

# Knowledge Access through Libraries in the United Digital Market

Eliza Stefanova<sup>[0000-0002-9697-2028]</sup>, Krassen Stefanov<sup>[0000-0002-4427-6098]</sup>

Faculty of Mathematics and Informatics, Sofia University St. Kliment Ohridski,  
5, James Bourchier Blvd., 1164, Sofia, Bulgaria  
eliza@fmi.uni-sofia.bg, krassen@fmi.uni-sofia.bg

**Abstract.** The European Open Science Cloud and the strategy for unified digital market are corner stones in the way of Europe's economy, industry, and society to the new digital era. In this paper, we will analyze how digital libraries can be best utilized for knowledge access and sharing in the new digital market.

**Keywords:** Open Access, Digital Libraries, Open Science Cloud, Research Data Alliance.

## 1 Introduction

European commission adopted the Digital market strategy on May 2015, updated 2021 (European Economic and Social Committee, 2021). The strategy has three main pillars: access; environment; economy and society.

The access pillar stated the need for better access to digital content and knowledge. The second pillar - environment, aims to allow such access, and includes digital libraries as the cornerstone for storing, sharing and disseminating the knowledge. The third pillar has as its main objective the creation of the European open science cloud.

The first 10-year digital agenda for Europe was set up in 2010 and it included as an integral part the digital libraries initiative. It was further improved through the digital single market strategy developed in 2015. In 2021, the strategy was complemented by the 10-year digital compass, which includes digital transformation of all business activities and digitalization of public services.

The digital libraries initiative includes one main objective: to make all cultural artefacts and scientific resources – books, journals, films, maps, photographs, music, *etc.* – accessible to all European citizens and preserve them for future generations. The initiative is focusing on cultural heritage through Europeana, and on access to scientific information through European Open Science Cloud (EOSC).

Open knowledge is one of the key concepts integrated under the umbrella concept the Open ecosystem, which includes other concepts like open science, open access, open research, open education and many more. All they advocate free access to all kind of results obtained from research projects and education activities at Universities and

all relevant scientific institutions. In fact, with the fast spreading of Internet, the traditional scientific publishing, used in centuries for dissemination of scientific results, is constantly replaced by new digital methods for dissemination through Internet. This new form of dissemination of scientific assets was named like Science 2.0, Open Research, and most recently as Open Science. The main goals for the Open science are to speed up the process of dissemination of research outcomes, to minimize the costs, to raise the quality and by shortening the path of every research outcome, to achieve vast speed up of the overall research productivity, and as a result significantly to shorten the path of innovations to society. This new trend is also beneficial for individual researchers, as they receive more visibility, with less costs and efforts.

There are different definitions of Open Science (Bartling & Friesike, 2014), (Open Access.nl, 2003), (Suber, 2012), but in this paper we will use the following one: “the dissemination of scientific knowledge that is as wide as possible, free of charge to all users, and accessible online”.

Open Science (Bartling & Friesike, 2014) includes other concepts such as Open Education, Open Software, Open Courseware and Open Research (having two important components: Open Data and Open Access).

Open Data (Open Knowledge Foundation, 2014), being major component of Open Science, is also freely available for everyone without any restrictions. One of the main features of Open Data is called FAIR (Findable, Accessible, Interoperable and Reusable) (FORCE11, 2016). While Open Data is FAIR data, not all FAIR data is Open Data. However, FAIR data could have clear and accessible data usage license.

Open Access (Open Access.nl, 2003), (Suber, 2012) means free, full and open online access to academic publications. There are two roads to publish scientific work with open access: golden road (in an open access journal) and green road (in a free public scientific repository).

Digital research repositories are already well established throughout many countries in the European Union. They become a permanent part of the scholarly communication and documentation research infrastructure.

Today all research institutions have developed their own digital repositories for storing and presenting their main research outcomes. The main trend is to increase the coverage of the existing digital repositories of published and unpublished textual research output, with new types of research outputs (e.g. images, video, and research data sets). Most of the European digital repositories (95%) support the Open Access (mainly through the Open Archive Initiative – OAI, <http://www.openarchives.org/>).

