Data from the Digital Repository of the Faculty of Mining and Geology in eScience (eNauka)

UDC 001.103.2:004.624(497.11)
DOI 10.18485/infotheca.2023.23.2.4

ABSTRACT: The paper describes linking the Digital Repository of the University of Belgrade, Faculty of Mining and Geology, with the eScience system in terms of transferring metadata about the results of researchers' scientific work. The steps taken to ensure a smooth harvesting of metadata are outlined. Additionally, a presentation of additional improvements to the OAI system is provided, aiming to contribute to the automatic linking of authors with their results in the eScience system.

KEYWORDS: repositories, eScience, transfer of metadata, Faculty of Mining and Geology

PAPER SUBMITTED: 25 October 2023
PAPER ACCEPTED: 24 January 2024

Biljana Rujević
biljana.rujevic@rgf.bg.ac.rs
ORCID: 0000-0002-9103-3902

Mihailo Škorić
mihailo.skoric@rgf.bg.ac.rs
ORCID: 0000-0003-4811-8692

University of Belgrade
Faculty of Mining and Geology
Belgrade, Serbia

1 Introduction

The Digital Repository of the University of Belgrade, Faculty of Mining and Geology[1] (FMG Repository) exists since 2019. It serves as a depository for the results of scientific and research work conducted by the faculty's staff. Starting from 2022, an additional collection named "Završni radovi" (Final Papers) is established within the repository, where all defended bachelor's and master's theses of the faculty's students are deposited.

The repository is hosted on an open-source platform for managing digital collections called Omeka[2] (OmekaS 2023). More details about the platform and its customization can be found in the work (Popović, Škorić, and Rujević 2020). The decision to choose Omeka as the platform, rather than another

1. Digital Repository of the Faculty Mining and Geology. Accessed on October 23, 2023
2. Omeka. Accessed on October 23, 2023

Current Research Information System (CRIS) platform, is a result of the development team's experience in adapting it to the specific needs of managing mining documentation (Tomašević et al. 2018), namely, for depositing documentation related to mining projects and regulations. Also, additional tools for efficient full-text search of digital objects in the Repositoryhave been developed and implemented (Tomašević et al. 2017).

Since 2023, the Repository has been integrated with the eScience system, which represents an efficient way of organizing data about the scientific results of researchers in Serbia. This integration involves the retrieval of data about these results from relevant sources andstreamlining the organization and accessibility of scientific information within the eScience framework.

2 eScience team requirements and their implementation

The eScience system enables the retrieval of metadata about scientific results according to the international standard for the exchange of library information, which involves the implementation of the OAI-PMH protocol (Lagoze et al. 2015). Any information system, database, or internal infrastructure within any institution capable of supporting the OAI-PMH protocol has the potential for integration into the eScience system. After an initial training session for metadata editors from scientific research organizations (SRO editors[3]) from which works are going to be downloaded into the eScience system, the OAI-PMH access point of the Digital Repository of the Mining and Geological Faculty was validated. The access point was implemented using the OAI-PMH Repository module for the Omeka S system[4]. The validation was conducted by the eScience development team using their OAI Validator[5], which allows for a transparent, public, and free semantic and syntactic verification of the correctness of the metadata downloaded via the OAI-PMH protocol. After the validation, the form for technical data related to the repository was completed, establishing contact between the development teams of eScience and the Mining and Geological Faculty's repository. Soon after, the process of finding a specific solution for retrieving data on

3. The term "SRO editors" refers to editors from scientific research organizations and is being used in the documentation https://enauka.gov.rs/help/index.html. Accessed on December 23, 2023

4. Omeka-S-module-OaiPmhRepository. Accessed on October 24, 2023

5. OAI Validator. Accessed on October 23, 2023

scientific results from the Mining and Geological Faculty commenced. Given the successful recognition of the existing OAI-PMH access point of the repository, the initial solution was found relatively easily. The procedure was then streamlined to fulfill three specific requirements set by the eScience development team. These requirements were aimed at achieving optimal representation of data from the repository and ensuring their effective recognition and retrieval.

The first request pertained to the existence of a unique identifier (URI) for individual works in the metadata schema of the FMG repository. As a result, in the "DC.IDENTIFIER" field, the existing path to the record on the API (Application Programming Interface) (e.g., http://dr.rgf.bg.ac.rs/api/items/8538) was replaced with the path leading directly to the record within the Repository itself (e.g., http://dr.rgf.bg.ac.rs/s/repo/item/8538).

