Workshop Report[1]
ENLIGHT-RISE Workshop on Digital Research Infrastructures

7 June 2022
Link to Workshop Site here
The workshop was recorded and the presentations are available on YouTube.
The links to the videos are attached to the workshop agenda below.

Introduction

Research increasingly depends on advanced and interconnected digital research infrastructures (DRIs) - to collect and store, process and share data. ENLIGHT partners engage in various ways in stepping up their capacities in this area: through own infrastructures, staffing and integration into national, European and global networks.

One big ship on the horizon is the European Open Science Cloud (EOSC), an endeavour that mobilizes developments and commitments through European projects and alliances as well as institutional, national and regional contributions. The workshop focused on inroads into the EOSC from different perspectives such as University´s perspective but also from national or large European infrastructures.

The purpose of the workshop was to provide an overview of the EOSC ecosystem and digital infrastructures that are or will be available, and to explore a range of approaches, "inroads", put into practice at ENLIGHT partner universities and their steps towards integration into the EOSC. As this report will outline, a lot of this work is still in progress and it is a complex task in a dynamic environment.

Workshop Participants

The workshop was attended by more than 60 participants, among them research infrastructure managers, librarians and researchers who are engaged in setting up infrastructure services within their institutions and through collaborations or who wanted to get an overview of current developments.

Agenda of the Workshop

9.00-9.15 **Welcome & introduction** - Birgit Schmidt, University of Göttingen; Margaretha Andersson, University of Uppsala
(online: https://www.youtube.com/watch?v=Rb4cNR2fYZA)

9.15-11.15 **Digital research infrastructures and the EOSC**

- **Opportunities to engage with the EOSC** - Sarah Jones, GÉANT, EOSC-AISBL Board of Directors Member
  (online: https://www.youtube.com/watch?v=Rb4cNR2fYZA)
- **Inroads to the EOSC from the SSH perspective** - Sally Chambers, University of Gent, DARIAH, SSHOC
  (online: https://www.youtube.com/watch?v=v3kk9Nz1HPY)

---

[1] Report written by Dr. Merle Schatz (UGOE).

10.15-10.30 Break

- **Inroads to the EOSC from the Life Sciences perspective** - Björn Nystedt, Uppsala University, ELIXIR Sweden
  (online: https://www.youtube.com/watch?v=rzuzBL19Krs)
- **A perspective on national research infrastructures and their links to the EOSC** - Regine Stein, University of Göttingen
  (online: https://www.youtube.com/watch?v=JKbC8uhkqcs)
- **Inroads to the EOSC through OpenAIRE** - Inge van Nieuwerburgh, University of Gent
  (online: https://www.youtube.com/watch?v=7j3zl1o446Y)

11.20-12.00 **Panel   discussion   with   speakers,   chaired   by   Hans   Karlsson**
(online: https://www.youtube.com/watch?v=DVnGvV5AAVY)

- What are the main benefits of the EOSC from the user and/or the provider perspective?
- What type of skills and training are needed to use the EOSC?
- What is the next big challenge for the EOSC from your perspective?
- What would be the most practical thing EOSC could deliver to you from your perspective, e.g. searching across datasets?

12.00 **Closing**

Workshop Report

Birgit Schmidt, University of Göttingen, and Margaretha Andersson, University of Uppsala, jointly organized and chaired the workshop. The workshop was opened with an introductory presentation by Birgit Schmidt on ENLIGHT and ENLIGHT RISE as its´ satellite project that aims at creating a joint research and innovation agenda with focus on OS incentives, connecting and sharing of digital research infrastructures, civil society engagement tools, and early career development instruments.

Sarah Jones, GÉANT, EOSC-AISBL Board of Directors Member, followed with an introduction on how the EOSC infrastructure and governance is conceptualized, what is currently available and what is under development, including its infrastructural connections to institutional, national and disciplinary resources and services. The EOSC Association, the European Commission (EC) and the EU member states are committed to Open data and FAIR principles as central aspects in their policies. Against this background, the audience of the workshop was walked through the challenging goal of making outputs of research - such as data, source code, methodologies, etc. - not only available for (re)use but also to connect them in an open way within the EOSC. The aim is to integrate the different repositories, infrastructures and services that are available in all the different member states via one platform. The FAIR data principles connect these data and services. However, with data sharing and infrastructural connecting processes, questions regarding political agreements and activities as well as questions regarding the development of EOSC governance and legal structure need to be solved.
The EOSC portal (https://eosc-portal.eu) is the main gateway to navigate resources and services integrated into the EOSC. Other core functionalities of the EOSC platform are still under development, e.g. federated identity mechanisms and search across data repositories. There are still many challenges in terms of long-term data availability, in particular if they are based on project funding. Regarding the EOSC, the goal is to achieve sustainability through a

shift from a project model to a procurement model, which covers storage and other services to any researcher.