The majority of the EU countries have developed rich research infrastructures from digital repositories, where most of the research institutes has implemented one or more digital repositories containing research and educational resources. This means that research infrastructures from digital repositories are largely established throughout the European Union. It is also clear from the country reports that many initiatives are under way to deploy repositories to other institutes.

European Commission is very active in relation with Open Science. The first official plan was announced on April 19 2016 to create an “Open Science Cloud”. Later on the Amsterdam Call for Action on Open Science was published. As a result, EOSC was established in 2019 (Research and Innovation department of EC, 2019).

In Bulgaria, the Implementation plan for the National strategy for the development of the scientific research 2017-2030 (<https://www.strategy.bg/File-Handler.ashx?fileId=9594>) and the national plan for the development of Open Science Initiatives (Bulgarian Ministry of Education and Science, 2021), approved in January 2021, provide the background policy documents for the transition towards open science. The Plan defines the strategic goals, the roadmap for the next period (2021-2025) and the instruments to encourage Bulgarian researchers and research institutions to implement Open Science principles in practice. It sets the role of the Bulgarian Portal for Open Science, the institutional collaboration with EOSC and the further research infrastructures, enabling better visibility and integration of the Bulgarian science.

Sofia University is the main active player in Bulgaria in the field of Open Science. It is the only Bulgarian member of the EOSC association. Sofia University is also the only Research Data Alliance (RDA) node for Bulgaria. Sofia University has one of the biggest and richest digital libraries in Bulgaria and is actively participating in many research and development activities related to Open Science and digital libraries.

In the next chapter we present the main activities performed by the Bulgarian RDA node for promoting the Open Science culture. After that, we present the vision for the development of the European Open Science Cloud and implementation of Bulgarian Open Science Cloud (BOSC) as an integral part of the EOSC. We present technical details related to the design and implementation of Bulgarian Open Science Digital Library as the first and main ingredient of the BOSC. In the last chapter some new challenges and responsibilities of libraries as institutions and librarians are discussed.

## **2 Bulgarian RDA Node**

The Bulgarian RDA Node (RDA-BG) is the most active unit in the process of implementation of the EOSC in Bulgaria. It was established in 2020 to support and coordinate the participation of Bulgarian institutions, organisations and individual members in the Research Data Alliance.

The general objectives of the RDA-BG are to promote Open Research principles in Bulgaria, to raise awareness about RDA and Open Science activities and to support the data sharing culture. The RDA-BG is providing a national platform for collaboration and cooperation of Bulgarian researchers on:

- Promoting RDA principles among the Bulgarian research community, across scientific domains and within diverse research data organisations;
- Representing Bulgarian researchers in the RDA Interest groups and Working groups and governing structures;
- Raising awareness of RDA activities, events and funding calls while encouraging active involvement of new Bulgarian members in RDA activities;
- Promoting RDA's nationally relevant outputs, recommendations, and technical specifications in order to stimulate their adoption in the Bulgarian research environment;

- Implementing the policy of open science and open access to research data at a national level;
- Interacting with national research funding bodies and other relevant government officials to influence the implementation of open research data policy and digital research agendas;
- Promoting and supporting data management best practices, standards and solutions in the Bulgarian research area;
- Promoting data citation, sharing rewards and crediting best practices amongst Bulgarian publishers and stimulating the development of journal policies of research data deposit in connection to a paper publication.

The RDA-BG is coordinated by Sofia University, Faculty of Mathematics and Informatics. The RDA team includes researchers, actively involved and committed to promoting FAIR data principles and EOSC activities among Bulgarian institutions and the research community. The setup of the RDA BG Node was supported by RDA Europe project RDA Europe 4.0 Grant Agreement no 777388, community and structures.

Bulgarian RDA node implemented a number of activities and efforts aimed to increase awareness and involve more Bulgarian institutions and organisations in the activities of the EOSC and RDA.