The second step involved creating a collection for harvesting metadata in the eScience system. Initially, the repository of the Mining and Geological Faculty was designed with each employee represented as a separate collection. Each individual collection was described by metadata including the employee's name, surname, institution, section,department, and chair of employment, various identifiers on scientific social networks and databases (ORCID, ResearcherID, ScopusID, Google Scholar, E-CRIS.SR), an identification number within the employee database at the Faculty of Mining and Geology, job title, and email address. Within each collection, all the papers of an author were included[6]. However, this structure posed challenges for stable integration with the eScience system, particularly in manual corrections required for modifications, such as when opening a profile (collection) for a new employee.After testing and internal discussions within the Faculty of Mining and Geological team, a stable solution was found. This solution involved creating a new "Radovi istraživača"[7] collection encompassing all records from the Repository containing a digital object, excluding doctoral dissertations or final assignments[8]. This approach addressed the challenges of

6. Collections of employees are presented on: https://dr.rgf.bg.ac.rs/s/repo/sets. Accessed on December 23, 2023

7. Collection "Radovi istraživača"

8. Entries without digital objects are entries retrieved from the local database FoDiB (Fondovska dokumentacija ibiblioteka), which was used to collect metadata on employees' scientific research results before the establishment of the Repository. This database served as the basis for developing the Digital Repository of the Faculty of Mining and Geology. Some of these records were later supplemented with digital objects, while around half of them still lack a digital object.

the initial structure and provided a more efficient way to manage collections in the eScience system.

For the validity of metadata, it was considered better to include only verified entries in the harvesting collection "Radovi istraživača". Verified entries in that collection refer to those where the repository administrator has reviewed the content, and these entries must contain the full text of the document (digital object). Doctoral dissertations were excluded, as they are harvested from the NaRDuS repository[9] (National Repository of Dissertations in Serbia)[10] by the eScience portal. Final assignments were also omitted as they do not represent the direct results of the research.

The collection "Radovi istraživača" is continually updated with new entries added to the Repository, ensuring that it remains a comprehensive and accurate representation of the research output from the Faculty of Mining and Geology.

The third request was to align the metadata schema used in the Repository with the schema used in the eScience system. The metadata structure in the Repository is based on the Dublin Core ontology[11] (DC 2012), with minor additions of fields from other ontologies that do not exist in this one. Dublin Core is one of the default ontologies supported by the eScience system for transferring metadata from other systems, which facilitated the process but still required specific field mappings, as described in Table 1. Additionally, the document typology appearing in the DCTERMS:TYPE field in the FMG Repository in Serbian is mapped to the English terminology used in the OAI-PMH protocol, as described in Table 2.

Table 1: Mapping fields from the metadata schema of the RGF Repository to the fields of the eScience system's metadata schema.

Mapping	
Fields from the RGF Repository	**Fields from the eScience system**
dcterms:alternative	dcterms:title

9. NaRDuS repository. Accessed on October 23, 2023

10. Doctoral dissertations defended at the University of Belgrade are processed in the COBISS system, from where metadata is transferred to the PHAIDRA repository, where they are permanently stored. Subsequently, this data is transferred to NARDUS.

11. Dublin Core ontology. Accessed on December 23, 2023

Continuation of the table 1	
Fields from the RGF Repository	**Fields from the eScience system**
rgf:typeVersion	dcterms:type
dcterms:spatial	dcterms:subject
rgf:subjectUdc	dcterms:subject
rgf:identifierCategory	dcterms:description
rgf:identifierSubcategory	dcterms:description
dcterms:tableOfContents	dcterms:description
dcterms:abstract	dcterms:description
rgf:contributorAdvisor	dcterms:contributor
rgf:contributorOther	dcterms:contributor
dcterms:created	dcterms:date
dcterms:valid	dcterms:date
dcterms:available	dcterms:date
dcterms:issued	dcterms:date
dcterms:modified	dcterms:date
dcterms:dateAccepted	dcterms:date
dcterms:dateCopyrighted	dcterms:date
dcterms:dateSubmitted	dcterms:date
dcterms:extent	dcterms:format
dcterms:medium	dcterms:format
dcterms:bibliographicCitation	dcterms:relation
rgf:identifierCobissid	dcterms:identifier
dcterms:isVersionOf	dcterms:relation
dcterms:hasVersion	dcterms:relation
dcterms:isReplacedBy	dcterms:relation
dcterms:replaces	dcterms:relation
dcterms:isRequiredBy	dcterms:relation
dcterms:requires	dcterms:relation
dcterms:isPartOf	dcterms:relation
dcterms:hasPart	dcterms:relation
dcterms:isReferencedBy	dcterms:relation
dcterms:references	dcterms:relation
dcterms:isFormatOf	dcterms:relation
dcterms:hasFormat	dcterms:relation
dcterms:conformsTo	dcterms:relation
dcterms:accessRights	dcterms:rights

