



SSHOC
social sciences & humanities open cloud

Social Sciences & Humanities Open Cloud

Position Statement

Table of Contents

Preamble	3
Introduction	3
EOSC	3
SSHOC	4
Inclusiveness	5
Tech and Human Centric	5
What is the research community expecting from EOSC?	6
SSHOC Data Infrastructure for Open Science	6
EU wide availability of high quality SSH data	6
EU wide availability of high quality “cloud ready” SSH tools	6
SSHOC Open-Source Software and Service Repository – the Market Place	6
Availability of an EU wide, easy to use SSH Open Market Place	6
SSHOC Secured Environment for Data Analysis	7
EU wide availability of trusted and secure access mechanisms for SSH data	7
SSHOC Engagement and Communication	7
Data sharing is accepted practice (the “new normal”) among the different SSH communities	7
State of the art advanced through dedicated SSH data pilots cluster projects	7
The Social Sciences and Humanities are seamlessly integrated in EOSC	8
What added value will EOSC bring to the research?	8
Increase the efficiency and productivity of researchers	8
Contribute to the creation of a cross-border and multi-disciplinary open innovation environment	8
New discoveries	8
Which main issues should be considered by EOSC executive and governing boards ?	9

Preamble

In this document, we address three questions:

1. **What is the research community expecting from EOSC?**
2. **What added value will EOSC bring to the research?**
3. **Which main issues should be considered by EOSC executive and governing boards?**

Introduction

EOSC

EOSC is about connecting Research Data with (e-Infrastructure) Tools & Services, following the FAIR principles – not just for research data, but also for software and the way science is carried out. EOSC is about breaking down silos and providing seamless access to research data, tools and facilities.

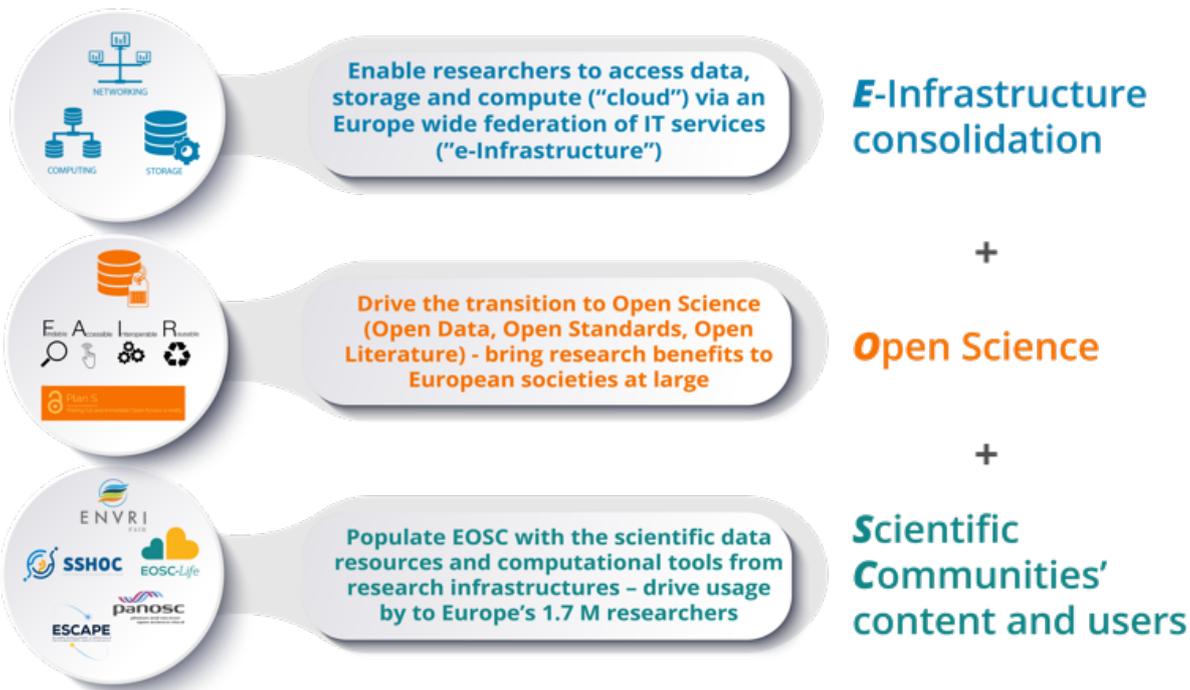
EC President Ursula von der Leyen at the World Economic Forum, Davos, 22 January 2020

We are creating a European Open Science Cloud now. It is a trusted space for researchers to store their data and to access data from researchers from all other disciplines. We will create a pool of interlinked information, a 'web of research data'.



Source image via: https://ec.europa.eu/commission/presscorner/detail/en/SPEECH_20_102

European Open Science Cloud =



Source: Blomberg, Niklas, & Petzold, Andreas. (2020, January 30). ESFRI thematic cluster view on EOSC. Zenodo. <http://doi.org/10.5281/zenodo.3631247>

SSHOC

The overall objective of the Social Sciences and Humanities Open Cloud (SSHOC) project is to realise the social sciences and humanities' part of European Open Science Cloud (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 scholars - especially 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.

