Social Sciences and Humanities Pathway Towards the European Open Science Cloud

Suzanne Dumouchel, Francesca Di Donato, Monica Monachini,
Yoann Moranville, Stefanie Pohle, Maria Eskevich
CNRS, Net7, CNR, DARIAH, MWS, CLARIN ERIC
suzanne.dumouchel@huma-num.fr, didonato@netseven.it,
monica.monachini@ilc.cnr.it, yoann.moranville@dariah.eu,
Pohle@maxweberstiftung.de, maria@clarin.eu

Abstract

The paper describes a journey which starts from various social sciences and humanities (SSH) Research Infrastructures (RI) in Europe and arrives at the comprehensive "ecosystem of infrastructures", namely the European Open Science Cloud (EOSC).
We highlight how the SSH Open Science infrastructures contribute to the goal of establishing the EOSC. First, through the example of OPERAS, the European Research Infrastructure for Open Scholarly Communication in the SSH, to see how its services are conceived to be part of the EOSC and to address the communities' needs. The next two sections highlight collaboration practices between partners in Europe to build the SSH component of the EOSC and a SSH discovery platform, as a service of OPERAS and the EOSC. The last two sections focus on an implementation network dedicated to SSH data *fairification*.

Keywords: SSH, Research Infrastructure, EOSC, data, FAIR

1. Introduction

The EOSC implementation plan (DG Research and Innovation, 2019) is based on a federated model, aiming at creating, stimulating and implementing synergies between existing scientific resources, primarily through the Research Infrastructures (RI), including e-Infrastructures, part of the Horizon 2020 Work Programme. This paper guides through a long journey, articulated in a path which starts from the OPERAS RI, and crosses various Social Sciences and Humanities (SSH) Research Infrastructures in Europe to arrive at the comprehensive "ecosystem of infrastructures", namely the European Open Science Cloud (EOSC). It makes several stops at different crossroads to highlight the steps which contribute to developing SSH research both at European and international levels. By depicting this scenario, we aim at drawing the picture of an ecosystem, the European Open Science Cloud (EOSC).

While the EOSC implementation is a multi-year undertaking which is being addressed in practice in several stages, different European infrastructures are currently engaged in the activities in the field of Open Science in the SSH. Most of them are dealing with data, especially to develop tools and guidelines for researchers to be able to share, use and host data, following the FAIR[1] principles. In all initiatives, needs of collaboration emerge in order to reinforce the links between data and publications, especially regarding Persistent Identifiers (PID), data journals, etc.

This paper highlights how the SSH Open Science infrastructures contribute on various levels to the goal of establishing the EOSC. First, through the example of OPERAS, the European Research Infrastructure for Open Scholarly Communication in the SSH, to see how its services are conceived to be part of the EOSC and to address the communities' needs. Then the paper points out collaboration practices between partners in Europe to build the SSH component of the EOSC (in the context of the SSHOC[2] H2020 project) and a discovery platform specifically conceived as an OPERAS service to be integrated into the EOSC (TRIPLE H2020 project). The last two parts of the paper focus again on collaborations: at a national level, through the EOSC-PILLAR project, and internationally, through an implementation network dedicated to SSH data fairification.

2. Crossroad 1: OPERAS-P and OPERAS

OPERAS-P[3] is a two-year, European Commission-funded project, aiming at the development of OPERAS - Open Scholarly Communication in the European Research Area for Social Sciences and Humanities - as a European Research Infrastructure[4].
OPERAS-P project will develop a protocol and a roadmap for the inclusion of the OPERAS Research Infrastructure

---

[1] https://www.go-fair.org/fair-principles/

[2] Social Sciences and Humanities Open Cloud project.
[3] H2020-INFRADEV-2019-2. See: https://cordis.euro pa.eu/project/id/871069.
[4] Created in 2015, OPERAS consortium comprises 40 organisations from 16 countries and is led by a Core Group consisting of 9 members.