Continuation of the table 1	
Fields from the RGF Repository	**Fields from the eScience system**
dcterms:license	dcterms:rights
dcterms:temporal	dcterms:coverage

Table 2. Mapping the values of the field DCTERMS:TYPE in the FMG Repository to those of the DCTERMS:TYPE in OAI-PMH.[13]

Value of the DCTERMS:TYPE in the FMG	**Value of the DCTERMS:TYPE in OAI-PMH**
Докторска дисертација	doctoralThesis
Саопштење са скупа штампано у извод Рад у зборнику	conferenceObject
Дипломски рад	bachelorThesis
Магистарска теза Мастер рад	masterThesis
Рад у часопису	journalArticle
Поглавље у монографији	bookPart
Књига Монографија Практикум Скрипта	book

The mapping shown in Table 1 has been implemented using the mentioned OAI-PMH Repository module. In addition to what is presented, the default mapping of concepts and properties from the *bibo* ontology[14] (D'Arcus and Giasson 2016) to the *dcterms* ontology has also been applied.

14. BibliographicOntology – *bibo* – The Bibliographic Ontology, bibo, is a bibliographic ontology developed for the Semantic Web to describe bibliographic resources such as books, journals, and similar materials. It is based on the RDF (Resource Description Framework) data model. It includes concepts and properties that facilitate the description of bibliographic sources, bibliographic citations, document classifications, or the description of any document in RDF. Available

The concepts and properties of the *bibo* ontology are used in the metadata schema of the Repository to describe the following metadata: the number of pages (BIBO:PAGESTART, BIBO:PAGEEND), volume and issue numbers of journals (BIBO:VOLUME, BIBO:ISSUE), digital object identifier (BIBO:DOI), ISSN of the journal (BIBO:ISSN), and URI identifier (BIBO:URI). Figure 1 illustrates an example metadata record in the FMG Repository and the structure of the metadata schema. An example of the same record in the OAI-PMH access point is available at https://dr.rgf.bg.ac.rs/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:dr.rgf.bg.ac.rs:5011. An alternative solution has been found because the OAI-PMH Repository module does not support document typology mapping. Another separate application has been deployed alongside the Repository on the server, acting as an adapter for the results sent by the original OAI-PMH access point. The solution consists of three steps (Figure 2):

1. The OAI-PMH Adapter application receives a query in the format expected by the OAI-PMH access point;
2. The OAI-PMH Adapter application forwards the unchanged query to the OAI-PMH access point of the FMG Repository (Appendix 3);
3. The OAI-PMH access point sends a response to the adapter application in the form of an XML string;
4. The adapter application further processes the received XML string, mapping document typology values according to the data shown in Table 2. Once the adapted response is ready, it is sent to the eScience portal to respond to the issued query.

The developed OAI-PMH Adapter application is open-source and available on GitHub[15]. It can be downloaded and customized for the needs of other repositories.

After creating a collection for downloading and processing data sent via the OAI-PMH protocol, the established linking works almost perfectly. When a work is verified in the Repository, it becomes part of the "Radovi istraživača" collection and is pulled into the eScience system at a specific time. An example displaying the record from the FMG Repository on the eScience portal is provided on Figure 3.

The only issue we have identified is the lack of information transfer regarding the number of pages for publications without Persistent Identifiers

at: https://www.dublincore.org/specifications/bibo/bibo/ (accessed January 22, 2024.

