

■ SOCIAL SCIENCES, HUMANITIES AND THEIR INTEROPERABILITY WITH THE EUROPEAN OPEN SCIENCE CLOUD: WHAT IS SSHOC?

by *Ivana Ilijašić Veršić and Julian Ausserhofer*

Abstract: *The EC H2020 cluster project SSHOC aims to provide a full-fledged Social Sciences and Humanities (SSH) Open Cloud where data, tools, and training are available and accessible for users of SSH data. The focus of the project is determined by the goal to further the innovation of infrastructural support for digital scholarship, to stimulate multidisciplinary collaboration across the various subfields of SSH and beyond, and to increase the potential for societal impact. The intention is to create a European open cloud ecosystem for social sciences and humanities, consisting of an infrastructural and human component. SSHOC will encourage secure environments for sharing and using sensitive and confidential data. It will contribute to the Open Science agenda and realization of the European Open Science Cloud (EOSC), as well as contribute to innovations stemming from the coupling of heterogeneous data types and work on the interoperability principle of FAIR.*

Keywords: *Open Science; federated & inclusive EOSC; Research Infrastructures; Social Science & Humanities; FAIR data*

DIE SOZIAL- UND GEISTESWISSENSCHAFTEN UND IHRE INTEROPERABILITÄT MIT DER EUROPEAN OPEN SCIENCE CLOUD: WAS IST SSHOC?

Zusammenfassung: *Das EC H2020-Clusterprojekt SSHOC zielt darauf ab, eine vollwertige Open Cloud für die Sozial- und Geisteswissenschaften (SSH) bereitzustellen, in der Daten, Tools und Schulungen verfügbar und für die Nutzerinnen und Nutzer von SSH-Daten zugänglich sind. Der Fokus des Projekts wird bestimmt durch das Ziel, die Innovation der infrastrukturellen Unterstützung der digitalen Wissenschaft sowie die multidisziplinäre Zusammenarbeit zwischen den verschiedenen Teilbereichen der SSH zu fördern und darüber hinaus das Potenzial für gesellschaftliche Auswirkungen zu erhöhen. Ziel ist es, ein europäisches offenes Cloud-Ökosystem für die Sozial- und Geisteswissenschaften zu schaffen, das aus einer infrastrukturellen und menschlichen Komponente besteht. SSHOC wird sichere Umgebungen für den Austausch und die Nutzung sensibler und vertraulicher Daten etablieren. Es wird zur Open Science-Agenda und zur Realisierung der European Open Science Cloud (EOSC) sowie zu Innovationen beitragen, die sich aus der Kopplung heterogener Datentypen und der Arbeit am Interoperabilitätsprinzip von FAIR ergeben.*

Schlagwörter: *Open Science; föderative & inklusive EOOSC; Forschungsinfrastrukturen; Sozial- und Geisteswissenschaften; FAIRe Daten*

DOI: <https://doi.org/10.31263/voebm.v72i2.3216>

© Ivana Ilijašić Veršić, Julian Ausserhofer



Dieses Werk ist – exkl. einzelner Logos und Abbildungen – lizenziert unter einer [Creative-Commons-Lizenz Namensnennung 4.0 International](#)

1. The emergence of the European Open Science Cloud

As part of the adoption of the Digital Single Market Strategy, on 6 May 2015 the European Commission announced the launch of one of the most ambitious packages of measures, digital Europe has seen in many years: the European Cloud Initiative (European Commission, 2015). A cornerstone of the European Cloud Initiative is the European Open Science Cloud – short EOOSC, an immense collaborative undertaking of researchers, research infrastructures, the public sector and industry. In order to reach the aim to set up the EOOSC, the European Commission will have invested EUR 600 million by the end of 2020 (European Commission, 2018). The EOOSC aims to create a trusted environment for hosting and processing research data to support EU science in its leading role. The cloud is closely connected to open science, which can be defined as the ongoing transition of how research is performed and how knowledge is shared. It includes open access, which is the practice of providing online access to scientific information free of charge to the user, and reusable. Open science and FAIR data (Wilkinson et al., 2016) have been endorsed not only at the European level – in the Conclusions of the 2016 Competitiveness Council in May (General Secretariat of the Council, 2016), but also internationally at the G7 Science Ministers Meeting in Turin in September 2017 (G7 Expert Group on Open Science, 2017).