Serving as a primary information desk for Bulgarian research institutions, researchers and industry members, the RDA BG Node regularly publishes at its website (<https://rda.bg>) up-to-date information about RDA and EOSC activities:

- Invitations for RDA webinars and conferences;
- Information about RDA open calls;
- Call for experts and participants in RDA Working groups;
- News about adopted EOSC documents and policy statements.

The RDA-BG mailing list is another major communication channel, with more than 60 active members receiving regular email newsletters on new developments and opportunities.

The RDA-BG team members are closely involved in the design and adoption of Open Science policy documents in Bulgaria, and contributed to structuring and adopting of the Bulgarian Open Science Cloud infrastructure. All team members took part in the promotion of the RDA outputs in Bulgaria and further connecting and involving Bulgarian researchers in other EOSC activities.

The RDA-BG team works in cooperation and actively contribute to the activities of other national institutions involved in Open Science, FAIR, and EOSC, such as:

- The NI4OS-Europe national dissemination event in Bulgaria was held online to bring together Open Science Cloud stakeholders in Bulgaria. During this event, main Open Science initiatives were presented: EOSC, NI4OS-Europe, Bulgarian Open Science Portal, National Science Program for ICT, RDA.bg, OpenAir.bg, and an overview of the Bulgarian activities in this field.
- RDA-BG hosted the event where the National Centre for Information and Documentation (NACID) presented the Bulgarian portal for open science at the Spring Science Session of the Faculty of mathematics and informatics at Sofia University.

- A documentary about Open science policies in Bulgaria and RDA was released by the RDA-BG staff.

The RDA-BG team members took part in several EOSC Working groups:

- Task Force Charter: Upskilling countries to engage in EOSC;
- Task Force Charter: Research Careers, Recognition, and Credit.

They presented their outcomes on the EOSC Symposium 2021.

As of today, the RDA-BG has about 34 registered members and about 60 associate members interested in the adoption and development of the FAIR, Open Science and Research Data Alliance initiatives.

The main goal now for the Bulgarian RDA node is to attract more potential members and to raise awareness focusing on regional centres and research partners in the country, promoting more actively RDA activities among Universities, Institutions and Industry partners, and especially gathering regional research communities.

One of the main aspects is to achieve a better level of advocacy and coordination activities with the new research infrastructures of the Centres of Excellence and Centres of Competence funded by Horizon 2020 and EU Structural Funds and focusing on the RDA activities impact on the innovation ecosystem, smart specialisation and regional growth.

In addition, RDA BG Node seeks to explore new partnerships and cooperation opportunities with other research and industry projects and initiatives in the regional context. It also aims to raise awareness and train more researchers, academic librarians, publishers and science communicators on applying the FAIR principles, RDM and Open science, thus supporting the reforms taking place in higher education and academia in Bulgaria.

### **3 Bulgarian Open Science Cloud**

At European level, the key initiative related to the progress and constant advance in the field of open science digital libraries is OpenAIRE (<https://www.openaire.eu/>). It provides standards, tools and services for Open Science implementation and to ensure their uptake on a global level, at least in Europe. OpenAIRE is one of the central initiatives behind the new ambitious plan of the EC to implement the so called European Open Science Cloud (Research and Innovation Department of EC, 2017), (Research and Innovation Department of EC, 2019), (About Plan S, 2018), (European Commission, 2018). The main goal is to move from Open Access to Open Science. In other words, to provide open access not just to publications, but to all types of scientific results, including Open/FAIR research data, open source software, open education including free educational resources, open services, open protocols, open methodologies. By linking together all possible research knowledge assets, to achieve the ultimate goal to open science to all European citizens.

Following the efforts of all other European scientists, in Bulgaria we also started to think and work on opening the Bulgarian science to the world (Bulgarian Ministry of Education and Science, 2021). The ultimate goal is to develop the so called Bulgarian

Open Science Cloud (BOSC), based on the same principles, standards and technologies and fully compliant with EOSC.