services for SSH into the EOSC portal. This protocol will be based on the Rules and Procedures already introduced by the EOSC, while taking into account the work in progress of the SSHOC project. The project will implement some of the following OPERAS innovative services, which will be integrated with the EOSC ecosystem:

a. **OPERAS Discovery service.** The TRIPLE project, described in detail below (see Section 4), will become the OPERAS discovery platform, which will provide access to SSH resources, such as data and relevant publications, researcher profiles as well as project descriptions.

b. **OPERAS certification service.** The Directory of Open Access Books (DOAB), which ensures discoverability of Open Access books and delivers global peer-review certification for funders and libraries, will be redeveloped to become a central service of OPERAS as an open source platform based on DSpace technology. This move is crucial for SSH researchers in the light of Plan S[5] and the global shift towards Open Science in Europe.

c. **OPERAS Metrics service.** The Metrics service collects usage metrics and altmetrics from many different sources (Google Books, Matomo analytics, World reader, etc.) about the usage of monographs. Measures are displayed in a light javascript widget, broken down into types and sources, with links to the description of each measure. Different components complement the service, including a data model, an open source tool suite to provide metrics to the service, a central OPERAS database as well as a dashboard and a javascript widget for visualisation.

d. **OPERAS Publishing Service Portal.** Due to the fragmentation of services and tools, SSH researchers in Europe struggle to define and implement their communication strategy in an uncoordinated communication landscape. The OPERAS-P project will implement a common access point to the publishing services offered by its members. This access point is a web portal listing the relevant services provided by the OPERAS infrastructure nodes and beyond. The portal will help researchers in selecting the appropriate publishing venue and defining their scholarly communication strategy.

e. **OPERAS check-in.** To support a transparent and seamless access to the OPERAS platforms and to external sources of data, the EGI check-in service will be adopted as authentication and authorization service within the OPERAS RI. The service provides an identity and access management solution that facilitates the access to services and resources using the federated authentication mechanisms, thanks to the implementation of Virtual Organisation common for OPERAS services and its users.

f. **OPERAS XML toolbox.** In SSH, the community has to overcome a specific obstacle, i.e. the juxtaposition of two standards: XML JATS, adopted by the academic publishing industry, and XML TEI (Text Encoding Initiative) adopted by the humanities research community for books and digital editions. OPERAS-P will provide tools to achieve interoperability between these two standards.

The innovation part of the OPERAS-P project is aimed at producing a robust, empirically tested and stakeholder-validated foundational body of knowledge relevant for the future development and functioning of OPERAS. This includes the development of sustainable models of governance for infrastructures, business models for open scholarly publishing,/groundbreaking concepts to address the fairification of SSH data, multilingualism, the future of scholarly writing as well as quality assessment of novel research outputs.

In sum, OPERAS-P means a process of transforming OPERAS to the status of a mature community, with a set of services compatible with EOSC, stable national nodes and innovative plans for future development.

3. Crossroad 2: Building the SSH component of the EOSC (SSHOC)

The overall objective of the SSHOC project[6] is to build the SSH component of the EOSC.

The project aims at realising the transition from the current landscape with disciplinary silos and separated e-infrastructure facilities into a cloud-based infrastructure where data are FAIR, and tools and training are available for SSH scholars who have adopted, or want to adopt, a data-driven scientific approach and who have an interest in the innovation and integration of their methodological frameworks.

The ambition of SSHOC is to:

a. Increase the efficiency and productivity of researchers - by providing a fully-fledged SSH Cloud where data, tools and services are easily and seamlessly discoverable, accessible and (re)usable.

b. Contribute to the creation of a cross-border and multi-disciplinary open innovation environment - by fostering the development of infrastructural support for digital scholarship.

c. Strengthen/encourage the collaboration between the partners involved in the SSHOC project that are representing the broad spectrum of the SSH community through the use and harmonisation of different technologies and services that are already available and also being developed within the course of the project.