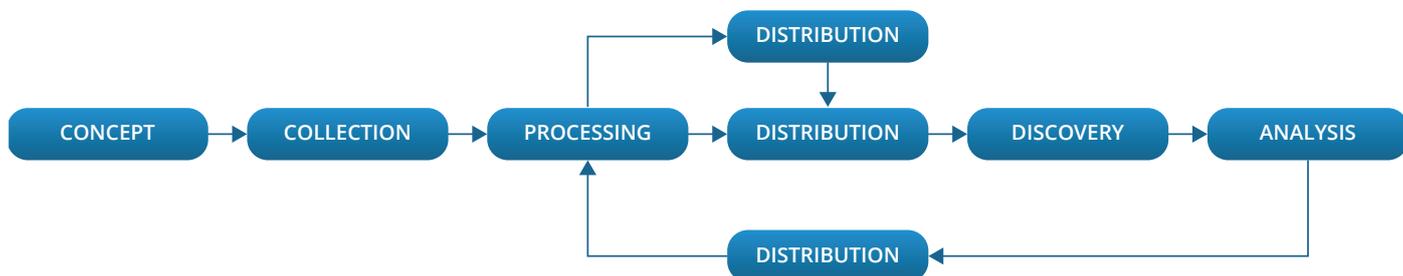
Specific Objectives

- Build the SSH Cloud
- Maximise re-use through Open Science and FAIR principles
- Interconnect existing and new infrastructures
- Governance for SSH-EOSC

The overall impact and end-result of SSHOC will be a SSH data ecosystem in which researchers and other interested parties have seamless access to high quality data.

Inclusiveness

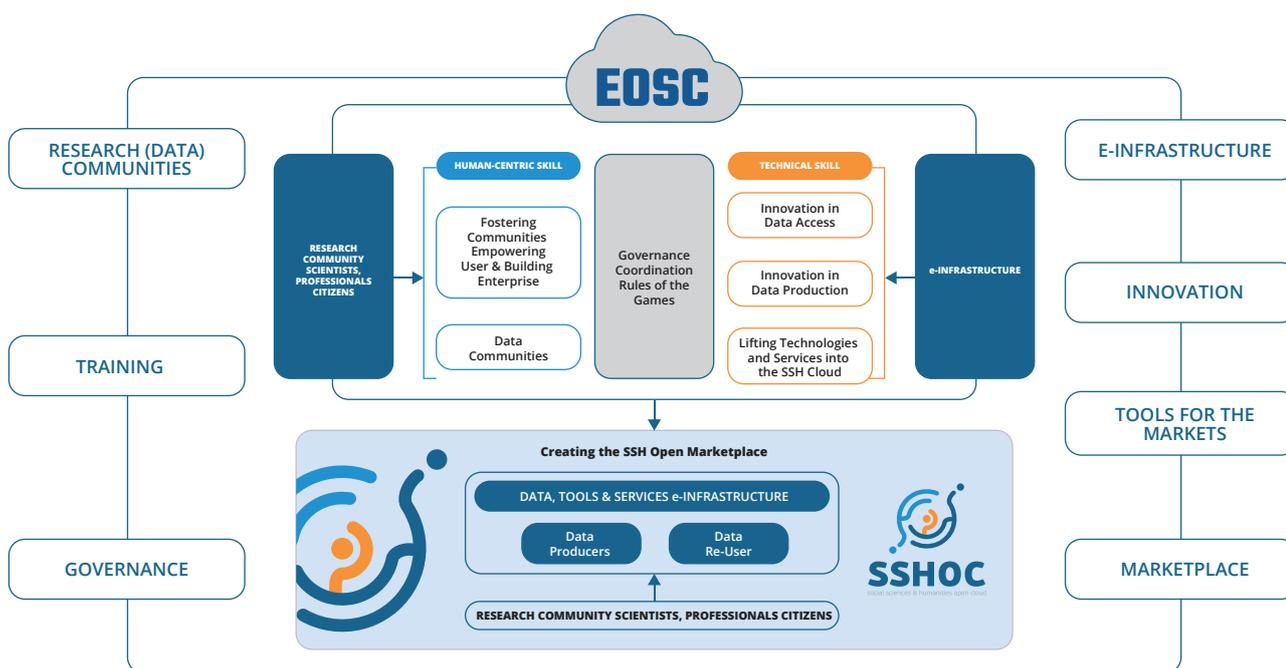
All SSH ESFRI Landmarks and Projects, as well as relevant international SSH data infrastructures and the association of European research libraries participate in this project. This will ensure an inclusive approach. Moreover, the consortium has the expertise to cover the whole data cycle: from data creation and curation to optimal re-use of data, and can also address training and advocacy to increase actual re-use of data.



