Role of a Croatian National Repository Infrastructure in Promotion and Support of Research Data Management

Posavec, Kristina; Celjak, Draženko; Jertec Musap, Ljiljana

Download date / Datum preuzimanja: **2024-05-05**

PRACTICE PAPER

Role of a Croatian National Repository Infrastructure in Promotion and Support of Research Data Management

Kristina Posavec, Draženko Celjak and Ljiljana Jertec Musap

Middleware and Data Services Department, University of Zagreb, University Computing Centre (SRCE), Zagreb, HR
Corresponding author: Kristina Posavec (kristina.posavec@srce.hr)

The paper will give an overview of national infrastructure for digital repositories, Digital Academic Archives and Repositories (DABAR), and its role as technology steward in raising awareness about research data management (RDM) and promoting good practices in the Croatian A&R community. The University of Zagreb, University Computing Centre (SRCE) is providing national infrastructure DABAR suitable for storing and dissemination of different types of digital objects. Through DABAR, all Croatian higher education and research institutions can establish their digital repository. A strong collaboration between SRCEs DABAR team and institutions repository managers has proven to be important in the process of disseminating knowledge about research data management among researchers and the scientific community at large. The paper will provide information about this collaboration during the project RDA Europe 4.0 – The European plug-in to the global Research Data Alliance (RDA). The main goal of this collaboration is to raise awareness about the importance of managing and sharing research data.

Keywords: repository; research data; datasets; RDM; data steward

Introduction

The main goal of this paper is to give insight on how SRCEs DABAR team and librarians from four university libraries[1] collaborated in a process of dissemination of the principles on research data management (RDM) and research data sharing. It is important to emphasize that all involved librarians are from institutions that already have hosted their repositories on the SRCEs DABAR platform. In addition, this paper will give an overview of the role that institutional repository managers have in process of spreading knowledge among researchers about the importance of RDM and research data sharing, i.e. the role of data steward for the research community. It is important to see what RDM is and why it is important for the research community. Whyte and Tedds (2011) define RDM as "the organization of data, from its entry to the research cycle through to the dissemination and archiving of valuable results and it aims to ensure reliable verification of results and permits new and innovative research built on existing information"(p. 1). University Library System (ULS) from the University of Pittsburgh defines RDM as a set of activities that correlates with the organization, storage, preservation, and sharing of data that are collected and used in a research project (ULS 2020). ULS points out a few examples of RDM activities that include the everyday management of research data during the lifetime of a research project and decisions about how data will be preserved and shared after the project is completed. In process of educating researchers about RDM, institutional repository managers, which are in Croatian case, mostly librarians of higher education and research institutions, can have a major role because they store, manage and archive objects, and one of them is research data. An institutional repository can be described as "an electronic system that captures, disseminates, and preserves

---

[1] National and University Library in Zagreb, University Library Rijeka, University of Split Library and City and University Library in Osijek.

Art. 48, page 2 of 7

Posavec et al: Role of a Croatian National Repository Infrastructure in Promotion and Support of Research Data Management

intellectual results of a group of universities or a single university" (Kamraninia, Abrizah 2017: 122). One of its main goals is to "make it easier for researchers to disseminate and share research outputs and thus support the open access goal of scholarly communication" (Hockx-Yu 2006: 233). Lynch (2003) defines institutional repositories as "a set of services that a university offers to the members of its community for the management and dissemination of digital materials created by the institution and its community members" (p. 328). Because of their knowledge about data, institutional repository managers are the ideal target group for contributing to RDM training and raising awareness on sharing research data. As Peng et al. (2016) are stressing "the open data policy and data sharing requirements have brought closer than ever two groups of people — data producers and data managers — who are often at separate stages of the lifecycle of scientific data products" (p. 2). In this case, institutional repository managers are data managers, and researchers are data producers who need guidance with data management. For successful collaboration between those two groups, qualitative communication needs to be established.

Digital Academic Archives and Repositories (DABAR)

In the 2015 University of Zagreb, University Computing Centre (SRCE) together with research and higher education institutions in Croatia built the national infrastructure for digital repositories DABAR.[2] Repositories in DABAR are intended for the preservation and dissemination of various digital objects produced by Croatian higher education and research institutions (Celjak et al., 2017).

