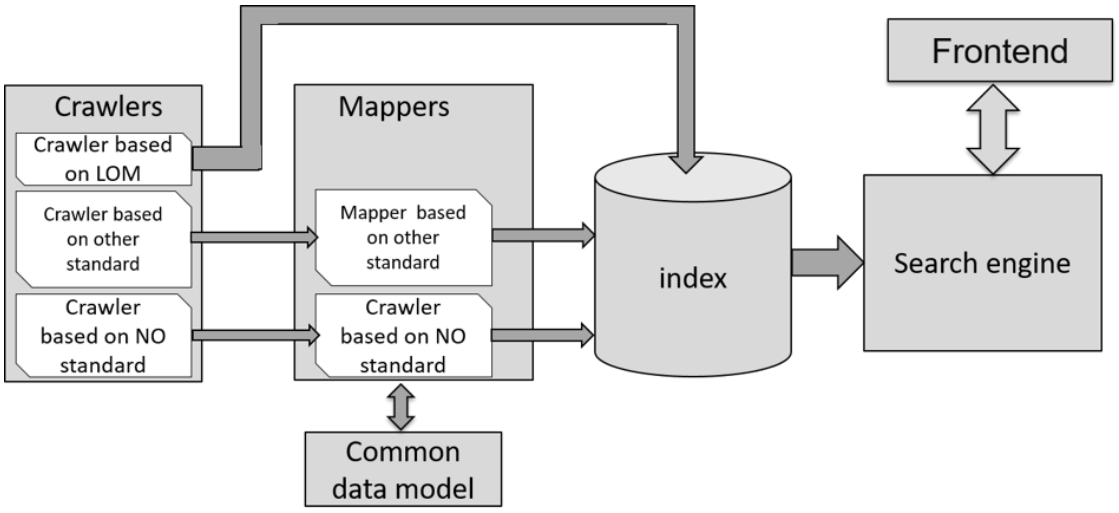

Fig. 4. The scenarios of capturing and mapping OER metadata modeled using the LOM standard, as well as other standards from capturing the metadata to storing the results and presenting them on the front end

.

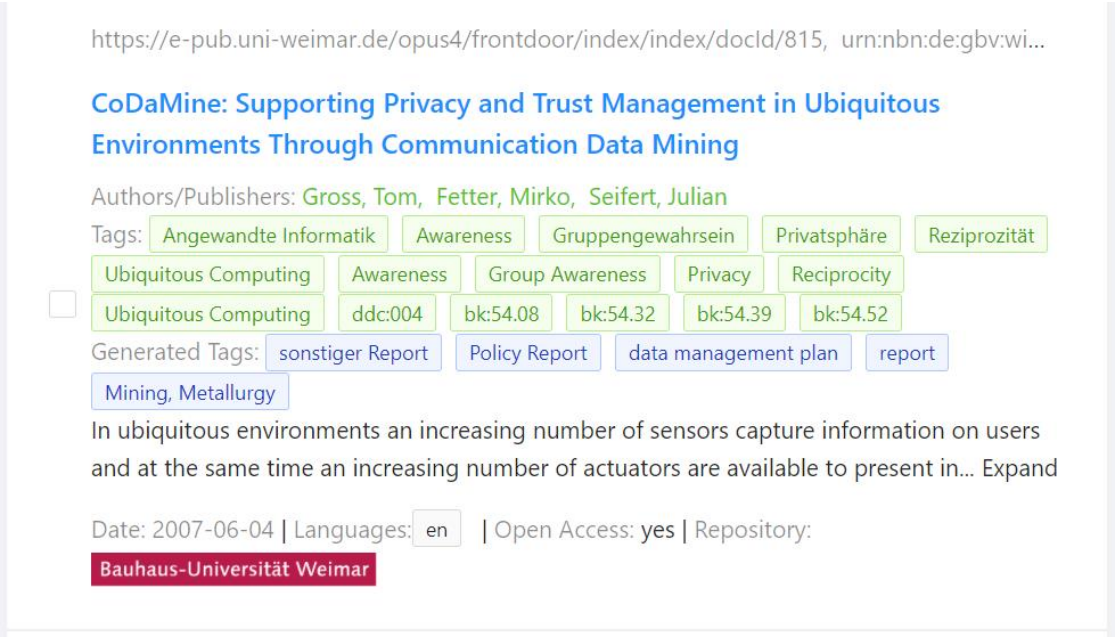

Fig. 5. Appearance of generated metadata (generated tags) on the search frontend of the EduArc platform

of a document (e.g., business, computer science, etc.). The Standard Thesaurus for Economics (STW) was used to generate the new metadata. The STW consists of nearly 6,200 keywords. In addition, the following subjects were used:

- Education Server Annotations: 53,580 German subjects.
- Confederation of Open Access Repositories (COAR): 234 German subjects and 274 English subjects.
- Skohub: 730 English and German subjects.

These thesauri were used to annotate the documents using various subject indexing models. The generated annotations for each document are added to the "generatedmetadata" field. The generated metadata is made available to data providers and other interested parties through the search frontend (Generated Tags) (Figure 5).

5 Conclusion

A federated infrastructure was developed for connecting various Open Educational Resources (OER) repositories. Six data providers were connected to the EduArc platform. The connection was made through EduArc's Common Data Model (CDM). Five of these repositories are related to educational and research institutions, and one is a source of figures. In addition, the platform automatically generates the topics of the documents in the index using learning-to-rank models. EduArc's platform includes four main services: search and advanced search with the ability to filter results, adding OER metadata to the index, storing resource metadata, and collecting OER repositories and adding the collected records to the index.

Reproducibility The EduArc platform was developed on top of Moodle. The platform's code repository has been published online[3]. A limited demo is accessable[4]. Interested universities and institutes are invited to use the platform.

[3] https://git.informatik.uni-kiel.de/asal/eduarcbox
[4] http://eduarc.zbw.uni-kiel.de

Acknowledgement This article is part of the project "Digital educational architectures: Open learning resources in distributed learning infrastructures – EduArc" funded by the German Federal Ministry of Education and Research (grant #16DHB2129).

References

1. V. I. Marín, M. Bond, O. Zawacki-Richter, C. H. Aydin, S. Bedenlier, A. Bozkurt, D. Conrad, I. Jung, Y. Kondakci, P. Prinsloo, *et al.*, "A comparative study of national infrastructures for digital (open) educational resources in higher education," *Open Praxis*, vol. 12, no. 2, pp. 241–256, 2020.
2. M. Menzel, "Lom for higher education oer repositories: Ein metadatenprofil für open educational resources im hochschulbereich zur förderung der interoperabilität von oer-länderrepositorien," *o-bib. Das offene Bibliotheksjournal/Herausgeber VDB*, vol. 7, no. 1, pp. 1–10, 2020.
3. D. E. Perry, S. E. Sim, and S. M. Easterbrook, "Case studies for software engineers," in *Proceedings. 26th International Conference on Software Engineering*, pp. 736–738, IEEE, 2004.
4. S. Lamnek and C. Krell, *Qualitative sozialforschung*. Psychologie Verlags Union München, 2005.
5. A. Saleh, I. Vagliano, T. Heck, and S. Kullmann, "Eduarc. a fair and user-centred infrastructure for learning resources," 2020.
6. M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, *et al.*, "Andra waagmeester," *Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons*, 2016.
7. C. Olston and M. Najork, *Web crawling*. Now Publishers Inc, 2010.