15. OAI-PMH Adapter. Accessed on October 25, 2023.

```
<dcterms:type> Рад у часопису</dcterms:type>
<rgf:typeVersion>објављена верзија</rgf:typeVersion>
<dcterms:language>енглески</dcterms:language>
<dcterms:creator  xml:lang="sr">Petar Popović, Mihailo Škorić, Biljana Rujević</dcterms
    :creator>
<dcterms:title xml:lang="en">The Use of the Omeka Semantic Platform for the Development of
    the University of Belgrade, Faculty of Mining and Geology Digital Repository</dcterms
    :title>
<dcterms:source xml:lang="en">Infotheca</dcterms:source>
<dcterms:publisher xml:lang="en">Faculty of Philology, University of Belgrade</dcterms
    :publisher>
<dcterms:issued>2021</dcterms:issued>
<dcterms:abstract xml:lang="en"> Under the regulations of the Ministry of Education,
    Science and technological Development, a digital repository based on the Omeka S data
    storage platform has been developed for the Faculty of Mining and Geology. The platform
    has been upgraded with the required modular extensions, Solr index and automatic OCR.
    Furthermore, document indexing and search have been fine-tuned with the aid of e
    -dictionaries of the Serbian language, which has brought about outstanding results in
    terms of usage facilitation and overall speed of document storage and search within the
    repository that is a part of the application. </dcterms:abstract>
<bibo:volume>20</bibo:volume>
<bibo:issue>1-2</bibo:issue>
<bibo:pagestart>136</bibo:pagestart>
<bibo:pageend>148</bibo:pageend>
<bibo:doi>10.18485/infotheca.2020.20.1_2.9</bibo:doi>
<bibo:issn>1450-9687</bibo:issn>
<dcterms:subject xml:lang="sr">Дигитални репозиторијум, Омека, претрега дигиталних
    библиотека</dcterms:subject>
<dcterms:subject xml:lang="en">Digital Repository, Omeka, Digital Library searching
    </dcterms:subject>
<bibo:uri href="http://dx.doi.org/10.18485/infotheca.2020.20.1_2.9">http://dx.doi.org/10
    .18485/infotheca.2020.20.1_2.9</bibo:uri>
<rgf:identifierCategory>M50</rgf:identifierCategory>
<rgf:identifierSubcategory>M53</rgf:identifierSubcategory>
<dcterms:format>.pdf</dcterms:format>
<dcterms:rights>Отворени приступ</dcterms:rights>
<dcterms:license>All rights reserved</dcterms:license>
```

Figure 1. The structure of metadata schema and values of an original entry in the FMG Repository

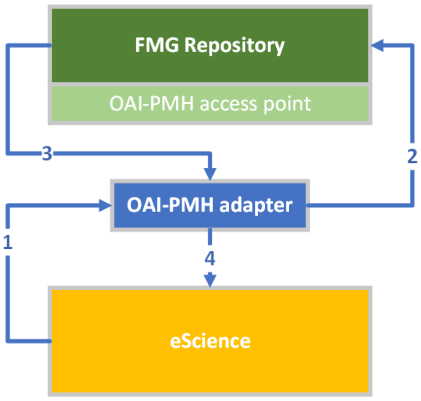

Figure 2. Structure of the OAI-PMH Adapter Application

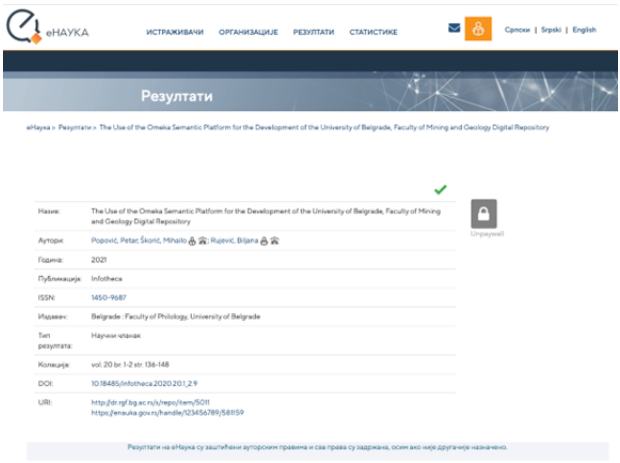

Figure 3. Example on Figure 1 in the eScience system

(PIDs). When a publication has a PID, the eScience system retrieves information about the number of pages from an appropriate external source.[16] In practice, this issue is most common with conference papers because they often lack PIDs, making it impossible to add page number information from other sources without editorial intervention.

3 Improvement of Metadata Structure

Over time, it was identified that metadata could be enhanced. Therefore, using the OAI-PMH Adapter application, additional processing of the following results was applied:

1. Individual data from the fields for listing authors (DC:CREATOR), keywords (DC:SUBJECT), and spatial coverage (DC:SPATIAL), which were originally entered in the Repository in the same field separated by

16. To enhance the system's interoperability, data related to Scientific Research Organizations (NIO), researchers, and results have been enriched with Persistent Identifiers (PIDs), such as ORCID, E-CRIS ID, APVNT, Researcher ID, COBISS, SR ID, DOI, Scopus ID, etc.

commas, were separated into distinct fields. For example, all authors initially added to the DC:CONTRIBUTOR field in the Repository as "PETAR POPOVIĆ, MIHAILO ŠKORIĆ, BILJANA RUJEVIĆ" (as shown in Figure 1). As a result of using the OAI-PMH Adapter application on the OAI-PMH access point were separated into distinct repeated fields DC:CREATOR with values "PETAR POPOVIĆ," "MIHAILO ŠKORIĆ," and "BILJANA RUJEVIĆ" (example on the OAI-PMH access pointhttps://dr.rgf.bg.ac.rs/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:dr.rgf.bg.ac.rs:5011).