One of the main missions of SRCEs service DABAR is to facilitate the establishment and maintenance of a large number of reliable and interoperable institutional and thematic digital repositories and archives (DABAR 2020; Celjak et al. 2017), it is an essential element of the national data infrastructure in Croatia. DABAR is PaaS (platform as a service) made of technical infrastructure and a team of experts that provide librarians, editors of repositories, researchers, etc. with training and workshops to inform and educate them about current and new functionalities in repositories. DABAR is developed on Islandora[3] which is an open-source software framework used for digital object management based on Drupal CMS (Content Management System), Fedora repository system, and Apache SOLR search platform. All repositories in DABAR are openly available, and to update their content users authenticate using the authentication and authorisation Infrastructure of science and higher education in Croatia AAI@EduHr.

Content that is stored in repositories is called a digital object and "is defined as a digital data file, a paper record, an image, an article, or a collection of any or a mix of those items" (Peng et al. 2016: 6). An important objective for this paper is data, e.g. datasets which Peng et al. (2016) describe as "model output such as forecasts, projections, analyses, or re-analyses" (Peng et al. 2016: 5). In addition, the Cambridge dictionary defines datasets as a collection of separate sets of information that is treated as a single unit by a computer. Other types of objects that can be stored in DABAR are preprint papers, reviewed articles, conference papers, datasets, dissertations, theses, books, teaching materials, images, video and audio files, presentations, digitalized materials, educational content, etc. It is important to mention that every object, which is stored in the DABAR repository, must contain a nationally agreed set of metadata (this is defined in collaboration with working groups), which makes objects more findable. The support for datasets is implemented following the FAIR data principles and each dataset is described with appropriate metadata, assigned a permanent identifier (URN:NBN), and is made available through repository search interface where metadata is exposed to other services such as OpenAIRE portal or Google Scholar (DABAR 2020). Metadata description of the dataset which could be stored in DABAR is developed combining properties defined in DataCite Metadata Schema and the local needs of repositories in Croatia.

All institutions that have established their repository within DABAR infrastructure, have total control over access rights and repository content usage, as well as a possibility of open access publishing and increasing the visibility of the content and the institution itself. DABAR offers institutions free and reliable solution for long-term preservation of various digital objects described by metadata, a possibility to modify the design and content of the repository interface, a possibility to establish a thematic repository for the research community (DABAR 2020).

---

[2] National infrastructure for digital repositories Digital Academic Archives and Repositories (DABAR) is available online on https://dabar.srce.hr/en.

[3] More about Islandora on https://islandora.ca.

Posavec et al: Role of a Croatian National Repository Infrastructure
in Promotion and Support of Research Data Management

Art. 48, page 3 of 7

Role of DABARs institutional repository managers as data stewards

Loshin (2009) defines a data steward as an "individual responsible for collecting, collating, and evaluating issues and problems with data" (p. 1). Peng et al. (2016) state that they are "responsible for ensuring compliance with data management standards, including community standards on data quality metadata and policies" (Peng et al. 2016: 9). Clare et al. (2019) describe data stewards as motivators for researchers and their mission is to "more effectively engage with their research community about data and to provide better RDM support for researchers" (p. 106). In addition, they "serve as bridges between the researchers and all other research support services, such as the library, ethics committee, ICT, privacy and legal teams and the daily job of the Data Steward is to respond to researchers' requests, advise them, and promote good RDM practices" (Clare et al. 2019: 113).

When considering data stewards Peng et al. (2016) are providing a classification of this term. Their classification consists out of three categories of stewards: "data steward – data management and preservation, scientific steward – scientific data quality management and usability, and technology steward – system engineering and software development" (p. 4). This paper is focused on the first category of data steward whose main role is to manage and preserve data during research projects. Some of the characteristics of data stewards are described by Clare et al. (2019) where it is stated that qualitative data steward includes high interpersonal skills. It is stressed that the most important skill is communication because data steward needs to establish a good relationship with the researcher and the best ground for that is an environment that is based on good communication and collaboration. Additionally, data stewards need to have experience with the research process because one of their tasks is to guide and give support to researchers during their projects and researches. In addition, it could be stated that SRCEs service DABAR has a role of technology steward because they provide technical support to all institutional repositories (within DABAR infrastructure) and continuously upgrade and improve functionalities for their users.

