

The use of the Omeka Semantic platform for the development of the University of Belgrade, Faculty of Mining and Geology digital Repository

UDC 027.7(497.11):004.738.5

DOI 10.18485/infotheca.2020.20.1_2.9

ABSTRACT: 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.

KEYWORDS: Digital Repository, Omeka, Digital Library searching.

PAPER SUBMITTED: 16 October 2020

PAPER ACCEPTED: 9 November 2020

Petar Popović

petar.popovic@rgf.bg.ac.rs

Mihailo Škorić

mihailo.skoric@rgf.bg.ac.rs

Biljana Rujević

biljana.rujevic@rgf.bg.ac.rs

*University of Belgrade,
Faculty of Mining and Geology
Belgrade, Serbia*

1 Introduction

In 2018, the Ministry of Education, Science and Technological Development of the Republic of Serbia prescribed the mandatory deposit of metadata of scientific research publications in the appropriate national, thematic or institutional repository for all institutions engaged in scientific research.¹ In accordance with that, in June of 2019, the Faculty of Mining and Geology of the University of Belgrade started developing a software solution for the repository. The platform was selected according to the Compliance with the

¹ Platforma za otvorenu nauku

Open Archives Initiative standards and the principles of FAIR metadata dissemination.

FAIR (*Findable, Accessible, Interoperable, Reusable*) principles ensure that metadata is visible, accessible, interoperable and durable. The basic FAIR principle is that data must have at least one unique identifier based on which it can be found in the repository in which it is stored or in some other repository to which it is linked. Such a unique identifier should thus enable the retrieval of data according to a standardized protocol. Data interoperability is achieved by referring to other data (metadata) and is measured by the performance of the repository in various data exchange methods, both with users and with other software.

In addition to many open source software solutions such as DSpace with Cris add-on, e-Prints, Omeka and others, we opted for the Omeka² platform due to previous positive experiences, ease of installation, flexibility and a large number of software add-ons to extend basic functions, which are also open access. A large number of institutions use the Omeka platform for various applications of digital archiving of diverse content, which also contributes to the credibility of the platform³ and a large user base is a prerequisite for an active user community gathered at the forum,⁴ where users can consult with colleagues from other institutions, add-on authors or the development team of the Omeka project itself. Creating a multi-purpose platform is exactly one of the goals we wanted to fulfill. It implies the creation of a repository of not only scientific papers published in foreign journals, but also professional and scientific papers in the Serbian language. An important functionality is the possibility of mutual thematic and spatial connection of research results related to a locality, which is very important in the field of mining and geology. The development of the repository and its software add-ons, as a rule, follows the principles of Agile Project Management and design that takes into account the needs of users (User-Centered Design) (Perry and Waggoner, 2018). Focusing on users, inclusive and iterative development are key components of the repository development process, which means close cooperation of the development team with fellow librarians and other faculty employees from the Mining and Geology Department whose feedback on software needs and testing contributed to the development process.

² Omeka

³ Omeka Directory

⁴ Omeka Forum

Examples of previous positive experiences of using the Omeka platform are ROmeka@RGF⁵ (Томашевић et al., 2017) and the digital library of the Criminal Police Academy⁶ (Vorkapić et al., 2017). Romeka@RGF served as an open source development platform on which query extension software and morphological dictionaries were successfully applied when searching the digital library, which showed how important these language resources are for the relevance of the obtained results, ie for increasing system response (Tomašević et al., 2018). A special OWL⁷ RuDokOnto ontology has also been developed for this digital library in order to better show users the relationship between the elements of the project documentation for which the library is intended to be stored (Лазивић и др., 2019). Another advantage of the Omeka platform is the number of users and their activity in further software development. Omeka Semantic,⁸ an improved version of the Omeka Classic platform, was used to create the repository. A detailed list of all add-ons for the basic version of Omeka Semantic can be found on the GitHub page of the project author.⁹ Omeka Semantic is intended for institutions that manage multiple sites that share resources with each other. Also, Omeka Classic uses only Dublin Core for metadata, while Omeka S (Semantic) supports multiple namespaces, linking to ontologies and creating dictionaries, and is also compliant with Linked Open Data principles. An example of using the Semantic version of the platform is Henri Poincaré’s Corpus of Digitized Correspondence with Famous Scientists, published using the Omeka S development environment, combining RDF (Rich Document Format) repositories and SPARQL¹⁰ queries (Bruneau et al., 2020). Numerous long-term solutions of digital archives based on the Omeka platform have indicated that the platform can be successfully used at the Faculty of Mining and Geology not only as a specialized digital library but also as a repository for scientific papers and various forms of digital resources. Omeka is also used as a teaching tool for independent research projects of students (Cuenca and Kowaleski, 2018). As a platform for different types of digital archives, it is widely used in the digital humanities, not only for written digital objects, but also for multimedia content (Gill et al., 2020).