Ideally, the EOOSC contributes to a public research commons, thereby tackling several issues such as the commodification of academia, the growing predominance of proprietary research infrastructures and a diverse landscape of legal frameworks. The EOOSC is “a process, not a project. [...] It envisions to interlink existing European data infrastructures, integrating high-capacity cloud solutions, and in due course, widening the scope of these services to include users from the public sector and industry. Efforts focusing on the development of sustainable collaborative ecosystems, as

foreseen by the EOSC, are particularly made in the fields of *Data Culture*, *Research Data Services*, *Federated Architecture* and *co-Funding*.” (Budroni, Burgelman, & Schouppe, 2019, p. 130).

The EOSC should be seen as not only a technical “cloud”, in a sense of available data storage and computing power. It is foremost a growing-together of the cultures and practices of different fields of research and its infrastructures, connecting different communities and disciplines. It is envisioned for EOSC to foster best practices of global data findability and accessibility (FAIR data), to help researchers get their data skills recognised and rewarded (careers, altimetric), to address issues of access and copyright (IPR) and data subject privacy, to allow easier replicability of results and limit data wastage e.g. of clinical trial data (research integrity), to contribute to clarification of the funding model for data generation and preservation, and finally, to reduce rent-seeking and priming the market for innovative research services, e.g. new business models (European Commission, 2016). While many of these goals and issues have been prevalent for many years, it is clear that the EOSC cannot solve these aspects conclusively; nevertheless, it is expected it will advance substantially the state-of-art in these domains.

In the Horizon 2020 Infrastructure and e-Infrastructure Work Programme 2018–2019, Implementation of the European Open Science Cloud (H2020-INFRAEOSC-2018-2020) supports actions to make the European Open Science Cloud (EOSC) a reality. It consists of calls and funding instruments, most of them being research and innovation actions. One of the calls is focused on the connection of the major European research infrastructures (ESFRI)¹ through cluster projects (INFRAEOSC 04-2018). Projects funded under this call should work against the fragmentation of the research data landscape and provide support to actions included in the 2016 Communication on the European Cloud Initiative. In detail, they should integrate and consolidate e-infrastructure platforms, connect the ESFRI infrastructures to the EOSC, and develop a European Data Infrastructure (EDI).

2. The social sciences and humanities (SSH) area of the EOSC

The project Social Sciences and Humanities Open Cloud (SSHOC)² is part of the activities foreseen under INFRAEOSC 04-2018. Other projects funded under this call are: ENVRI-FAIR for environmental research, PaNOSC for multidisciplinary scientific analysis, ESCAPE for astronomy and particle physics, and EOSC-Life for life sciences.

The overall objective of the SSHOC project is to realise the social sciences and humanities' part of European Open Science Cloud (EOSC). It has started on 1 January 2019, and throughout 40 months it shall provide a significant contribution towards achieving the vision put forward by the European Cloud Initiative that was set up by the European Commission (2016). 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 scholars from those domains in the social science and humanities that have adopted a data-driven scientific approach and that have an interest in the innovation and integration of their methodological frameworks.

All Social Sciences and Humanities (SSH) research infrastructures established as ESFRI Landmarks and Projects³, as well as relevant international SSH data infrastructures and the association of European research libraries (LIBER), participate in this project. This ensures a federated and inclusive approach as required by the EC Strategic Working Document (European Commission, 2018, p. 4).