During the project RDA Europe 4.0 – The European plug-in as part of the global Research Data Alliance[4] SRCE became a Croatian national RDA node. As one of the project activities, SRCE started to disseminate RDA outputs and recommendations within the Croatian research community to raise awareness among researchers about the importance of managing and sharing research data in which the data management plans (DMP) has an important role. Another goal was to educate the research community on how to manage their research data during the project's life cycle. These goals are fully achieved in collaboration with four major university libraries (National and University Library in Zagreb, University Library Rijeka, University of Split Library, and City and University Library in Osijek) with which SRCE had signed a Memorandum of Understanding. One of the tasks of this collaboration was to adapt RDA outputs to the Croatian community. The main focus was on FAIR data and principles, RDM, DMP, 23 Things: Libraries for Research Data,[5] Engaging Researchers with Data Management: The Cookbook,[6] CoreTrustSeal certification of trustworthy repositories,[7] types of data licensing, sharing research data, and practical demonstration of uploading dataset into DABAR. Main adaptations were made on document 23 Things: Libraries for Research Data which is translated to Croatian language and some recommendations are replaced with available tools and information for RMD that are more appropriate for Croatian context and community, e.g. recommendations for the long-term preservation are directing reads to the national DABAR platform, the section about metadata is adjusted so that it complies with DABAR and Croatian infrastructure. In addition, information about FAIR principles is added to the document. The adjusted version is appropriate for librarians, students, institutional repository managers, and researchers in Croatia.

Associates from four institutional libraries and SRCE's DABAR team developed content that other institutional repositories managers can use to educate researchers and scientists about research data management and the importance of sharing their data with the wider community. The role of the DABAR team in this collaboration was important because it was a liaison between institutional repositories managers and

---

[4] The Research Data Alliance is defined an international member-based organisation focused on the development of infrastructure and community activities to reduce the social and technical barriers to data sharing and re-use and to promote the acceleration of data driven innovation and discovery worldwide. More about RDA is available on https://www.rd-alliance.org/rda-europe.

[5] Developed by RDA's Libraries for Research Data Interest Group and research datasets, this output was adapted for Croatian community. Original document is available on https://bit.ly/2Uh1UiY.

[6] Book is available on: https://www.openbookpublishers.com/product/1080.

[7] It is a community-based non-profit organization promoting sustainable and trustworthy data infrastructures. More about it is available on: https://www.coretrustseal.org/ and https://www.rd-alliance.org/coretrustseal-certification-launched.

Art. 48, page 4 of 7

Posavec et al: Role of a Croatian National Repository Infrastructure
in Promotion and Support of Research Data Management

their engagement in RDA. All materials are stored in institutional repositories under an open-access license so anyone can use it.[8]

In February 2020, SRCE and associates from four university libraries held an event about research data at the University campus Rijeka (Croatia), where they presented some of RDA outputs and recommendations to librarians and researchers. After the event, all participants were asked to fill in an online survey about the level of satisfaction and usefulness of topics that were presented. Out of 32 participants, only 25% (n = 8) filled out the survey. Analysis of participants responses is conducted on sample of 62,5% (n = 5) librarians and 37,5% (n = 3) researchers. Half of the respondents (50%) considers that event fulfilled their expectations, 12,5% of respondents were not satisfied with the event, and 37,5% of respondents were somewhat satisfied with the event. Half of the respondents (50%) consider that presented content would be useful for their work, 12,5% of respondents presented content considers unuseful, and 37,5% of respondents think that content is somewhat useful. All respondents would recommend this event to their colleagues. The most useful presentations which were presented during event were as follows: Croatian Science Foundation (CSF) plans for DMPs (62,5%), data licencing (62,5%), FAIR data (62,5%), importance of data management (62,5%), practical examples of RDA (50%), demonstration of uploading research data into institutional repository (50%), 23 Things: Libraries for Research Data (37,5%), informing and engaging researchers into RDM (37,5%) and CoreTrustSeal (37,5%). It can be seen that the majority of respondents found interesting topics related to CSF's plans regarding the implementation of mandatory DMP into project proposals. Other interesting topics are the importance of managing research data, how to choose the right license, and topics about FAIR data and principles. Nevertheless, all other topics are equally important for librarians and researchers but it can be stated that in the Croatian community awareness about sharing research data is still low. Further education about all segments of RDM needs to be disseminated so that librarians and researchers could benefit from it. Participants provided positive feedback about the event and in comments, they emphasize the importance of organizing educational webinars and the development of materials about RDM. They stated that it would be useful to include more examples of short and long-term storage of different datasets from various science fields. In addition, the current state of the Croatian community regarding open science is shown on information which were collected as part of the NI4OS-Europe project[9] in a survey in which participated 575 participants from Albania, Armenia, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Georgia, Greece, Hungary, Moldova, Montenegro, North Macedonia, Romania, Serbia, and Slovenia, out of which 76 response come from Croatia. Survey results show that 43 participants in Croatia are unfamiliar with FAIR data and principles and only 22 participants are familiar with the term EOSC (European Open Science Cloud). The majority of participants (about 90% from all countries) stated that training about making data FAIR is much needed.[10]

These two surveys show that there is interest about sharing research data and management but there is much work to be done for establishing quality communication between librarians as data managers and researchers as data producers.

