The Implementation of Dais Repository in ISL SASA

Vladimir Živanović

vladimirludwig@yahoo.com

Institute for Serbian Language of SASA

Belgrade, Serbia

ABSTRACT: The paper presents the involvement of the Institute for the Serbian Language of SASA (ISL SASA) into the world of digital repositories. The ISL SASA was one of the first institutions in Serbia to have an institutional repository (2017). Before that, scarce research outputs of the ISL SASA were available online whereas the available resources remained invisible to major international infrastructures. In developing the repository, the role of the ISL SASA librarian has been crucial. It involved repository management, verifying the metadata accuracy, preserving digital content and further dissemination. More than 100,000 pages (2500 full-text records in the repository) were digitized in one year. This has significantly improved the international visibility of Serbian linguistic humanities.

KEYWORDS: Digital Repository, Institutional Repository, DAIS, DSpace, Open Science, Self-Archiving, Digital Humanistics.

1 Introduction to the digitization process in ISL SASA

The Ministry of Education, Science and Technological Development adopted the Open Science Platform in 2018 (Platforma, 2018), This Platform, which is based on the principles of Open Science and European Commission guidelines in this field, has introduced an obligation for individuals and institutions to make their own scientific production, which is publicly funded, also

publicly available and to deposit it in a digital repository. In order to meet the "transparency of scientific communication and methodology" and to ensure the availability of scientific production, it was necessary to develop the appropriate digital infrastructure, which in Serbia did not exist.

The open approach in this document is defined as "the right of every Internet user to read, download, store, print and use digital content of publications without financial expense". In addition to these rights, users will respect copyrights, i.e. a license that is associated with deposited work and with proper source guidance participate in the dissemination of open science, using some of the "institutional / thematic / national repositories". The type of repository for the deposit of scientific production is not strictly determined, thus leaving it to institutions to find an adequate solution in accordance with their capabilities. Under this Platform, each of the institutions involved is obliged to adopt a "local" policy, that is, an appropriate document on the implementation of the Platform that will regulate this process at the institutional level.

At the same time, during the preparation of this Platform, the Serbian Academy of Sciences and Arts (SASA), in cooperation with the University of Belgrade Computer Center (UBCC), has worked on developing its own repository designed for the Academy and related institutes. The repository was named Digital Archive of the SASA and the Institute – DAIS.[1] According to Branin's definition, repositories represent "models of systems and services designed to collect, organize, store, share and store digital information and knowledge of the institution" (Branin, 2004-2005, 237). The purpose of the DAIS digital repository is to enable the Academy, as well as the Institutes founded by the Academy, to permanently preserve their scientific production, thus publicly presenting the results of its scientific work.

The development of communication processes leads to the increasing availability of scientific papers in full text. General availability is still an ideal to strive for, but the scientific institutions that have been instigated by numerous open source initiatives, as well as the practice of large tech company services (Google Books, etc.) have already begun with setting up their digital repositories. One of the results of the successfully managed repository is the increase of visibility of scientific papers, especially in the field of Serbian humanities. Due to the commercial potential of technical and natural sciences, their visibility in the world greatly exceeds that of humanities. Such

---

[1] DAIS (on-line)

a disparity is not justified, given that it represents a culture at the global level.

Unlike a standard library, the formation of a digital repository requires technical preparation. DAIS was created by customizing open source software DSpace.[2] Customizing this software implied its localization, adaptation of the user interface, alignment with the guidelines of the OpenAIRE consortium for digital repositories, integration with ORCID and Altmetric.com, and the development of additional external applications that allow normative base, as well as the downloading of records and the massive input and correction of metadata.[3]

The reason for creating such software is based on the development of library management. Baudoin and Branschofsky from MIT argue that university communities depend on their libraries that allow them constant access to research and scientific work, and enable them to seek solutions to the problem of storage and take-over of intellectual work in the long run (Baudoin and Branschofsky, 2004, 32). The implementation and development of software, as well as its adaptation to the specific needs of Serbian scientific production, was performed by UBCC. The computer center is constantly working on maintaining the repository and its continuous improvement.[4]