⁵ ROmeka@RGF

⁶ Digital Library of the Criminal Police Academy

⁷ OWL RuDokOnto Ontology

⁸ Omeka Semantic

⁹ Omeke Semantic - GitHub

¹⁰ SPARQL

The repository of the Faculty of Mining and Geology¹¹ was built as a standalone application on a server within the institution. It can be accessed by all users via the Internet, and can be used via a standard user interface through a web browser and through an API (*Application Profile Interface*). The OAI-PMH¹² (*Open Archives Initiative - Protocol for Metadata Harvesting*) has also been established, so that metadata from the entire repository is available this way as well. The API is particularly important because it allows for further upgrades over the system, and the OAI-PMH protocol is one of the more well-known ways of automated data exchange between a metadata aggregation system and a repository. The metadata of all papers are publicly available, and the availability of the full text of the papers is defined by the application of various licenses in accordance with the authors of the papers.

2 Technical aspects of the repository

The operating system on which the repository was built is *Ubuntu*¹³ distribution of the Linux operating system, version 19.04. For the needs of the Omeka platform, a 5.7 version of the MySQL¹⁴ server has been installed, and *Apache Solr*¹⁵ is also used to search full text.

The current repository¹⁶ contains a total of 3874 records of 10 different types of papers: in-proceedings (1333), journal articles (1052), doctoral dissertations (688), master's thesis (621), chapter in the monograph (113), conference prints (20), specialist paper (18), monograph (1), practicum (1) and habilitation paper (1). All papers for which the original document is available, in addition to containing metadata, the text is available and searchable over the full text, and this functionality is enabled by using the SOLR index whose integration is available as a plug-in (*plug-in*) for Omeka platform.

Additional Omeka functionalities are provided by software add-ons, and the modules used are:

1. *Advanced Search Plus module* - Module for additional fields that are searchable via advanced search (search for records by creation date, availability and resource type, etc.);

¹¹ The repository of the Faculty of Mining and Geology

¹² OAI-PMH protocol

¹³ Ubuntu

¹⁴ MySQL

¹⁵ Apache Solr

¹⁶ Given data on the 15.10.2020

2. *Bulk export* – A module that allows mass export of metadata in various formats;
3. *Citation* – Module for displaying a string for quoting papers in bibliographic sources;
4. *Clean Url* – A module that displays URLs as simple links suitable for SEO (Search Engine Optimization);
5. *CSS editor* – A module that adds options for entering external CSS styles on the administration page;
6. *Custom Ontology* – A module that allows the creation of custom ontologies to describe resources that cannot be described by standard ontologies;
7. *Custom Vocab* – A module that allows you to create your own dictionaries for describing objects, which allows us to create drop-down lists when entering metadata and provide uniform values of certain fields with metadata;
8. *Derivative Images* – Module for automatic creation of images for display of papers – e.g. automatic generation of a small image of the home page of the paper for which a source in pdf format is available;
9. *Group* – Module for grouping users and resources in order to determine access rights and visibility;
10. *Collecting* – Module for adding forms through which users who are not repository administrators can add records;
11. *Guest User* – Module for creating user profiles for users who are not administrators;
12. *OAI-PMH Repository* – A module that makes the repository accessible via the OAI-PMH protocol for data exchange;
13. *Search* - Module for advanced search parameters;
14. *Search adapter for Solr* - Integration module with Solr index that allows searching papers in full text;
15. *Auto Suggest* – A module that allows automatic display of metadata suggestions for object description.