How did Croatian RDA node disseminate and promote RDM in the Croatian community

Croatian RDA node activities during the RDA Europe 4.0 project included the organization of one on-site event and three webinars about RDM for Croatian researchers and librarians.

One of the RDA outputs is a book "Engaging researchers with data management" that provides 24 case studies from different universities around the globe, on how did librarians engage researchers into RDM. The majority of librarians and data stewards emphasize the importance of educating researchers and scientists about RDM in a form of webinars or workshops (Clare et al. 2019). Croatian RDA node chose this approach for dissemination, promotion, and education of the Croatian research community about the importance of RDM.

Some universities had taken a step forward and integrated RDM into their curriculums. For example, University UiT The Arctic in Norway has Ph.D. and open courses that include "seminars with a focus on academic integrity and open science, available to all Ph.D. students and since 2019 seminars are obligatory for

---

[8] Materials are developed on Croatian language and are available on https://www.srce.unizg.hr/rda/dogadjaj-rijeka.

[9] More about project NI4OS-Europe is available on https://ni4os.eu/. Dataset: Kosanović, Biljana, Ševkušić, Milica, & Otašević, Vladimir. (2020). Open Science stakeholders in Albania, Armenia, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Georgia, Greece, Hungary, Moldova, Montenegro, North Macedonia, Romania, Serbia and Slovenia [Data set]. Zenodo. http://doi.org/10.5281/zenodo.3722930.

[10] Visualizations and other results are available on https://ni4os.eu/survey-results/.

Posavec et al: Role of a Croatian National Repository Infrastructure
in Promotion and Support of Research Data Management

Art. 48, page 5 of 7

law students" (Clare et al. 2019: 61). The University of Minnesota in the USA developed methods class outreach in which they integrated RDM into existing classes, e.g. in "regular research methods courses, the RDM training usually lasts between 60 to 90 minutes with a class size of five to 20 students" (Clare et al. 2019: 57). Their courses have the same structure, which includes naming data files, describing files organization, explaining how to share and archive data and how to secure their data. Structure and content of the events/webinars were developed based on this information and approach were practical examples and information about RDM was included in materials. Events covered the following topics: FAIR data and principles, the introduction of DMP as part of project documentation in Croatia, types of data licensing, best practices of research data management during and after the research data cycle which include naming conventions, versioning, licensing, long term preservation of data, available and safe online storage for Croatian researchers during the data collection and presentation of adapted RDA output "23 Things: Libraries for Research Data" to the Croatian context.

An experience that institutional repositories managers have about managing, storing, and archiving data can help researchers with their research data management so they are a key point for raising awareness about RDM in the research community (Katayoon and Abrizah, 2017). The motivation of researchers could be increased through different types of education processes, e.g. workshops, webinars, training, tips, and tricks, etc. One of the key motivating factors for researchers to engage more in sharing their research data is a citation (Clare et al. 2019). If researchers share their data they can raise the visibility of their projects, findings, and publications, throughout they can improve other research projects and they can have higher citation rank. This was the inspiration for the Croatian RDA node to invite and include researchers and university professors to the events program so they could share their experience with RDM with participants and emphasize the importance and benefits for researchers to manage and share their data.

Croatian RDA node has involved three eminent Croatian university professors to participate in the events that practice RDM in their work. The aim was for participants to see how RDM is conducted in practice and to get the first-hand experience from the representatives of the research community. In two events participated associates from the main Croatian research funding organization (Croatian Science Foundation, CSF), who presented to the participants the work of the CSF. In addition, the emphasis was put on the importance of a data management plan (DMP) and the CSF's plans to introduce DMP as a mandatory part of project applications. The last webinar was the demonstration of long-term preservation of dataset on test repository in DABAR platform which was demonstrated by a researcher from the School of Medicine, University of Split where the researcher used his research data and codebook. During the demonstration researcher explained and presented his own research and process of long-term storage on to the platform. Dissemination and communication plan of Croatian RDA node for the events was based on the use of practical examples for RDM (e.g. how to use a naming convention, versioning, data licensing, etc.), involvement of researchers from the Croatian research community who are practicing RDM in their work with aim to share their experience with participants (e.g. how to manage and organize research data, what are the benefits of RDM for researchers, etc.), development of useful materials for Croatian research community with goal to encourage, motivate and help researchers.