One needs to note the difference between a digital collection and digital repositories. While a collection remains local – for example, the collection of digital objects that institutions or individuals upload on their own internet sites – it does not meet the standards of interoperability (DOI number, other persistant identifiers, metadata and attributes, structures and standards for describing a document, metadata exchange protocol, etc. (Van de Sompel and Nelson, 2015)) The Digital Repository, however, has presented a digital collection organized in such a way that it can be archived, preserved and disseminated according to its defined goals and protocols compatible with other databases. Its main purpose is "the presentation of intellectual conduct" and the promotion of accessibility (Winter and Bowen-Chang, 2010, 320).

---

[2] Dspace open source software was developed by the Massachusetts Institute of Technology (MIT) together with Hewlett-Packard in 2002. Its quality is seen in an easy and open approach to various types of digital objects: text, images, audio or video material.

[3] For more on DSpace software, as the backbone for the development of digital repositories, see (Rajović and Ševkušić, 2018)

[4] Customizing open source software to local needs is the biggest challenge in the initial stages of organizing the repository and requires high-quality technical support. The software code is continuously adapting.

The ISL SASA experience in organizing and managing a digital repository can be useful to other institutions involved in organizing digital full-text collections in open access.

2 Structure

In order for an institutional repository to become functional, the filling of repository is the first major obstacle to overcome. That's why in ISL SASA we had two phases of filling. A properly organized repository increases the visibility of the scientific production of a particular institution, thus enabling better citation of researchers (Piwowar and Haustein, 2018).

The first phase is the collection, digitization and installation of the digital content of the journal and the ISL SASA monograph. The history of periodicals of the Institute is very long. Two of the four scientific journals published by the Institute have been published for more than a hundred years.[5] Published volumes of these journals have several hundred volumes. In addition to digitazion of the serial publications in the first phase in the institutional repository, we have digitazed the scientific papers of prominent scientists from the Institute. There are colleagues who are retired and some of them are deceased. By introducing their digitized works, several successful criteria for the functioning of the repository are fulfilled: the good practice of future work is established, it participates in the formation of the customs of future users, and the scientific production is presented in the best light.[6] This also covers the diachrony of the scientific production of the Institute.

The second phase, which represents a synchronous level, concerns the digitazion of the current scientific production, at the latest 18 months from the date of the publication of a scientific article or monograph, and is defined by the Rulebook adopted at the institute level. This fulfills the conditions provided by the Platform of the Ministry.

Institutions that engage in the formation of a repository should define their goals and clearly identify the set of traits of a well-trained staff involved in the process of maintaining a repository (Winter and Bowen-Chang, 2010, 323). Training a personnel to work in the repository can be a challenge.

---

[5] The first issue of the journal *Srpski dijalektološki zbornik* was published in 1904, *Južnoslovenski filolog* in 1913, the journal *Naš jezik* in 1933, and the *Lingvističke aktuelnosti* 2000.

[6] Unlike in technical sciences, in linguistics, many old scholarly papers do not lose importance because they are still relevant for the research.

Lack of initiative and enthusiasm in accepting a different approach to scientific publications, as well as insufficient awareness of its significance, can be reflected in the first stages of organizing the repository. It is very important that managerial staff, as well as associates of a scientific institution, recognize the importance of establishing a repository.

All researchers in the Institute are obliged to depose their scholarly work in the process of self-archiving. Training is organized for the researchers in workshops that contain practical work (personal input of metadata and deposition of objects), and a detailed explanation is prepared in the Power-Point presentations that they can use for their own work. Once a researcher deposes work and describes all the associated metadata, the librarian as an administrator checks the accuracy of entries and approves the record, since only he is in charge of subsequently changing metadata or removing digital object.[7]

A metadata collection involves entering basic document data, text summary, and keywords, as well as determining the level of document availability and the type of license. Also, providing data on the project within which the work is funded is mandatory. The same pertains to data on the type and version of the document, the availability of content and licenses, which are entered in accordance with the current guidelines for the digital repositories of the OpenAIRE consortium. The domain that is becoming increasingly important is monitoring copyright compliance and contacting publishers of the journals for copyright regulation.