Having learned about the complex EOSC ecosystem and governance structures the following presentations focused on inroads into the EOSC from disciplinary perspectives. **Sally Chambers, University of Gent,** introduced inroads to the EOSC from the Social Science and Humanities Perspective using the SHHOC Project as an example.

One of the main deliverables from the project is the SSH Open Market place. It feeds also into the EOSC market place with tools and services, training materials, publications, datasets and workflows. A demonstration of the interrelationship between the different types of resources from data to workflow within the SSHOC was given to then explain how this can be connected with EOSC via the onboarding process: 1st step: joining EOSC as a provider, EOSC has a dashboard for that. 2nd step: follow the instructions to onboard resources into the EOSC catalogue.

Sally Chambers then went on explaining, that the true value of this is the interconnection with the various core services or more generic services, for example the EOSC authentication and authorization (AAI), Monitoring, Help-desk, accounting for services. Therefore, not the onboarding of the sources into EOSC alone, but additionally the subsequent connection within the EOSC ecosystem is a very important part of the process.

The relevance of SSH to participate in EOSC at all was explained: it helps magnify the voice at the highest levels of European policy formation and makes sure, that the needs are recognized with the EOSC. One essential factor she came to talk about in this context is the relevance of cultural heritage data from archives, museums and libraries for humanities research, and the necessity to interconnect with and within the research data ecosystem. A European data space could be a valuable way to do that, and indeed, in November 2021 the European data space for cultural heritage was announced. Sally Chambers emphasised, that libraries, archives and museums are very important EOSC providers as well.

A question from the audience on how the very different structures of EOSC should be connected, how collaboration should be enabled was raised. Here, several plans and initiatives are in place and are being developed: Right now, national level analysis is made to see what can be compared and shared. Challenges are, that the data needs to be made machine readable, additionally there are cross border difficulties regarding questions of copyright. Also training needs to be offered for the people who are involved with onboarding. Within the SHHOC marketplace-collection as one example, there is a data initiative to build a workflow that helps cultural heritage institutions to be able to publish their data as datasets as well. Europeana will be the foundation for the European data space for cultural heritage data.

Björn Nystedt, Uppsala University, followed by presenting inroads to the EOSC from the Life Sciences perspective by introducing ELIXIR, the national digital research infrastructure managing data from life science organisations in Europe. With regard to the issue, that a lot of data that needs to be handled and used can be a burden, he suggested that digital research infrastructure is indeed a part of the solution when looking at the complex data landscape. ELIXIR has five technical platforms for computer, data, tools, interoperability and training (all this is already part of EOSC). ELIXIR is also interacting with other initiatives in Europe, and additionally the support of the community is an important factor when it comes to data production, data sharing, data storage. ELIXIR is investing here in community upskilling by offering e.g. training, courses, and mentorship programs.

The purposes of ELIXR and EOSC are quite similar, there is a lot in place, the challenge is the mapping exercise, because ELIXIR is differently structured, funded and governed than EOSC. One important question is, how both can be made compatible. Björn Nystedt elaborated more on this challenge by raising related topics such as:

- When sensitive data is being generated there is no obvious way how it can be offered for research. Trying to integrate this into a legal and technical framework is challenging. And to bring it from national level to European level is even more difficult.
- One way to do this is to go through a federation with shared metadata: FAIR principles and secure and sustained. There is the need to find a funding model that allows all this.
- Challenges and opportunities: linking EOSC to existing efforts is key; the fragmented landscape with many stakeholders and initiatives is a challenge; competence build-up is central; access only is not enough, to make data reusable, annotation/ metadata/ documentation is crucial; and ethical and legal issues need to be balanced with community value and with personal risk/integrity.

A question was raised regarding legal issues when working with sensitive data. So far, there are no task forces in EOSC for that, but it is crucial. For sharing data, the questions regarding legal challenges across borders need to be addressed. If not harmonizing it, then at least monitoring it would be important for understanding what at least is possible. This is a difficult task, since data sharing between different institutions only on national level already is very difficult.

Regine Stein, University of Göttingen, presented her perspective on national research infrastructures and their links to EOSC. The Göttingen State and University Library is active in national level initiatives like the German Initiative for Network Information (DINI), DARIAH-DE, NFDI, Open-Access.network, on the European level with projects such as DARIAH ERIC, OpenAIRE, EOSC Future, FAIRsFAIR, and on international level with the Research Data Alliance (RDA), and the Conferation of Open Access Repositories (COAR).