Institutional repositories managers must share their knowledge and engage more with research communities. The dissemination process that emphasizes the importance and usefulness of RDM and data sharing is crucial for motivation along with examples of best practices and case studies. Through this approach researchers can see how other researchers are managing their research data and what benefits they have from it. Networking is another way to engage the research community where researchers with more experience in RDM can share their experience and knowledge with researchers that have less or none experience.

Due to the high interest of the Croatian community for the webinars (on all events participated 499 participants), the Croatian RDA node developed a handbook, "How to Manage Research Data?"[11] which is published in open access and contains practical guidelines and recommendations about RDM during the whole research data lifecycle. The handbook is created for the Croatian research community and includes available Croatian e-infrastructure for RDM in the Croatian context. After each chapter, there is a list of free online tools for RDM.

---

[11] Handbook "How to Manage Research Data?" is published in open access and it is available on https://urn.nsk.hr/urn:nbn:hr:102:810090.

Art. 48, page 6 of 7

Posavec et al: Role of a Croatian National Repository Infrastructure
in Promotion and Support of Research Data Management

Conclusion

Institutional repositories could be a central point for engaging researchers into RDM and data sharing through various activates. As it is shown on the example of Croatian RDA node, successful collaboration among associates from university libraries and DABAR's team produced two types of materials, one for researchers (1) and one for the librarians (2). Materials that are developed for researchers (1) aim to engage researchers with RDM and sharing data with the community through their institutional repositories or other available services of their choosing. Materials that are developed for librarians (2) aim to disseminate knowledge about different ways how to become data steward and how to motivate and engage researchers with RDM within their institutions. These two types of materials were foundation for development of handbook about RDM which is intended for the Croatian community and infrastructure as useful literature with practical examples and guidelines.

All preformed activities contributed to building stronger relationships between institutional repositories managers and the research community, establishing better communication channels for sharing knowledge and expertise about the importance of RDM and data sharing in the Croatian community.

From the experience of the Croatian RDA node, it can be concluded that for successful promotion and dissemination of RDM on a national level focus must be put on the development of educational materials and trainings which include practical examples so that users see how RDM could be done in practice, not only in theory (1), the experience of researchers from the community with RDM is useful because users can see benefits of RDM from first-hand experience (2), usage of examples which could be applied in real-life situations which are understandable not just exclusively to researchers but also the students because they are the future research community (3) and qualitative data stewards who understand research data lifecycle and have experience in research themselves (4).

Acknowledgements

Competing Interests

The authors have no competing interests to declare.

References

Celjak, D, Bekić, Z, Cundeković, M, Jertec, LJ, Milinović, M and **Zubić, A.** 2017. DABAR – the national infrastructure for digital repositories. *EUNIS 23rd Annual Congress: Shaping the Digital Future of Universities: Book of Proceedings*, pp. 16–24.

Clare, C, Cruz, M, Papadopoulou, E, Savage, J, Teperek, M, Wang, Y, Yeomans, J, et al. 2019. *Engaging Researchers with Data Management: The Cookbook (epub).* Open Book Publishers. DOI: https://doi.org/10.11647/OBP.0185

Dabar.srce.hr. 2020. *What Is DABAR? | Digital Academic Archives And Repositories* [Online]. Available at: https://dabar.srce.hr/en/dabar [Last accessed 9 March 2020].

Hockx-Yu, H. 2006. Digital preservation in the context of institutional repositories. *Program: electronic library and information systems*, 40(3): 232–243. DOI: https://doi.org/10.1108/00330330610681312

Katayoon, K and **Abrizah, A.** 2017. Librarians' role as Change Agents for Institutional Repositories: A Case Of Malaysian Academic Libraries. *Malaysian Journal of Library & Information Science*, 15(3): 121–133.

Peng, G, Ritchey, NA, Casey, KS, Kearns, EJ, A Privette, JL, Saunders, D, Ansari, S, et al. 2016. Scientific Stewardship in the Open Data and Big Data Era-Roles and Responsibilities of Stewards and Other Major Product Stakeholders. *D-Lib Magazine*, 22(5/6): 1–24. DOI: https://doi.org/10.1045/may2016-peng

University of Pittsburgh Library System (ULS). 2020. *Libguides: Research Data Management @ Pitt: Understanding Research Data Management.* [online] Available at: https://pitt.libguides.com/managedata#s-lg-box-4890337 [Last accessed 24 March 2020].

Whyte, A and **Tedds, J.** 2011. Making the Case for Research Data Management. *DCC Briefing Papers.* Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/briefing-papers [Last accessed 14 August 2012].

Posavec et al: Role of a Croatian National Repository Infrastructure
in Promotion and Support of Research Data Management

Art. 48, page 7 of 7