The current digital research libraries in Bulgaria are dispersed, not well integrated and using different data models and standards. Most of these libraries are not compliant with existing models and standards adopted by OpenAIRE and approved by EOSC. In fact, only three such digital libraries are listed as compliant.

The other important problem is related with adopted standards and practices from Bulgarian National Centre for Information and Documentation (NACID) (NACID, 2022), supporting the main general registries with all relevant scientific results available from all Bulgarian scientific organizations.

So, we developed new model for storing all research knowledge assets in current Bulgarian Open Science Digital Libraries (BOSDL), compliant with OpenAIRE and EOSC, and to design and implement various transition schemes, in order to transfer all available information from national registries supported by NACID, into new BOSC.

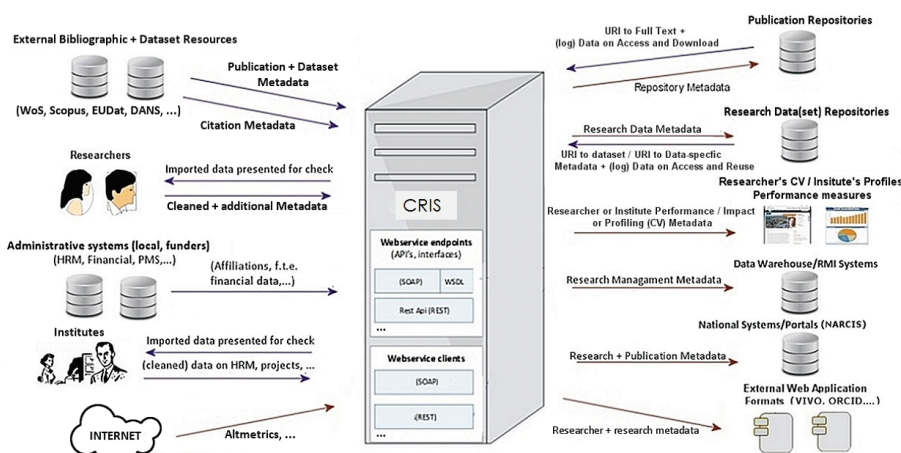
The main work was focused on the development of open research digital repositories, preserving research outcomes and assets, working in multilingual mode and storing the full assets (either text, data or programs), and following well established metadata standards from OpenAIRE initiative (<https://www.openaire.eu/>) and supported by EOSC.

The main portal and the necessary repository were designed and the relevant models for research asset description and storage were developed and implemented. The model chosen is in full compliance with models offered by OpenAIRE and CRIS (euroCRIS, 2002). This data model (Figure 1) relies on a set of basic entities as defined by the Common European Research Information Format (CERIF) model (euroCRIS, 2020) maintained by the non-profit organization euroCRIS.

On the base of this model, the first software prototype was developed, utilizing the open source system DSPACE-CRIS (4Science, 2016).

The existing registries of NACID were analysed and used to populate the prototype. The first version of automatic tool for extracting metadata from existing sources was developed. The information was further checked in the official public registries using DOI and ISBN, and also using the Crossref API (Crossref, 2020).

In addition, the available full text sources for all research assets were analysed from their metadata descriptions, and relevant automatic tools for their extraction and storage in the prototype of the BOSC were implemented.



**Fig. 1.** CRIS and CERIF model in action

A dedicated workflow for transferring entities from the Bulgarian Current Research Information System, supported by NACID into the Bulgarian Open Science Digital Library (BOSDL), was developed. This workflow was designed as a multi-stage process, similar to the process of extraction, transformation and loading (ETL) data in data warehouses. The multi-stage process designed consists of the following steps:

Step 1: Metadata harvesting. BOSRL objects are harvested using the REST API provided by NACID National Research Information System in JSON format and stored in a single MongoDB database organized in collections.