The German National Research Data Infrastructure (NFDI) consists of up to 30 consortia. The consortia are connected and need to address common infrastructures of the following sections:

a) Section-infra: identity and access management, persistent identifiers (PID), research software;

b) section metadata: (meta)data terminologies and provenance;

c) section ELSA: ethical, legal and social aspects;

d) section edutrain: training and education.

The collaboration of the consortia is one aim. The NFDI association acts as mandated organisation representing the German research data community in the EOSC Association. Right now, there are four humanities-related NFDI initiatives: NFDI4Culture; Text+, NFDI4Memory, and NFDI4Objects. Connecting cultural heritage data and research data is one main focus point for the four initiatives.

Regine Stein then elaborated further on the example of Text+ which is a consortium for text and language-based research data. The core of the research data strategy here are both, the data domains and thematic clusters. The three data domains collections, lexical resources, and editions bundle activities related to specific subtypes of data and research methods in one data domain. Each of these clusters consists of specialized data and competence centres and offer services which build on generic services provided by the infrastructure operations task. The close involvement of the research communities in the development of the research

infrastructure is part of developing these infrastructures. Infrastructure components and services such as AAI, PID, repositories, storage, computing resources, data domains and thematic clusters, and collections are distributed. Also, there is a cross domain connectivity such as standardization activities, software activities, community activities to address the domain and cluster specific work and to contribute to cross domain work.

The "infrastructure operations task area" is the main glue to provide cross-data domain generic services that foster interoperability and connects Text+ to NFDI and also to the EOSC. Regine Stein then looked deeper into the connection to EOSC: The main interaction from the German perspective lies between the NFDI and EOSC. EOSC has done a lot of ground work for the development of federate service infrastructure which resulted in the EOSC interoperability framework. The task now is to adopt these existing components which consist of standards, policies and procedures and best practices. Linking the structures is a complex task in a dynamic environment since the experience proves, that connecting the own communities only, already is quite a challenge.

Inge van Nieuwerburgh, University of Gent, talked about inroads into the EOSC through OpenAIRE and explained, how the connection to OpenAIRE can help to link to EOSC, because OpenAIRE can be one way for universities and other institutions to become a provider for EOSC.

OpenAIRE is a membership organization with the mission to shift scholarly communication towards openness and transparency and facilitate innovative ways to communicate and monitor research outputs and their links and impact. The focus therefore lies on scholarly communication and on how to disseminate research knowledge. There are three pillars of action: services, training and policies.

Showing the OpenAIRE Research Graph she explained the interoperability and connection of networks and data providers, as well as the sources that are linked to them and that can be used by researchers. The OpenAIRE Research Graph includes metadata and links between scientific products (e.g. literature, datasets, software, and "other research products"), organizations, funders, funding streams, projects, communities, and (provenance) data sources.

A practical example of the Research Graph was given by looking at a publication repository. Information that is available in the research graph shall automatically go into services like the EOSC, here, again, the challenge is to get the infrastructures connected. But there are already several connections between OpenAIRE and EOSC: OpenAIRE operates at the national-EU-International levels, with federated activities in three areas:

a) Policy alignment – supporting organisations and countries setting up fit-for-purpose policies that serve local needs;

b) Support and Training - utilising the network to provide hands-on support and educate researchers how to practice Open Science. For example, there is training on how to get valuable information in the EOSC and work with it. In the community of practice, a lot of best practices are exchanged. RDA is present as well as SSHOC, ELIXIR, and others;

c) Service Provision - providing the services and tools to all actors in the research ecosystem to practice and monitor Open Science. The Services shall help researchers and institutions to comply with EOSC principles and recommendations for end users and service providers on Openness, FAIRness and the EOSC architecture.

Additionally, there is the EOSC observatory: The EOSC association, the EU Commission, the member states observe how things are evolving, and how member states infrastructures and EOSC association are developing the EOSC together.

At the end of her talk, Inge van Nieuwerburgh recommended the presentations of the EOSC provider days that addressed topics on how to become an EOSC provider, and she emphasized the essential role of libraries as providers, repositories as providers, research projects as providers, HPC centres as providers.

A question came up regarding publications since so far EOSC is rather data focused. At the moment there is good infrastructure for research data and open access to publication, however, all this needs to be closer aligned and interconnected, such that both infrastructures can be used seamlessly. A researcher should be able to go to a publication and access the data related to it.