The whole system consists of an application written in the PHP programming language, a MySQL database for metadata and a Solr index for storing full-text documents. In addition to all the functionalities offered by additional modules, the repository has the ability to automatically read resources (scanned documents) that have not been previously machine-readable and save their text, which allows searching the entire text of documents, and which is especially suitable for stored pdf files created by scanning. Google’s

open source program *Tesseract*¹⁷ is used for automatic optical character recognition (in the background) or OCR (*Optical Character Recognition*). The search has been improved and enhanced with the possibilities of smart search for all inflected forms of words, so that the requested term will be found even though it does not appear in the form specified in the query, which will be discussed in more detail later in this paper.

3 Entering data into the repository

There are two types of entries in the user interface - bulk and standalone. The repository supports the principle of self-storage, which means that authors of papers can store their own papers after logging into the system. The structure of user accounts firmly follows the data structure of other parts of the information system of the Faculty of Mining and Geology, so that for each user has a certain department and chair where it belongs, which enables more detailed filtering and analysis of data in the system. Each user receives a username and password from the administrator, who assigns them access rights and authorizations to paper in the repository. There are three categories of users:

1. The author – researcher – can add his papers, and in accordance with the authorizations can see other papers in the repository;
2. Librarian – manages all records in the repository and controls the entry and import of data by authors, exports data from the repository;
3. Administrator – performs complete site administration, manages user accounts, metadata and pages.

The user interface is adapted to the employees at the faculty, so they can easily archive their papers on their own. The simplest example of entering a paper in the repository looks like this: after the author logs in to the system, he chooses the type of his paper from the drop-down menu and fills in the form with the required information (metadata about the paper) and uploads a file if necessary (Figure 1), and then the librarian, after the control, accepts it, and if necessary supplements it with metadata and verifies it so that it becomes visible in the repository. Selecting a paper type from the drop-down menu generates a form specifically designed for the required paper type, which makes it easier for users to enter data. Another way to make it easier for users to enter data and to control the quality of the data entered

¹⁷ [Tesseract](#)

are drop-down lists with predefined values designed for each form separately, where feasible. Mandatory fields when entering data are marked with a “*” and the system certainly does not allow the storage of incorrectly entered record. Also, next to each input field there is an example of correctly entered data. A total of 270 user accounts were created and several workshops were organized where members of the development team trained users for self-archiving of papers.

The screenshot shows a web form for adding a work to a journal. The form is titled "Рад у часопису" (Work in journal). It includes several input fields with labels and examples of correct data entry. The fields are: "Врста публикације" (Publication type) with a dropdown menu showing "Рад у часопису"; "Верзија рада" (Work version) with a dropdown menu showing "објављена верзија"; "Језик рада" (Work language) with a dropdown menu showing "енглески"; "Аутор(и)" (Author(s)) with a text input field containing "Aleksandra Tomašević, Blijana Lazić, Dalibor Vorkapić, Mihailo Škorić, Lijljana Kolonja"; "Наслов рада (Наслов - поднаслов)" (Work title (Title - subtitle)) with a text input field containing "The Use of the Omeka Platform for Digital Libraries in the Field of Mining"; "Наслов часописа" (Journal title) with a text input field containing "Infotheca"; "Издавач (Београд : Просвета)" (Publisher (Belgrade : Prosвета)) with a text input field containing "Faculty of Philology, University of Belgrade"; "Година издавања" (Year of publication) with a text input field containing "2018"; and "Сажетак на српском језику" (Summary in Serbian language) with a text input field. At the top of the form, there is a search bar with the text "10.18485/infotheca.2017.17.2.2" and two buttons: "Добави податке" (Add data) and "Провери категорију (ПроРеФ)" (Check category). The form is part of a larger interface with a green header and a navigation menu.

Figure 1. Form for posting a paper published in a journal

Significant acceleration of data entry was achieved by integrating meta-data retrieval tools from the DOI (*Digital Object Identifier*)¹⁸ database. If the paper has been previously assigned a DOI identifier, the user can press the button to “download data from the DOI database” and automatically fill in the form with the downloaded data. Data entry can also be performed through a mass data entry add-in, where the data must be saved in a single CSV (*Comma Separated Value*) file in which the input field data is separated by commas. Users of the *librarian* and *administrator* categories have the privilege of such data entry, in order to avoid misuse or incorrect data entry.