Furthermore, the librarian uses additional DAIS services, such as a metadata editing service, a normative database, and a service for downloading the already existing digital content from other repositories. The UBCC Development Team has independently developed the Ellena2 application, which includes a normative database, a service for massive metadata correction, and a mass-feed service for metadata. In the early stage of its development, there were two separate applications: Ellena (normative file and massive metadata correction) and MultiLoad (downloading records from other DSpace repositories and massive import of metadata in XML or RIS format). At the beginning of 2019, these two applications were integrated into the Ellena2 application. The reason for the creation of an independent application is the lack of a DSpace platform, which in itself does not contain a normative

---

[7] The DAIS repository consists of multiple levels of access and content management permissions, and the administrator has the ability to restrict the user's authority when self-archiving, thereby ensuring quality control, accuracy of data and database integrity.

file module. In addition, initial tests have shown that the system for down-loading metadata that was already embedded in DSpace does not provide satisfactory results.

Content management allows researchers to use the repository in a wider range than the Open Source Platform requires because DAIS supports the input and processing of multiple types of documents and formats. In addition to scientific articles, chapters in a monograph, monographs, etc., the associates are able to deposit and pre-print versions of their work, doctoral or master thesis, reports and all other contents called "gray" literature.[8] DSpace indexes all readable text from the deposited documents, so it can also be searched. Ranking indexed documents is done through formal parameters, i.e. by relevance, title and date. DSpace supports the storage of virtually all formats. DSpace does not have a system for viewing documents in the way it works in Omeka and other platforms primarily intended for displaying cultural heritage and digitized material. DSpace is designed to present scientific content, whereas stored files can be saved to the user's computer and then opened through appropriate application: PDF, JPG and similar formats are opened in the new browser window. DSpace can be up-graded to display content within the platform itself (for example, listening books), but for users in ISL SASA, it is not necessary at this moment.

3 Practical work

ISL SASA started the work on DAIS towards the end of 2017. Considering the relevance of ISL SASA throughout the Slavic world, a part of the year-books has already been scanned within the Google Books service but has not been made publicly available.[9] Several US universities have digitized Serbian journals as part of their collections in cooperation with Google. These digital copies were collected and they represented approximately 50% of the total volume of the Journals. The rest was to be digitized. The work, which involved organizing, scanning, redesigning digital copies, working on the optical character recognition of content, and depositing it in the database, has

---

[8] For more on scientific "gray" literature see (Ferreras-Fernández and Merlo-Vega, 2015) and (Ćirković, 2018).

[9] One of the major Google Books missions is the organization of world-level information. By digitizing the world's library heritage, Google becomes a universal library of world-class knowledge. For more about the partner program with Google Books, see (Шевкушић, 2013)

begun. This way over 50,000 pages of text were prepared in a year. The lack of infrastructure in certain segments of the work was replaced by cooperation with Google Books: thanks to the fact that ISL SASA made scanned publications available through this service, we managed to reduce the scanned material, which was very large in memory,[10] to the optimal size, and we provided optical text recognition.[11]

In the first phase, we made the entire existing production of the Institute publicly available by depositing every journal on the annual basis, i. e. as a single document, while it remained to deposit each article separately, thus making it more visible. In addition to the journals, the monographic production is included in DAIS, and it contains more than 60 volumes. This part of the first phase was finished and the Institute presented a retrospective of its publishing activity in DAIS (250 volumes of magazines and monographs).

The second segment in the first phase of the establishment of DAIS was the collection, preparation and distribution of articles of prominent ISJ SANU scholars. In the first year, more than 1,200 scholarly articles entered the repository.[12] A small number of these articles already existed in digital form and the whole digitization work was supposed to go through the process of physical preparation and work on scanning, editing scanned documents, and creating metadata. According to Foster and Gibson, the institutional repository "without content, is the same as a series of empty

---

[10] In order to make our documents readable we reduced the size of the files (because some also had more than 1 GB). The size of the document burdens the repository and its resources as well as the end user who wants to get an optimal size document in a short time. For example, a year's magazine has an average of 200 and 300 Mb, but after Google's optical character recognition, the document has been reduced to approximately 30 Mb.