The project therefore aims for synergies across disciplines and work towards a clustered cloud infrastructure that makes use of common elements, such as secured login, storage and computing power, and other e-infrastructures. The project is very well connected to national activities,

---

thanks to the participation of all five SSH ERICs (European Research Infrastructure Consortium). Furthermore, salient pan-European and global data surveys participatie in the project. SSHOC also participates in international activities such as the Research Data Alliance and other initiatives of a similar nature. The SSHOC ecosystem will use the existing infrastructures that are already provided by the project partners and will improve the findability of make existing tools and services for diverse communities of potential use better available. In particular, the SSHOC approach is to develop, enhance, integrate a set of tools and services for managing and processing SSH research data that are central to the communities of use in SSH, based on existing tools and functionalities, and requirements for interoperability. Existing tools and services will be adjusted and enriched, making connections to EOSC-hub e-infrastructure for the sharing and use of tools and services useful for SSH. Special attention is given to cross-disciplinary use of services e.g. providing language technology for social -sciences and humanities scenarios of use.

The SSHOC project will cover the full Research and Development and ready-to-market cycle: in particular, the SSH Open Marketplace platform will contain solutions, training materials, tools and services for researchers, all contextualised within one another. The lack of a central place integrating assets from all SSH-related project websites, service registries and data repositories is what drove the creation of this Marketplace. The choice was made to provide datasets via the Marketplace only when relevant in the context of tools, trainings or other materials[7]. The Marketplace has always, since itsbeginning, been conceptualised as a community-oriented platform where the community can directly take part in the curation of its data. The leveraged services will deeply embed Open Science and FAIR principles by making data Findable, Accessible, Interoperable and Re-usable.

4. Crossroad 3: Building a European discovery service for SSH data (TRIPLE)

SSH research is divided across a wide array of disciplines, sub-disciplines and languages. While this specialisation makes it possible to investigate the extensive variety of SSH topics, it also leads to a fragmentation that prevents SSH research from reaching its full potential. Use and reuse of SSH research is suboptimal, interdisciplinary collaboration possibilities are often missed, and as a result, societal, economic and academic impacts are limited (Dallas C., 2017).

The TRIPLE project[8], which consists of a consortium of currently 19 partners from 13 countries, is a practical answer to the above issues, as it aims at designing and developing a multilingual and multicultural discovery platform dedicated to SSH resources at European scale. TRIPLE will improve the accessibility and dissemination of SSH resources through a single access point which allows free access to circa six million documents in the domain of Social Sciences and Humanities, including peer reviewed journals, articles, books and blog posts, as well as to research data, projects and researcher profiles.

The TRIPLE solution will provide linked exploration thanks to (1) the ISIDORE search engine[9], and (2) a variety of connected innovative tools, which include visualisations, a web annotation service, a trust building system, a crowdfunding system and a recommender system.

TRIPLE main objective is then to enable researchers to discover and reuse SSH data macro-typologies, related not only to publications, but also to people and projects.

The integration of TRIPLE into the EOSC will be performed according to EOSC general principles and to the set of recommendations and guidelines, structured under the six priorities, i.e. Landscape, FAIR, Architecture, Rules of Participation and Sustainability, Skills and Training, which are coordinated by the relative EOSC Working Groups.

A major strength lies in the composition of the TRIPLE consortium: Not only are the main RIs for SSH project partners, but several partners also play an active part in the EOSC implementation. Moreover, specific synergies are developed with SSHOC, and Memorandums of Understanding (starting with SSHOC) are planned.

The TRIPLE solution is envisaged to be a major component of the SSH Open Marketplace, which will be the entry door to the EOSC for all the different SSH services.

The TRIPLE consortium is also experimenting with new forms of engagement and community-building through the TRIPLE Forum, which will bring together relevant stakeholders. Linked to the SSHOC community and the ones served by the Research Infrastructures, TRIPLE Forum will contribute to bringing the researchers into the EOSC and more largely into the Open Science movement.

5. Crossroad 4: Beyond national services, how SSH open collaborations

The EOSC-Pillar project (https://www.eosc-pillar.eu/) aims to identify, coordinate and harmonize existing national initiatives for the national coordination of data

---

[7] TRIPLE could overcome potential gaps by providing access to other datasets, see https://doi.org/10.5281/zenodo.3547649 and Section 4

[8] Funded under the European Commission program INFRAEOSC-02-2019 "Prototyping new innovative services".