¹⁸ Digital Object Identifier

4 Data display

The home page, also the repository search page (Figure 2) consists of three parts:

1. Main menu bar – which contains links to an advanced search page, a page for posting a new paper (for authorized users only), and a page for viewing a list of authors and their works. Next to the field for entering search terms, there is a drop-down menu from which the category to be searched is selected (one specific metadata category or full text);
2. Left menu for filtering search results and sorting found results;
3. Search results display – consists of the highlighted title of the paper, the author, the button for generating a string to cite the paper, the abstract and the generated image of the title page of the paper. During the detailed presentation of the work (which is obtained by clicking on its title), you can see the display of all metadata and resources, if they are set. As a result of the search of the text of the document, instead of the abstract, excerpts of the text (*snippets*) are displayed in which the found terms are emphasized.

For each resource in PDF format added to the repository (after being verified by the librarian), a new title page is automatically generated with the repository logo, basic metadata (author(s), title, year of publication, etc.) and a citation string, which becomes first page of the document.

In addition to the user interface for those searching the repository, Omeka's admin panel is also used, a special interface through which the repository is administered. These pages are available in the user categories *librarian* and *administrator*, and the ease of use of the admin panel is one of the main advantages of the Omeka platform.

Changes in the source code of the application and the use of language resources have significantly improved the system in two directions:

1. Search assisted by electronic dictionaries – improvement of Solr system;
2. Snippets presentation – source code enhancement and several add-ons for Omeka S.

Based on the Serbian electronic dictionaries (Krstev, 2008), lists of related terms were first generated, which were then imported into the Solr environment on the repository server. This type of connection enables a smarter search of the repository and obtaining better quality results. Searching a

Др РГФ - Репозиторијум РГФ Радови Пријављени као Petar Popović Омега

Др РГФ - Репозиторијум РГФ: Радови Претрага Напредна претрага Постави рад Аутори

Претрага радова Сва поља

Релевантност 1 од 194 1–20 од 3876 извоз резултата

Филтер

Тип

- Рад у зборнику (1333)
- Рад у часопису (1054)
- Докtorsка дисертација (688)
- Магистарска теза (621)
- Поглавље у монографији (113)
- Саопштење са скупа штампано у изводу (20)
- Специјалистички рад (18)
- Монографија (1)
- Практикум (1)
- Хабилитациони рад (1)

Медиј

- pdf (372)
- doc (2)
- htm (1)

Одсек

- Геолошки одсек (1151)
- Рударски одсек (743)
- Календар и кабинетни општих предмета (271)
- Административно-техничка служба (20)

Департаман

- Департаман за минералогiju, кристалографију, петрологију и геохемију (594)
- Департаман за хидрогеологију (359)
- Департаман за геотехнику (261)
- Департаман за регионалну геологију (214)
- Департаман за економску геологију (176)
- Департаман за палеонтологију (83)
- Департаман за геобизнис (79)

Еоенска фауна југозападног дела Сирт басена (Либија) Eocene fauna of the southwestern part of Sirt basin (Libya)

Лукић Рундић, Мери Ганић (2020) [цитирај извор](#)

Палеогене формације заузимају велико распрострањење на простору Либије. Осим стратиграфског значаја, важне су у истраживању нафте и гаса и генетски су везане за домен некадашњег океана Тетиса. Током вишегодишњих теренских истраживања југозападног дела Сирт басена (N 25°00' - 28°00'; E18°00' - 19°30') прикупљени су бројни фосили из Wadi Thamus формације која одговара старјени средњем (Јурајем-Бартонјем) и горњем еоцену (пријабонјем). У асоцијацији фауне доминирају остаци мезушњак (гастроподи, бивалвије, цефалоподи и др.) и крупне формицифере. Разноврсност фосилне фауне зависи преважно од ...

