

Bulgarian Open Science Cloud

Peter Stanchev^{1,2}, Krassen Stefanov³

¹ Institute of Mathematics and Informatics, BAS, Bulgaria

² Kettering University, Flint, USA

³ Faculty of Mathematics and Informatics, SU, Bulgaria

pstanche@kettering.edu, krassen@fmi.uni-sofia.bg

Abstract. The paper presents the current state of Building the Bulgarian Open Science Cloud (BOSC).

Keywords: Open Science, European Open Science Cloud, Bulgarian Open Science Cloud

1 Introduction

With the huge explosion of Internet communications, the way people perform research activities has changed dramatically. Now researchers can have constant access to research outcomes of other researchers in real time. This increases the effectiveness and the speed of all new research outcomes. But with the explosion of all big research data available around the globe, it is of critical importance for people to have access to the most relevant data and information in the fastest possible way. And here we see the role and the purpose of the European open Science Cloud (EOSC) initiative (Communication from the commission..., 2016). The EOSC should be such an infrastructure, which will make it possible for all researchers to share and find all the relevant research information as fast as possible, and to support and enable researchers and different research teams to join their efforts in solving complex and difficult problems.

What should be involved in such an infrastructure? First of all, research literature – papers, reports and books. So, this infrastructure should unite all digital repositories with science literature into one big super library. But this is not all. The current research experiments usually involve huge datasets that cost a lot of time and efforts to collect and prepare. So, these datasets are also very valuable assets that need to be collected, shared and accessed by all researchers.

In order to process such huge datasets, researchers create modern and very productive research infrastructures, like clusters for high performance computing. And it is very important that all such infrastructures are also part of the EOSC, and are used and shared between all researchers. Most research teams are specialized in providing specific services. And when very complex problems need to be solved, usually we need to combine the expertise of several distinctive research groups, combining the services

they offer. So, all such services are also very valuable assets, which need to be provided by the new EOSC infrastructure.

EC started to work on building the EOSC several years ago initiative (Communication from the commission..., 2016), with first specific consultations starting in 2016 (Realizing the European open science cloud..., 2016), followed by formulating its goals, main building principles (EOSC Declaration, 2017), and roadmaps for its development (Commission Staff Working Document..., 2018), adopted in March 2018.

2 Building Bulgarian Open Science Cloud (BOSC)

The Executive Board of the EOSC (Setting up the Expert Group..., 2018) has been established by the Commission Decision in August 2018. In November 2018, the European Commission launched the European Open Science Cloud (EOSC) in Vienna¹. The EOSC envisions establishing a European data infrastructure, integrating high-capacity cloud solutions, eventually widening the scope of the services described in Figure 1.

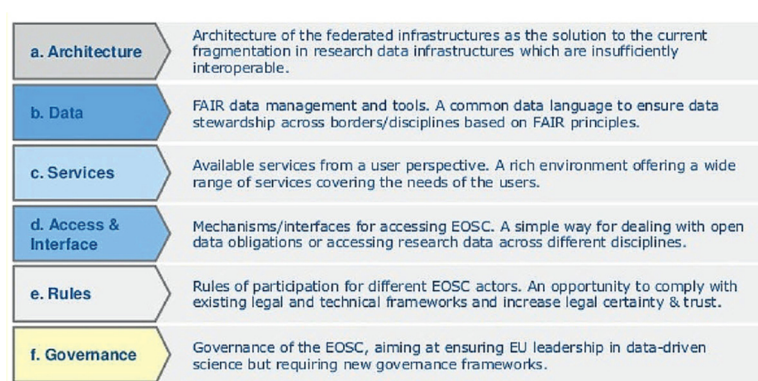


Fig. 1. EOSC Model

(Source: <http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud>)

In the beginning of 2018, the first idea related to the development of the Bulgarian Open Science Cloud (BOSC) was formulated during the Bulgarian Presidency of the Council of the European Union, adopting the principles of Open Access, Open Science and Open Education. As a consequence, a new national scientific program was initiated (“ICT for science, education and security” - ICT_SES²) with one working package “Open Science for Society” dedicated to the development of new methods and the elab-

¹ November 2018, the European Commission launched the European Open Science Cloud (EOSC) in Vienna: <https://eosc-launch.eu/home/>

² Available at: <https://npict.bg/>

eration of standards for facilitating open access to research data. In this paper we present the first results from this new scientific research program targeting the development of the BOSC, which will be based on the principles, standards, technologies and architecture of the European cloud (EOSC), will be fully compatible and an integral part of it. The researchers from ICT_SES are in a constant dialogue with the Ministry of Education and Science (MES, <http://mon.bg/en/100000>), the Bulgarian National Science Foundation (BNSF, <https://www.fnf.bg/>) and the National Centre for Information and Documentation (NACID, <http://nacid.bg/bg/>) in order to identify the user needs for functional specifications for BOSC and the optimal model for its implementation. Access will be available through an entry point, a portal, via a web-based interface. This portal, following the principles of EOSC, will bring together access to the different entry points of the different scientific infrastructures in Bulgaria, enabling users to access the individual resources offered by BOSC. This portal will automatically communicate with individual data and resource servers and will derive the necessary metadata to provide services to users, all of which will be communicated through selected EOSC protocols and standards and OpenAIRE (<https://www.openaire.eu/>) standards.

The basic principles of building BOSC include:

- Building an architecture for unified research infrastructures;
- Providing data based on FAIR principles (accessible, interoperable and reusable) (<https://www.force11.org/group/fairgroup/fairprinciples/>);
- Provisioning of services that meet the needs of the consumers;
- Using an interface to work with data in different disciplines;
- Creating rules for participation in the BOSC.