Moreover, the consortium has the expertise to cover the whole data cycle: from data creation and curation to optimal re-use of data. It can also address training and advocacy to increase actual re-use of data. The consortium is also very well placed to address SSH specific challenges such as the distributed character of its infrastructures, multi-linguality, huge internal complexity of some of the data it deals with, and different models of access to sensitive data. The project creates, pools and harmonises easily usable tools and services that allow the research community and other interested users to deal with the vast heterogeneous collections of data available, to process, enrich, analyse and compare it across the boundaries of individual repositories or institutions. SSHOC – by the partners in the project – could be one of the thematic EOSC federated centres (European Commission, 2018, p. 10).

The project plans to harness the innovation potential of infrastructural support for digital open scholarship, stimulate multidisciplinary collaboration, increase the potential for societal impact and provide readiness for mission-oriented research and innovation (Mazzucato, 2018).⁴

3. Building the SSH Cloud

Currently, SSH infrastructures are distributed and sometimes scattered or only locally available, lacking appropriate documentation and metadata,

which limits their re-use. By realising a common SSH Cloud that provides a recognisable and accessible environment for data, tools and services, and training those limitations should be reduced. By involving researchers in the development of this environment and by showcasing best-practices of data sharing the problem that a lot of SSH data is not shared will be tackled.

In line with the idea of a cloud, the project will not only create one centralised infrastructure; rather it will develop a virtual infrastructure with existing and new infrastructures connected. This approach is congruent with the character of SSH and more flexible than one single, centralised infrastructure. Thereby it will improve the sustainability and the innovative capability of the SSH Cloud.

4. Maximising reuse through Open Science and FAIR principles

Another major aim of the project is to embed Open Science and FAIR principles initiatives in the project structure. The project will generate services for optimal re-use of data by making data Findable, Accessible, Interoperable and Re-usable (FAIR). This objective is challenging since SSH data is very heterogeneous in its form and purpose. In order to realise open and FAIR data, the project aims to set standards for metadata, provide a common service catalogue, provide secure environments for sensitive data with easy single sign-on, experiment on connecting data using semantic techniques⁵, and include a human-centric approach by providing training to experts and data producers and data users.

The project will apply the FAIR principles to the creation of data in order to increase the efficiency and ease of creating and re-using new data. By making use of existing (and already tested) survey questions, multilingual question banks, common vocabularies, and proven sampling methods, SSHOC will ensure that newly generated data are cheaper to produce, that they are embedded more easily into the existing data infrastructures and have increased re-use because their questions are comparable to other data collections.

5. Interconnecting existing and new infrastructures

The SSHOC project can also be considered as an ecosystem that will use partnering infrastructures and make existing tools and services better

available. The SSHOC consortium brings together all the key players in the European Social Sciences and Humanities as well as additional important contributors, which are needed to achieve the project aims of making data accessible as widely as possible across SSH subdomains and beyond. The consortium is formed of organisations from 16 European countries, covering all SSH ESFRI Landmarks and Projects, as well as important international SSH data infrastructures and associations. In total 45 partner organisations participate in the project: 20 beneficiaries and 25 linked third parties. It aims for synergies over disciplines and works towards a clustered cloud infrastructure making use of common elements, such as secured log-in, storage and computing power, and other e-infrastructures.



Fig. 1: The beneficiaries and linked third parties of SSHOC

Furthermore, the project will deal with interdisciplinary use of data e.g. facilitating combination of social and environmental data, social and health data, and especially in heritage science there will be combinations of SSH, physics and environmental data⁶. The SSHOC members will use their expertise on distributed infrastructures to connect new and existing infrastructures. The project will also include an initiative to connect to the other Clusters (Physics, Environment, Life & Health, etc.)