Анализа ефикасности рада роторних багера SRS2000.32/5-VR агањаваних на површним копovima Elektroprivrede Србије

Филп Милетић, Марко Лaziћ, Предраг Јованчић, Стеван Денадић, Драган Љгњатовић (2020) [цитирај извор](#)

Велика капитална улагања при куповини основне опреме за императору имају што веће историфиђење. Један од показатеља ефикасности рада великих система на нашим површним копovima је остварена производња. Поред нје, изузетно битан сегмент представљају и кварови који изазивају веома велике трошкове услед стајања система и њиховог одклањања. У овом раду извршена је упоредна анализа два роторна багера типа 2000 и то са аспекта ефикавног времена рада, остварене производње и застоја система са циљем да се укаже на утицај појединачних фактора ...

Determination of the groundwater-leakage mechanism (binary mixing) in a karstic dam site using thermometry and isotope approach (HPP Visegrad, Bosnia, and Herzegovina)

Љијана Васић, Сања Милановић, Анита Пускаћ-Презнер, Лазло Палцу (2020) [цитирај извор](#)

This paper presents a study of karst conduit genesis as well as establishing a groundwater circulation model: an isotopic approach for revealing binary mixing of cold and the thermal water underlying hydropower plant (HPP) at Visegrad, Bosnia, and Herzegovina. The study was conducted with the aim of water-leakage mechanism in the porous karst. Before the first filling of the HPP Visegrad reservoir, water from various directions of the dam had a temperature

Figure 2. Home page

word automatically takes into account all its inflectional forms (tenses, number, cases, etc.), synonyms (if given in the database), as well as transliterated pairs of those words (it does not matter whether the search is performed in Cyrillic or Latin). Once imported, the lists also enable smart indexing in the Solr system, which reduces the namespace, index size and thus greatly speeds up the repository search process, although it takes a few extra seconds during the initial indexing of papers. All this together ensures that the user gets to the requested information quickly, avoiding multiple repository searches on any particular query.

Changes to the source code of appendices, among other things, include displaying marked parts of the text in which the requested terms were found (Image 3), functionality of reading scanned documents, automatic filling of input forms from the DOI database, etc. As a complementary software, when implementing background processes, the Redis¹⁹ server – Server for managing multiple background processes is also used. This ensures the smooth operation of the Omeka application while in the “background” the system solves more hardware-intensive tasks such as OCR of attached documents.

Figure 3 shows an example of search results for the keywords “underground water source.” As we can see in the passages of the text, the inflectional forms of the words “source,” “underground” and “water” are also emphasized. The search parameters define that all metadata fields are searched, including the entire text of the document, so the most relevant search result is considered to be the document in which the variant forms of the requested terms appear together several times. Among the search parameters, we can choose the search only by metadata, by title, keywords, abstract, author, year of publication, mentor. These fields were considered as the ones that users could most often search for, so they were set aside for quick search on the home page. Other fields can be searched via the advanced search page, where advanced users are enabled to create more complex queries via a graphical interface. The MySQL database is used for search by metadata and the Solr index is used for full-text search. The results of the obtained search can also be reported as a text file in the form of a list of references to papers. It is an efficient way to easily print a bibliography, and in combination with filtering the results with the help of facets, it becomes a particularly useful tool for different needs.

Also, a very important functionality when searching the repository is the ability to filter the obtained results, using facets on the left side as shown in Figure 3. It is possible to narrow the search by several parameters such

¹⁹ [Server for managing multiple background processes – Redis](#)

The screenshot shows a search results page from the RPF repository. The interface includes a search bar at the top with the query 'izviske podzemnih voda' and a search button. Below the search bar, there are navigation tabs: 'Pretraga', 'Napredna pretraga', 'Postavi rad', and 'Autori'. A filter for 'Relevantnost' is set to '1 od 28' results, with a total of '1-20 od 555' items. A sidebar on the left contains various filters categorized by 'Tip', 'Majst', 'Odelj', 'Departman', and 'Kategorija'. The main content area displays two search results. The first result is titled 'Stanje i perspektivnost vodosnabdjevanja Obrenovca i naselja optine posle poplava, 2014. godine' by Hajdin Bojan, Palomčić Dušan, Bajić Dragojka (2017). The second result is 'Upravljanje resursima podzemnih voda severne Bačke' by Bojan B. Hajdin (2013-03-07). A sidebar on the right shows a '4 godina' badge and a 'Prijavi grešku' button.