[9] ISIDORE is a large-scale discovery service, developed by the TGIR Huma-Num (CNRS) since 2009 (https://isidore.science/).

infrastructures and services that recently started in many Member States (MS) as one of the founding pillars for the development and the long-term sustainability of the EOSC. The idea is, thus, leveraging national initiatives of the MS and Thematic Initiatives (TI) developed by research communities working in national and European collaborations to build a future based on Open Science and FAIR data practices.

Concretely, that implies to:

a. Support the coordination and harmonization of mature national initiatives for open data, open science services, cloud and data infrastructures.
b. Facilitate the adoption and compliance with EOSC standards… while proactively providing feedback to the EOSC governance…
c. Contribute to the creation of an achievable cutting-edge, end user-oriented environment for European data-driven science, through the promotion of FAIR practices and services.

The Federation of National Initiatives will be the catalyst for trans-national open data and open science services (common policies, FAIR services, shared standards, technical choices). The project gathers representatives of the fast-growing national initiatives for coordinating data infrastructures and services in Italy, France, Germany, Austria and Belgium. In this framework, the French Very Large Research Infrastructure Huma-Num and the Center for Direct Scientific Communication (CCSD), who created the HAL open archive and is now in charge of its development and management, together with the conference management platform SciencesConf.org and the hosting platform of epi-journals, decided to join their effort to propose a Proof of Concept (POC) around two of their services for SSH. This POC will link the Huma-Num repository NAKALA to the HAL open archive to address the need for SSH to be able to prove the authenticity of data, and to guarantee accessibility to raw data which are at the root of research and innovation - this approach being in a perspective of reproducibility of the research.

In EOSC-Pillar, the SSH community is built from regional areas. It highlights practices and opens opportunities for new collaborations with other disciplines, so as to bring researchers to new networks and innovative research projects.

6. Crossroad 5: Beyond the EOSC, implementing SSH data FAIRification

CO-OPERAS is an Implementation Network within the context of the GoFAIR initiative (https://www.go-fair.org/implementation-networks/overview/co-operas/). It aims to bring SSH data into the EOSC, helping communities to FAIRifying them, and, in turn, to enrich the FAIRification process and registries with specific SSH standards. "Define FAIR for

implementation" is also the first Recommendation of the DG Research and Innovation, 2018.

The network was created and launched in 2019, and is one of OPERAS' building blocks connecting European and international research communities through the FAIR principles as a common ground. In that sense, within the OPERAS environment, CO-OPERAS' activities represent a reciprocal movement towards and from the research infrastructure: on the one hand, it brings feedback and suggestions from specific communities in order to implement the services; on the other hand it brings coordination to fragmented and heterogeneous communities.

CO-OPERAS stands right at the crossroad between data and publications, and it perfectly fits in the OPERAS ecosystem as it more than integrates data and publications. As a community-based network, CO-OPERAS' first aim is to define the term "data" in the field of SSH. To this purpose, regional and national workshops in different languages (e.g. Italian, German, French…) are being organized. Researchers are asked to provide their definitions of "data", and then to assess the level of FAIRness maturity of the data they are using and creating. Diversity comes along with fragmentation of practices and lack of standards. Then, the SSH community needs to converge around shared expertise and practices. To do so, the FAIR principles are one of the most valuable tools as they are able to be broadly applied and widely shared.

Identifying the gaps and the critical issues is crucial in order to plan new useful services or to create new standards and promote their adoption. In parallel, OPERAS' services and related projects such as TRIPLE and SSHOC will offer a field of application for concrete and improved FAIR data curation, discovery, harvesting, and reuse in the SSH.

7. Conclusions

Building EOSC components implies to be well-organised and coordinated at a European scale. For the Social Sciences and the Humanities, often fragmented also from a linguistic point of view, the challenge is quite high.

The above surveyed initiatives focus each both on general and on specific aspects which, in the end, contribute to define a set of rules and guidelines for the implementation of the SSH components of the EOSC.

This is why there is a strong need for collaborations between European Research Infrastructures, as well as for interoperability of the services. But what is most important is to share a common goal and to work in the same direction.