2. For authors added to entries as resources from the Omeka system, ORCID identifiers were imported if they exist. In the example provided in https://dr.rgf.bg.ac.rs/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:dr.rgf.bg.ac.rs:5011, the DC:CREATOR field includes the HREF attribute with the value being the path to the author's resource in the Repository from where author's (researcher's) ORCID ID is retrieved if exists. As a result of the OAI-PMH Adapter application's work, in the DC:CREATOR field instead of HREF attribute is added ID attribute with the value being the ORCID ID of the given author (researcher) which can bee seen in example https://dr.rgf.bg.ac.rs/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:dr.rgf.bg.ac.rs:5011. This way, the entry harvested from the Repository is automatically attributed to the profile of a specific researcher in the eScience system. In cases where authors are not added as separate resources, the DC:CREATOR field is split using commas to ensure that each author gets their own field, as illustrated in the example in the preceding paragraph (Figure 1 and example https://dr.rgf.bg.ac.rs/oai?verb=GetRecord&metadataPrefix=oai_dc&identifier=oai:dr.rgf.bg.ac.rs:5011).

4 Conclusion

Six months after the launch[17] of the eScience portal and the establishment of automatic metadata transfer from the FMG Repository, it is evident that this served as an additional encouragement for Faculty members to add their papers to the Repository. There has been an increase in the number of

17. Portal eScience was officially launched on July 3, 2023. Faculty members objectively had less time to use the eScience portal as its launch coincided with the period of collective annual leave in higher education institutions.

new entries, likely because employees realized that depositing papers in the Repository is beneficial. During 2021, 412 entries were deposited; in 2022, 437 entries; and in 2023, when the eScience system was active, the number increased to 655 entries. This number is anticipated to continue to grow, considering that eScience is currently the most systematic way of collecting and presenting information about scientific production in Serbia. It is believed that the system will become even more valuable and functional in the following stages of development.

In the future, efforts will be focused on refining metadata in the Repository to ensure that values of all metadata fields are transferred to the eScience system properly (e.g., page numbers). Simultaneously, communication with the eScience development team will continue, to ensure that all system improvements are implemented in a timely manner.

References

D'Arcus, Bruce, and Frédérick Giasson. 2016. *Bibliographic Ontology (BIBO)*. Preuzeto 22. 01. 2024, https://www.dublincore.org/specifications/bibo/bibo/.

DC. 2012. *Dublin Core™ Metadata Element Set, Version 1.1: Reference Description.* Preuzeto 23. 10. 2023, https://www.dublincore.org/specifications/dublin-core/dces/.

Lagoze, Carl, Herbert Van de Sompel, Michael Nelson, and Simeon Warner. 2015. *The Open Archives Initiative Protocol for Metadata Harvesting.* Preuzeto 24. 10. 2023, http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm.

OmekaS. 2023. *Omeka S User Manual.* Preuzeto 23. 10. 2023, https://www.dublincore.org/specifications/dublin-core/dces/.

Popović, Petar, Mihailo Škorić, and Biljana Rujević. 2020. "The Use of the Omeka Semantic Platform for the Development of the University of Belgrade, Faculty of Mining and Geology Digital Repository." *Infotheca* 23 (1-2): 136–148. https://doi.org/https://doi.org/10.18485/infotheca.2020.20.1_2.9.

Tomašević, Aleksandra, Biljana Lazić, Dalibor Vorkapić, Mihailo Škorić, and Ljiljana Kolonja. 2017. "The Use of the Omeka Platform for Digital Libraries in the Field of Mining." *Infotheca* 17 (2): 26–49. https://doi.org/https://doi.org/10.18485/infotheca.2017.17.2.2.

Tomašević, Aleksandra, Ranka Stanković, Miloš Utvić, Ivan Obradović, and Božo Kolonja. 2018. "Managing mining project documentation using human language technology." *The Electronic Library* 36 (6): 993–1009. https://doi.org/https://doi.org/10.1108/EL-11-2017-0239.