Figure 3. Search results view

as type of paper, format of the resource stored, department and chair of the Faculty to which the author belongs, narrower (eg. M14, M21, M22, . . .) and wider scientific paper category (M10, M20, M30, . . .), author and period of publication. Also, next to each facet, there is a number of documents that belong to a certain category, which can also help the user when finding the desired document or creating a list for export. On a specific example, it is possible to report papers for an individual researcher in a certain period, which is especially important when collecting data for election to the title.

5 Conclusion

As Omeka is an open source platform, many adaptations and additions to the functionality are possible, which in the case of the Digital Repository of the Faculty of Mining and Geology has been improved in the best possible way with the help of existing language resources for the Serbian language. These extensions to language resources significantly contribute to the quality of search results and a more user-friendly interface that is intended for non-IT users, and allows quick and easy storage, retrieval and use of metadata of entered works. Likewise, the simple user interface for administrators and librarians offered by Omeka brings many possibilities for editing access rights and site control, which enables easy work for librarians without the need for constant support from IT experts. In line with the needs of Omeka users, new open source modules are constantly appearing to improve existing ones and add new functionalities. They can, with administrator privileges, be added to the system without further modification of the source code, which is an additional benefit. As a further direction of the repository development, it is planned to switch to newer versions of Omeka and implement new add-ons for various functions if needed, as well as implement tools for extracting full-text information, such as named entities (locations, events, minerals, ore) to enrich the search capabilities.

References

Bruneau, Olivier, Nicolas Lasolle, Jean Lieber, Emmanuel Nauer, Siyana Pavlova et al.. “Applying and developing semantic web technologies for exploiting a corpus in history of science: The case study of the Henri Poincaré correspondence”. *Semantic Web* no. Preprint (2020): 1–20

- Cuenca, Esther Liberman and Maryanne Kowaleski. “Omeka and Other Digital Platforms for Undergraduate Research Projects on the Middle Ages”. *Digital Medievalist* Vol. 11, no. 1 (2018)
- Gill, Hannah, Jaycie Vos, Laura Villa-Torres and Maria Silvia Ramirez. “Migration and Inclusive Transnational Heritage: Digital Innovation and the New Roots Latino Oral History Initiative”. *The Oral History Review* (2020)
- Krstev, Cvetana. *Processing of Serbian. Automata, Texts and Electronic Dictionaries*. Faculty of Philology of the University of Belgrade, 2008
- Perry, Susan Chesley and Jessica Waggoner. “Processes for User-Centered Design and Development: The Omeka Curator Dashboard Project”. У *Developing In-House Digital Tools in Library Spaces*, 37–58. IGI Global, 2018
- Tomašević, Aleksandra, Ranka Stanković, Miloš Utvić, Ivan Obradović and Božo Kolonja. “Managing mining project documentation using human language technology”. *The Electronic Library* Vol. 36, no. 6 (2018):993-1009
- Vorkapić, Dalibor, Aleksandra Tomašević, Miljana Mladenović, Ranka Stanković and Nikola Vulović. “Digital Library From A Domain Of Criminalistics As A Foundation For A Forensic Text Analysis”. In *International Scientific Conference “Archibald Reiss Days” Thematic Conference Proceedings Of International Significance, Belgrade, 7-9 November 2017*, 2017
- Лазих, Биљана, Александра Томашевић и Михаило Шкорић. “Дигиталне библиотеке у рударству и геологији са посебним освртом на представљање сиве литературе”. У *Научна конференција Библиоинфо – 55 година од покретања наставе библиотекарства на високошколском нивоу, Београд 18. мај 2017*. Филолошки факултет Универзитета у Београду, 2019
- Томашевић, Александра, Биљана Лазих, Далибор Воркапић, Михаило Шкорић и Љиљана Колоња. “Употреба веб платформе Омека за дигиталне библиотеке из домена рударства”. *Инфотека* год. 17, бр. 2 (2017): 27–51, <http://infoteka.bg.ac.rs/pdf/Srp/2017-2/infoteka-2017-17-2-2.pdf>