In general, strong synergies are in place between all the described initiatives and projects:

- the main RIs for SSH, i.e. CLARIN, DARIAH and CESSDA, are TRIPLE project members, and all the five

ERICs (the three above plus SHARE and ESS) are SSHOC project members;

- specific synergies are developed between TRIPLE and SSHOC, where the coordinator (CESSDA) is a TRIPLE partner, and CNRS and CNR are SSHOC partners;

- Memorandums of Understanding are planned;

- EGI partnership within TRIPLE ensures that the technology will be fully interoperable with other e-infrastructures services, especially regarding AAI technology and resource discoverability;

- the collaboration between TRIPLE, SSHOC and the CO-OPERAS Implementation Network, in which, respectively, 12 TRIPLE partners and 3 SSHOC partners are part of, builds a bridge between SSH data and the EOSC, widening the concept of "research data" to all types of digital SSH research outputs;

- numerous discussions about the EOSC are linked to FAIRification of data, in the STM[10] especially focusing on big data. In the SSH field, data does not always fit the definition of "big data", but it still requires specific management and solutions. The CO-OPERAS work on SSH fairification, and specifically on FAIR Implementation Profiles and FAIR Data Objects, can be relevant for SSH initiatives.

SSH contribution to the EOSC definition and implementation draws upon the strong efforts made within the different projects and initiatives to build a strong SSH community. Within the SSH, communities of practice are very fragmented but with a high willingness to share practices and knowledge and to build upon the existing commonalities. Links are strengthened between humanities, social sciences, cultural heritage, scholarly communication communities.

All the above described initiatives show a common vision and complementarity while sharing common challenges, such as overcoming fragmentation and the lack of a single, central solution, addressing common issues such as multilingualism, interoperability, fairification, the EOSC marketplace, language and discovery services, and the connection to national and international activities..

These different initiatives could overlap in their activities at some point. However, this is not an issue. SSH are well-known for their diversity of interpretation and their critical dimensions. What is presented in this paper is a federation of SSH facets which contribute to avoid simplification and reduction in order to deploy complexity at a large scale through the different initiatives. This is where SSH, thanks and through the multiple facets, can play a strong role in the building of the EOSC: they anchor a practice in a history, in an area, in a future.

9. Acknowledgements

10. Bibliographical References

Burgelman J-C, Pascu C, Szkuta K, Von Schomberg R, Karalopoulos A, Repanas K and Schouppe M (2019) Open Science, Open Data, and Open Scholarship: European Policies to Make Science Fit for the Twenty-First Century. *Front. Big Data* 2:43. doi: 10.3389/fdata.2019.00043.

Barbot L, Moranville Y, Fischer F, Petitfils C, Ďurčo M, Illmayer K, … Karampatakis S (2019). SSHOC D7.1 System Specification - SSH Open Marketplace (Version 1.0). Zenodo, 10.5281/zenodo.3547648.the

Directorate-General for Research and Innovation (2018). Turning FAIR into reality, 1-78, European Commission, Brussels, 978-92-79-96546-3, doi: 10.2777/1524.

Directorate-General for Research and Innovation (2019) European Open Science Cloud (EOSC) strategic implementation plan, 1-48, European Commission, Brussels, 978-92-76-09175-2, doi: 10.2777/202370.

OPERAS Consortium. (2018, July 30). OPERAS Design Study. Zenodo. http://doi.org/10.5281/zenodo.1324055

Von Schomberg, R. (2019). "Why responsible innovation?" in *International Handbook on Responsible Innovation A Global Resource*, eds R. Von Schomberg and J. Hankins (Cheltenham: Edward Elgar Publishing), 12–32. doi: 10.4337/9781784718862 .

Wenger, Etienne (1998). Communities of Practice: Learning, Meaning, and Identity. Cambridge: Cambridge University Press;

Wenger, Etienne; McDermott, Richard; Snyder, William M. (2002). Cultivating Communities of Practice (Hardcover). Harvard Business Press; 1st edition.

---

[10] Scientific, Technical and Medical sciences