[11] So far, the evaluation of the success of optical text recognition has not been done. When services of a domestic initiative that deals with the optical text recognition, especially the Cyrillic are available and affordable, we will gladly re-enter the OCR and repossess the documents with more precisely recognized text into the repository.

[12] They include Academic Irena Grickat, Academic Mitar Pešikan, Dr. Ivan Popović, Dr. Berislav Nikolić, Scientific adviser dr. Egon Fekete, Prof. dr Dimitrije E. Stefanović, Academic Aleksandar Loma, Scientific advisor dr. Jasna Vlajić-Popović, Scientific advisor dr. Stana Ristić, Scientific advisor dr. Sreto Tanasić and others.

shelves" (Gibbons and Foster, 2005, par 4).[13] Therefore, the establishment of user habits and setting a good example is essential to empower the initial dynamics of a repository.

Creating and organizing digital content in DAIS is in line with the requirements of the European Commission concerning open access to publications, and it enables the further dissemination of scientific information through European and international portals such as OpenAire, BASE, CORE and Google Scholar. One of the fields for entering metadata is the name of the project within which the work was created. This field, through the link, connects all the works of a particular project to DAIS, but also in the OpenAire database, which in one place collects data of the results of all projects of the European Commission, as well as data on the results of national projects from about ten countries, one among them being Serbia (OpenAire, 2019). Also, using the ORCID (Open Researcher and Contributor ID) identifier, all of the author's publications are combined and linked in one place, and published under different variant names.

When determining the profiles of the digital collection, ISL SANU decided to provide full text in open access for all of its editions, as well as for the published work of the employees, with restrictions that determine the copyright of other publishers. As for the ISL editions, it was decided to apply CC licenses. This simple and standardized way of copyright regulation has multiple levels. The institute opted for the CC BY-NC-ND module (Attribution – NonCommercial – NoDerivatives), which implies that it is mandatory to indicate the name of the author and that the author gives permission to download and further distribute the work. This most restrictive of free licenses does not allow further processing of the work or its commercial use.

In its architecture, DSpace software allows creating collections in the digital repository. These collections categorize and sort content thematically, formally or otherwise. At present, ISL SASA in DAIS has seven collections. In addition to the four collections that classify the periodicals of the Institute, the two are created for monographic series, while the remaining one is general. The general collection collects the Institute's Proceedings and other materials, as well as the work of researchers published outside the Institute. The problem of classification always imposes various solutions. The material

---

[13] As an example of good practice, we can mention the repository of the Massachusetts Institute of Technology (MIT), which has owned over 106,000 digital records in its repository by 2019 (DSpace, 2019)

could have been classified differently, but the optimal approach also takes into account the easy navigation by the end users.

4 Conclusion

During the relatively short period of time in the field of digital humanities, ISL SASA has undergone major changes. The institutional repository has come to life and serves a number of purposes. The main purpose is to promote the concept of Open Science and to represent the scientific production of the Institute, which fulfills the criteria provided by the Ministry's Platform. What is publicly funded should be publicly available. The institutional repository is also developing the presentation of retrospective scholarly production. Almost everything published by ISJ SANU has been put on the repository. This way the old scholarly papers that are still relevant become easily accessible. The library has thus become a participant of scholarly research. According to Baudoin and Branschofsky, the creation of a digital repository changes the way we think about the life cycle of scientific research (Baudoin and Branschofsky, 2004, 43).

Thanks to the successful cooperation of the library and the management of the Institute, as well as research associates, the digital repository is established, filled with scholarly papers and ready to follow future scientific production. Through the successful cooperation of all factors in this complex process, the credibility of the established repository is being built and promoted. It is important to note that the ISJ researchers have shown willingness to engage in the work, which had a significant influence on the results. Now the researchers' production is almost completely in free access. The percentage of open access content in the ISL SANU collections is 100%, which is a great success, bearing in mind that the open approach to humanities in the world is much less represented and develops more slowly than in other disciplines. According to reports made for the needs of the European Commission, the greatest restriction in open access is within social sciences and humanities (Archambault and Roberge, 2014, 20).