6. Governing Social Sciences and Humanities in EOSC

The SSHOC project will set up an appropriate governance model for the SSH part of the EOSC, in collaboration and seeking commonalities with other cluster projects, EOSC Working Groups and the EC. The project will actively involve and support national decision-makers, service providers, researchers and policy makers as the most important stakeholders. The presence and involvement of five Landmarks and one project in the ESFRI Social and Cultural Innovation (S&CI) landscape, large and stable international data infrastructures and the LIBER-network, will ensure sustainability of the SSH platform and direct contact with researchers and other professional users.

Ivana Ilijašić Veršić, MA MSc
CESSDA – Consortium of European Social Science Data Archives
E-Mail: ivana.versic@cessda.eu

Dr. Julian Ausserhofer
ORCID: <https://orcid.org/0000-0001-6782-5988>
University of Vienna, AUSSDA – The Austrian Social Science Data Archive
E-Mail: julian.ausserhofer@univie.ac.at

References

- Budroni, P., Burgelman, J.-C., & Schouppe, M. (2019). Architectures of Knowledge: The European Open Science Cloud. *ABI Technik*, 39(2), 130–141. <https://doi.org/10.1515/abitech-2019-2006>
- European Commission. (2015). *A Digital Single Market Strategy for Europe (COM(2015) 192 final)* [Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions]. Retrieved from <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2015:192:FIN>
- European Commission. (2016). *European Cloud Initiative: Building a competitive data and knowledge economy in Europe (COM(2016) 178 final)* [Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions]. Retrieved from <https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:52016DC0178>
- European Commission. (2018). *Implementation roadmap for the European Open Science Cloud (SWD(2018) 83 final)* [Commission Staff Working Document]. Retrieved from <https://ec.europa.eu/transparency/reg-doc/rep/10102/2018/EN/SWD-2018-83-F1-EN-MAIN-PART-1.PDF>
- G7 Expert Group on Open Science. (2017). *WG Open Science*. Retrieved from http://www.g7italy.it/sites/default/files/documents/ANNEX%204_WG%20Open%20Science/index.pdf
- General Secretariat of the Council. (2016). *The transition towards an Open Science system* (Council Conclusions No. 9526/16 RECH 208 TELECOM 100). Retrieved from <https://data.consilium.europa.eu/doc/document/ST-9526-2016-INIT/en/pdf>
- Mazzucato, M. (2018). *Mission-oriented research & innovation in the European Union: A problem-solving approach to fuel innovation-led growth*. <https://doi.org/10.2777/360325>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>

- 1 European Strategy Forum on Research Infrastructures
- 2 <https://sshopencloud.eu/>
- 3 <http://roadmap2018.esfri.eu/projects-and-landmarks/>
- 4 Example activities include the development of the SSHOC Marketplace, a registry, where tools, services and other resources brought in by project partners will be listed and contextualized. The SSHOC Marketplace will be constructed in a way that it can be harvested by the to-be-developed EOSC catalogue (<https://www.sshopencloud.eu/marketplace>). Other project activities, which foster cooperation and innovation, are workshops and training collaboratively facilitated by infrastructures from different domains (<https://www.sshopencloud.eu/training>).
- 5 For instance, Task 9.3 of the project (“Data Community Project: Electoral Studies”) aims to generate an open research knowledge graph in the field of electoral studies by enriching electoral behaviour (mainly survey data) with taxonomies and ontologies. This will help researchers outside of electoral studies to use data generated in that field.
- 6 Heritage science is an interdisciplinary undertaking. Chemists, physicists, material scientists, engineers, archaeologists, digital humanists, conservators, literary scholars, historians and art historians – to name most involved disciplines – collaborate to explore and preserve cultural and natural heritage (collections, buildings, archaeological sites, digital and intangible heritage). In this context, Task 9.4 of the project („Heritage Science and Humanities”) will highlight where different notions and practices of different disciplines will need to be aligned to fit the EOSC requirements and landscape.

Funding

Social Sciences and Humanities Open Cloud (SSHOC) has received funding from the EU Horizon 2020 Research and Innovation Programme (2014-2020); H2020-INFRAEOSC-04-2018, under the agreement No. 823782.