Step 2: Extraction and finding of persistent digital identifiers DOI and ORCID. First, we are extracting DOI identifiers from publications using regular expressions (Crossref, 2020). Then we find DOIs of publications referenced by a given reference (unstructured) using Search-Based Matching with Validation (SBMV) algorithm (CrossRef, 2019). On the base of available information, we find authors' ORCID identifier by personal names and DOI identifier of an own (author) publication (ORCID, 2017).

Step 3: Extracting and enriching bibliographic metadata and content. Each DOI is associated with bibliographic metadata about the object, including one or more URIs where the object can be found. Using the DOI resolution service (<http://dx.doi.org/>) bibliographic metadata for each object is retrieved and stored. Using the bibliographic metadata all URIs are extracted and crawled for pdf content, if any, and downloaded.

Step 4: Validation. In order to validate the metadata record, we need to match the bibliographic metadata from DOI resolution service and source data from BOSRL. This step ensure that the extracted and found DOI identifiers are relevant for the corresponding documents. The match is performed by measuring the Levenshtein distance (Levenshtein, 1966) between the source and the target with pre-set threshold for measuring the similarity between two strings.

Step 5: Transformation. In contrast to the ingest of flat metadata formats like Dublin Core, the import of CRIS objects requires a denormalization of the entities in order to add to a given entity (e.g. publication) the properties of related entities (e.g. information from the person entity). Spreadsheets are generated that reflects the structure of the

DSpace-CRIS data model. Finally, the content files and metadata spreadsheets are transformed into a Simple Archive Format Package (DSpace, 2019) for batch import to the BOSDL repository.

The current version of BOSC portal is maintained by the National Centre for Information and Documentation NACID and is in compliance with Commission recommendation (EU) 2018/790 of 25 April 2018 on access to and preservation of scientific information.

#### **4 New Challenges and Responsibilities for Libraries**

In these new realities, libraries and librarians have new challenges and responsibilities. Digital libraries should serve the role of digital hubs providing a rich set of services to their clients. These services include high speed Internet access, training and education, access to cultural heritage and scientific resources all over the world.

In order to make this happen, they need to satisfy several requirements:

- to provide support for rich set of standards for presenting and sharing of digital resources
- to have much more powerful computing devices and much more flexible computer infrastructure, in order to react on time on the rapid development of IT products and services;
- to have much more IT-trained personal in order to guarantee full time access with high quality and zero time service interruption;
- to be able to transfer as fast as possible all their normal library assets in best possible digital representation formats.

The new digital market requires from libraries new business models for guaranteeing sustainability and self-efficacy. This means that the roles, numbers and competencies of librarian staff should be adjusted in order to meet all these requirements and challenges.

As a consequence, this means that education and training of library staff should change dramatically in order for libraries to answer to all new challenges.

That applies especially, when we are speaking about university libraries. They need to provide access to scientific knowledge, performing new for them functions.

These functions appear because of the need qualified specialists to maintain the data and metadata in institutional repositories with open resources. At that point, the librarians could assist scientists in professional manner in the process of providing research results and the data on which they are derived.

In addition, our observations at Sofia University show that in some cases researchers do not publish their research in the institutional repository because of uncertainty related to license agreements. Usually, the researchers themselves have a lot of questions: *Which open access repositories can I publish to? With what license to publish in open access repositories? What can I publish in open access repositories?* Researchers expect the librarians to be able to provide them the consultations regarding sharing in open data repositories depending on license agreement. With accumulated expertise



librarians could provide also trainings after which more researchers to increasing publication competence will be comfortable to publish their work with open access.

Through these new roles, libraries contribute significantly to open access of the society and the business to the knowledge provided directly from the source. They become mediator between science and society.

## 5 Conclusions

Technological changes move the society in new directions. In that paper we presented contemporary developments in Bulgaria related to the open science, looking on it not just from technological point of view, but also from organizational and institutional side. We demonstrated work done in the direction of open access to the research and data in the country, but a lot of work is ahead – at the personal and institutional level. The most difficult part of it is in the manner of work, culture and thinking.