References

"DSpace". Duraspace. Accessed: 12.03.2019, https://duraspace.org/dspace/about/

"OpenAire". European Union/European Commission. Accessed: 08.03.2019, https://www.openaire.eu

"Платформа за отворену науку". Министарство просвете, науке и технолошког развоја. Accessed: 21.02.2018, http://www.mpn.gov.rs/wp-content/uploads/2018/07/Platforma-za-otvorenu-nauku.pdf

Archambault, É., D. Amyot, P. Deschamps, A. Nicol, F. Provencher, L. Rebout, and G. Roberge. "Proportion of Open Access Papers Published in Peer-Reviewed Journals at the European and World Levels – 1996–2013". *Science-Metrix* (2014). Accessed: 20.03.2019, http://science-metrix.com/en/publications/reports/proportion-of-open-access-papers-published-in-peer-reviewed-journals-at-the

Baudoin, Patsy and Margret Branschofsky. "Implementing an Institutional Repository: The DSpace Experience at MIT". *Science & Technology Libraries* Vol. 24, no. 1/2 (2004): 31–45.

Branin, Joseph. *Encyclopedia of Library and Information Science*. Institutional Repositories. New York, N.Y.: Marcel Dekker, 2004-2005.

Ferreras-Fernández, Tránsito, Francisco J. García-Peñalvo and José A. Merlo-Vega. "Open Access Repositories as Channel of Publication Scientific Grey Literature". In *TEEM '15 Proceedings of the 3rd International Conference on Technological Ecosystems for Enhancing Multiculturality*, 419–426. New York, NY, USA: ACM, 2015.

Gibbons, Susan and Nancy Fried Foster. "Understanding Faculty to Improve Content Recruitment for Institutional Repositories". *D-Lib Magazine* Vol. 11, no. 1 (2005). Accessed: 25.03.2019, http://www.dlib.org/dlib/january05/foster/01foster.html

Piwowar, Heather, Jason Priem, Vincent Larivière, Juan Pablo Alperin, Lisa Matthias, Bree Norlander, Ashley Farley, Jevin West and Stefanie Haustein. "The State of OA: A Large-Scale Analysis of the Prevalence and Impact of Open Access Articles". *PeerJ* Vol. 6 (2018). Accessed: 01.04.2019, https://peerj.com/articles/4375/;6:e4375, DOI:10.7717/peerj.4375

Rajović, Vasilije, Biljana Kosanović and Milica Ševkušić. "DSpace – institutional repositories – dissemination of research results: A local case study". In *Primena slobodnog softvera i otvorenog hardvera PSSOH 2018*. Belgrade, Serbia: University of Belgrade – School of Electrical Engineering and Academic Mind, 2018.

Van de Sompel, Herbert and Michael L. Nelson. "Reminiscing About 15 Years of Interoperability Efforts". *D-Lib Magazine* Vol. 21, no. 11/12 (2015). Accessed: 15.03.2019, http://www.dlib.org/dlib/november15/vandesompel/11vandesompel.html

Winter, Marsha and Portia Bowen-Chang. "Dealing with DSpace: The Experience at the University of the West Indies, St. Augustine". *New Library World* Vol. 111, no. 7/8 (2010): 320–332.

Ćirković, S. "Grey Literature – The Chameleon of Information Resources". *Infotheca - Journal For Digital Humanities*, Vol. 18, no. 1 (2018): 75–83. doi:10.18485/infotheca.2018.18.1.5

Шевкушић, Милица, "Партнерски програм Гугл књиге као платформа отвореног приступа у научним библиотекама". In *Отворен приступ знању у библиотекама, организатори конференције Библиотекарско друштво Србије [и] Народна библиотека Србије*, 187–207. Београд: Библиотекарско друштво Србије, 2013.