BOSC should be a federation of existing and new research infrastructures, adding a layer to connect them and turn them into a Bulgarian research infrastructure. BOSC will essentially include a single gateway and a variety of pooled infrastructure research data committed to providing services as part of the EOSC.

The BOSC Data will lead to:

- Development of a better culture of management of research data and practical skills among scientists and innovators in Bulgaria;
- Development tools, specifications, catalogues and data standards for FAIR (accessible, interoperable and reusable) data (FAIR principles, n.d.), (Turning FAIR into reality..., 2018) within a secure and reliable environment, as well as supply-side services to assist scientists and innovators;
- Stimulating demand for FAIR data (FAIR principles, n.d.), (Turning FAIR into reality..., 2018) through incentives to find data in Bulgaria.

The BOSC General resources will cover:

- Finding catalogues of data/services and metadata;
- Interacting with compliant standards and common metadata;
- Reusing common intellectual property rights and legal provisions.

BOSC Services will offer the following five main types of services for Bulgarian researchers:

- Unique identification and authentication service, as well as access point and routing system to the BOSC resources;

- Secure and personalized work environment/space (e.g. diary, settings, compliance documents and unresolved issues);
- Access to relevant service information like BOSC status, list of unified data infrastructures, policy-related information, description of the compliance framework, and specific guidance;
- Services for finding, accessing, reusing and analyzing research data generated by others, available through appropriate data set catalogues and data services (e.g. analysis, synthesis, processing);
- FAIR Data Conversion services, for storing and securing long-term storage.

3 The First Prototype of Building Bulgarian Open Science Cloud (BOSC)

During the first six months of the execution of the national research program ICT_SES, researchers from the leading partners (Faculty of Mathematics and Informatics at Sofia University “St. Kliment Ohridski” and Institute of Mathematics and Informatics at the Bulgarian Academy of Science), together with all other project partners, started to work on the first prototype of the new BOSC portal (<https://cris.fmi.uni-sofia.bg/>), together with several important digital repositories and services.

In the beginning of 2019 the overall vision for the structure and functionalities of BOSC were formulated. This vision was based on the provided expert opinions, the first specification for the new portal for BOSC was developed and further used from NACID.

Researchers together with policy experts visited several EC funded meetings in order to enable the application of all relevant EC standards in a harmonized way. The overall goal was to enable as easily as possible the integration of the available data (even not in the needed format) from all existing repositories and registries and make it open and easily accessible.

The experience of other relevant research activities, related to the integration of research infrastructures, was also taken into account. On the base of all such meetings, investigations and experience, the first specific plan for the implementation of BOSC was developed.

The main work was focused on the development of open research digital repositories, preserving research outcomes and assets, working in multilingual mode, storing the full assets (either text, data or programs), following the well-established metadata standards from OpenAIRE initiative (<https://www.openaire.eu>), 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 CRIS³. This data model relies on a set of basic entities as defined by the Common European Research

³ CRIS - data model used for the definition of research information systems: <https://www.eurocris.org/>

Information Format (CERIF) model (<https://www.eurocris.org/cerif/main-features-cerif>), maintained by the non-profit organization euroCRIS⁴.

On the base of this model, the first software prototype was developed, utilizing the open source system DSPACE-CRIS (<https://dspace-cris.4science.it/>; <https://wiki.duraspace.org/display/DSPACECRIS/Dspace-CRIS+Home>).

The existing registries of NACID were analyzed and used to populate the prototype. The first version of an 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 (<https://www.crossref.org/services/metadata-delivery/rest-api/>).

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

4 Conclusion

The first prototype can be accessed at: <https://cris.fmi.uni-sofia.bg/global-search>.

Five of the project partners' repositories were implemented, while all of the rest are under development and should be ready till the end of 2019. In such a way during the first year of the program ICT_SES we will have a proof-of-concept prototype for the BOSC, which will be further developed and improved in the next two years of the program. The portal will be fully compatible with the registries supported by NACID and will involve all research outputs coming from projects funded by MES and BNSF.

References

Commission Staff Working Document - Implementation Roadmap for the European Open Science Cloud, SWD(2018) 83 final (2018). Available at: <http://ec.europa.eu/transparency/regdoc/rep/10102/2018/EN/SWD-2018-83-F1-EN-MAIN-PART-1.PDF>

Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions: European Cloud Initiative - Building a competitive data and knowledge economy in Europe, COM(2016) 178 final (2016). Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1555074889405&uri=CELEX:52016DC0178>

EOSC Declaration (2017). Available at: https://ec.europa.eu/research/open-science/pdf/eosc_declaration.pdf#view=fit&pagemode=none

FAIR principles, (n.d.). Available at: <https://www.force11.org/group/fairgroup/fair-principles>

Realizing the European open science cloud. First report and recommendations of the Commission high level expert group on the European open science cloud (2016).

⁴ Available at: <https://www.eurocris.org/>

Available at: <https://publications.europa.eu/en/publication-detail/-/publication/2ec2eced-9ac5-11e6-868c-01aa75ed71a1/language-en/format-PDF/source-89516661>

Setting up the Expert Group - Executive Board of the EOSC and laying down rules for its financing C(2018) 5552 final (2018). Available at: <https://www.eosc-portal.eu/sites/default/files/C20185552-EC-DECISION-EOSC-Excecutive-Board.pdf>

Turning FAIR into reality. Final report and action plan from the European Commission expert group on FAIR data (2018). Available at: <https://publications.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF/source-89516642>

Received: June 25, 2019

Reviewed: July 10, 2019

Finally Accepted: July 20, 2019