## References

- 4Science. (2016). *DSPACE-CRIS*. Retrieved from <https://dspace-cris.4science.it/>.
- About Plan S. (2018). *Plan S: Making full and immediate Open Access a Reality*. Retrieved from <https://www.coalition-s.org/>.
- Bartling, S., & Friesike, S. (2014). *Opening Science – The Evolving Guide on How the Internet is Changing Research, Collaboration and Scholarly Publishing*. Heidelberg, Germany: Springer. Retrieved July 16, 2022, from <http://www.openingscience.org/get-the-book/>
- Bulgarian Ministry of Education and Science. (2021). *Bulgarian national plan for the development of Open Science Initiatives*. Retrieved from [https://www.mon.bg/upload/24848/plan-otvorena-nauka\\_130121.pdf](https://www.mon.bg/upload/24848/plan-otvorena-nauka_130121.pdf)
- Bulgarian Ministry of Education and Science. (2021). *Bulgarian Open Science Initiative*. Retrieved from <https://rda.bg/en/node/47>.
- CrossRef. (2019). *Search-Based Matching with Validation (SBMV) algorithm*. Retrieved from <https://github.com/CrossRef/search-based-reference-matcher>.
- Crossref. (2020). *Crossref REST API*. Retrieved from <https://www.crossref.org/services/metadata-delivery/rest-api/>.
- DSpace. (2019). *Simple Archive Format Package*. Retrieved from <https://github.com/DSpace-Labs/SAFBuilder>.
- euroCRIS. (2002). *CRIS - data model used for the definition of research information systems*. Retrieved from <https://www.eurocris.org/>.
- euroCRIS. (2020). *Common European Research Information Format (CERIF)*. Retrieved from <https://www.eurocris.org/cerif/main-features-cerif>.
- European Commission. (2018). *The European Commission launched the European Open Science Cloud (EOSC) in Vienna*. Retrieved from <https://eosc-launch.eu/home/>.
- European Economic and Social Committee. (2021, April 17). *The digital single market - trends and opportunities for SMEs (own-initiative opinion)*. Retrieved from

<https://www.eesc.europa.eu/en/our-work/opinions-information-reports/opinions/digital-single-market-trends-and-opportunities-smes-own-initiative-opinion>

- FORCE11. (2016). The FAIR Data Principles. Retrieved from <https://www.force11.org/group/fairgroup/fairprinciples>
- Levenshtein, V. I. (1966). Binary codes capable of correct-ing deletions, insertions, and reversals. *Soviet Physics Doklady*, 10(8), 707–710.
- NACID. (2022). *Bulgarian National Centre for Information and Documentation*. Retrieved from <http://nacid.bg/bg/>.
- Open Access.nl. (2003). *What is open access?* Retrieved July 13, 2022, from <http://www.openaccess.nl/en/what-is-open-access>
- Open Knowledge Foundation. (2014). *Open Data definition*. Retrieved from <https://okfn.org/opendata/>.
- ORCID. (2017). *ORCID Public API V 2.1*. Retrieved from [https://pub.orcid.org/v2.0/#!/Public\\_API\\_v2.1/searchByQueryV21](https://pub.orcid.org/v2.0/#!/Public_API_v2.1/searchByQueryV21).
- Research and Innovation Department of EC. (2017). *EOSC Declaration*. Retrieved from <https://eosc-launch.eu/declaration/>
- Research and Innovation Department of EC. (2019). *European Open Science Cloud*. Retrieved from [https://ec.europa.eu/info/publications/european-open-science-cloud-eosc-strategic-implementation-plan\\_en](https://ec.europa.eu/info/publications/european-open-science-cloud-eosc-strategic-implementation-plan_en).
- Suber, P. (2012). *Open access*. Cambridge: The MIT press. Retrieved from <http://library.oapen.org/handle/20.500.12657/26065>

Received: July 05, 2022

Reviewed: July 20, 2022

Finally Accepted: July 30, 2022