Panel Discussion

In the following panel discussion, **chaired by Hans Karlsson**, University of Uppsala, the speakers addressed questions from the chat and discussed topics such as
- What does the EOSC currently offer to researchers and research infrastructure providers?
- What is the current state and are the next steps of developing federated infrastructures in the EOSC, and how can institutions engage?
- What are concrete steps institutions can take now towards integrating (some of) their DRIs into the EOSC? Important steps to be "EOSC-ready"?
- In the context of disciplinary and generic DRIs: What has been achieved and what are current hurdles to use and/or contribute services by/to the EOSC?

There were questions raised among the topic of research data on request, that is linked to necessity of communication which can delay the process of getting data. Still not everything is "open" but most funders push more for simple ways of availability.

Researchers are the main users, and from their perspective, making things globally findable is important and that data is re-used. But also discovering the possibilities and potentials of an EOSC can encourage their working together and sharing data by using a non-domain specific way.

Another topic discussed was if EOSC, so far European, can be expected to open to other countries as well. Here already first activities are in place: There was a DARIAH beyond Europe workshops, e.g. US, Australia, and is active in RDA. Research doesn't stop with European borders, and there are already a lot of interconnecting activities, but cross-border sharing of data has legal and even ethical challenges.

The different challenges and the aim to build a system that works easily for the users require different forms of training. The specific needs of the users must be identified in order to then provide the trainings and support. FAIR data is for many researchers a new concept. Ideally, EOSC must become a natural component at an early stage, therefore it must be implemented into every digital literacy pathway, and this should start early, students need to be trained from Bachelor onwards. The aim is to develop EOSC in a way, that training at the end shall not be necessary - but available in case of need. Other voices stated, that training is important and must be sustainable, sufficient staff is needed. It is, as one idea raised in the discussion,

possible to establish a nation-wide or European-scale training hub, e.g. through ELIXIR for the life sciences

Regarding the data, in particular full text, questions around copyright remain a huge challenge. And there is still quite a challenge on interoperability and the linkage of and within the national infrastructures, time is needed to solve these issues. The topic of sustainability was highlighted again, in addition to the efforts already underway, EOSC is one more thing on top of everything else, sustainability must be considered structurally during the development. Additionally, it needs to become easier to submit datasets to repositories, etc.
There is a need for policies for all the data domains and thematic clusters, but the accomplishment of all this on the different layers is a task and a challenge for research infrastructures as a network. And researchers could not like to move data, and again here: sharing data might be different to a Chemist or a humanities researcher, how does all this meet the aim of open accessibility?

The discussion closed with remarks on legal issues that are sometimes underestimated with regard to sharing research data. A single authentication and authorization mechanism across Europe would be a real milestone. But not one single place is the solution, a federated solution is better, that is competence centres distributed across Europe are needed. These competence centres could be virtual, e.g. providing services and support for the analysis of cultural heritage data, OCR, sharing algorithms, sharing outcome data, etc. This also involves questions of lobbying for the sharing of research data - in any case, for all these areas more legal expertise is needed as well.

Links Provided in the Workshop´s Chat
(DARIAH-EU) https://enlight-eu.org/index.php/university-about-us/news-events/158-news/567-rise-workshop-on-digital-research-infrastructures#agenda

Some cultural heritage related links:
https://collectionsasdata.github.io
https://data.bnl.lu
https://data.nls.uk
Jupyter notebooks and cultural heritage:
 https://glamlabs.io/computational-access-to-digital-collections/

OpenAIRE
About OpenAIRE: https://www.openaire.eu/about
OpenAIRE Research Graph: https://graph.openaire.eu/
OpenAIRE & EOSC: https://www.openaire.eu/openaire-and-eosc
OpenAIRE catalogue: http://catalogue.openaire.eu/home
OpenAIRE guidelines: https://guidelines.openaire.eu/en/latest/
OpenAIRE monitor: https://monitor.openaire.eu/dashboard/ec
Community of Practice training coordinators: https://www.openaire.eu/cop-training
EOSC observatory: https://eoscobservatory.eosc-portal.eu/home
EOSC Future: http://EOSCfuture.eu
EOSC future provider days: https://eoscfuture.eu/eventsfuture/provider-days/

OpenAIRE is also investing in an open platform for training, Open Plato:
https://openplato.eu/

Publications
- Engelhardt, C. et al. (2022). *How to be FAIR with your data: A teaching and training handbook for higher education institutions.* Göttingen University Press, Göttingen. https://doi.org/10.17875/gup2022-1915
- OECD (2020). Building digital workforce capacity and skills for data-intensive science. *OECD Science, Technology and Industry Policy Papers*, No. 90, OECD Publishing, Paris, https://doi.org/10.1787/e08aa3bb-en. https://doi.org/10.1787/e08aa3bb-en

Regarding Training
TeSS: ELIXIR's Training Portal, https://tess.elixir-europe.org/