Source: DDI Alliance

Tech and Human Centric

Although often overlooked, the community aspect and mutual understanding that we term “human centric” is the key for successful uptake of the infrastructure. Therefore, development of synergies and complementarity between involved research infrastructures is of the utmost importance, thus contributing to the development of a consistent SSH research infrastructures ecosystem. Community building around the cloud, as well as a strong connection with end users and intermediaries (e.g. research libraries and their institutions), and social networks are essential for its success.



SSHOC Project Structure

The SSH cloud infrastructure will be distributed, which will be beneficial to its scalability. It will foster data, tools & services in different domains and elaborate on connecting data to increase reusability. The consortium is 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 secured access to sensitive data.

The project will pool, harmonise and make easily usable tools & 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.

What is the research community expecting from EOSC?

In line with the objectives of Open Science, the SSHOC project will improve access to data and provide tools, enabling new and interdisciplinary research leading to new insights and innovation for society. The overall impact and end-result of SSHOC will be a SSH data ecosystem in which researchers and other interested parties have seamless access to high quality data and this system will be integrated in the European Open Science Cloud.

SSHOC Data Infrastructure for Open Science

EU wide availability of high quality SSH data

SSHOC will develop shared web-based tools and services to assist data producers at different stages of the data lifecycle to collect, process, archive and share high-quality cross-national research data and metadata in a cost-effective and streamlined way.

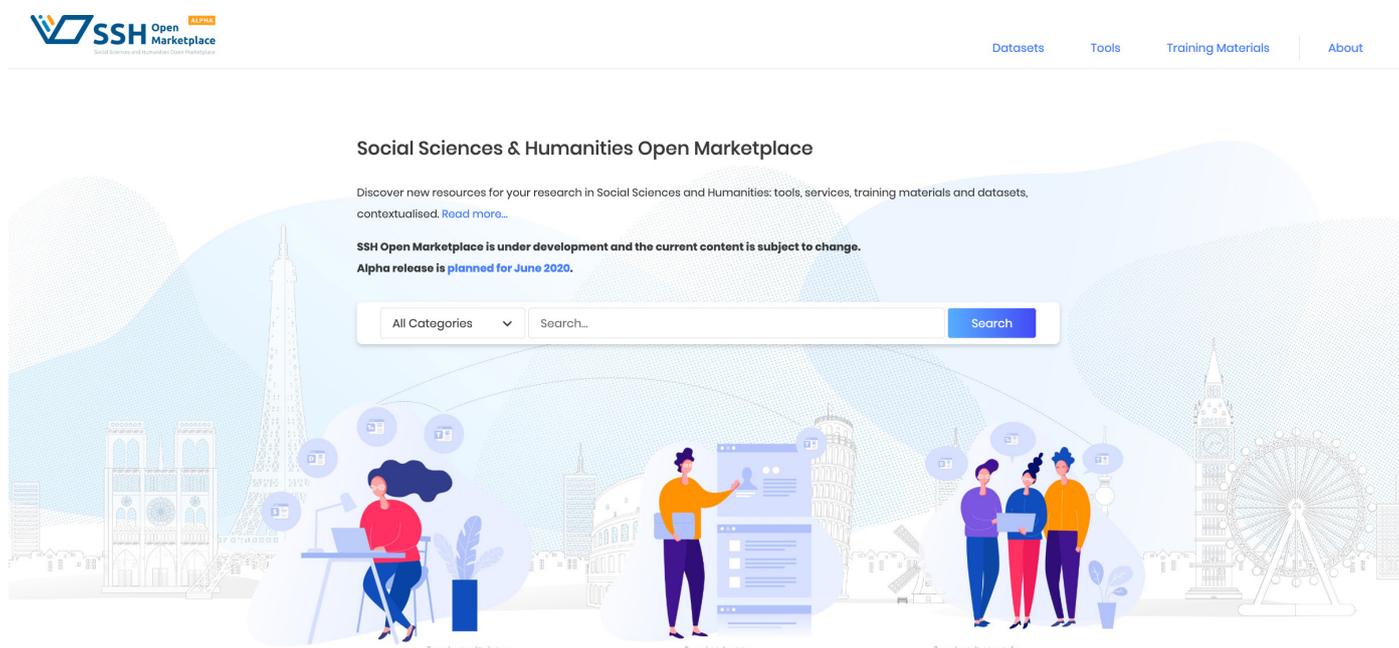
Secondly, SSHOC will explore how data producers can employ technology to add value to existing SSH primary data collections by, for example, automating the gathering and incorporation of contextual information. Such innovations will encourage re-purposing and reuse of data across disciplines. It will benefit data users (analysts) and policy makers as well as the data producers who are able to deliver more with less and increase the visibility and value of data to the scientific and policy communities. The results will be made openly available, unless privacy regulations require secured access.

EU wide availability of high quality “cloud ready” SSH tools

Many tools for data managing and processing already exist, but only some are ready for deployment in the cloud. SSHOC will therefore adjust and enrich existing tools and services for managing and processing SSH data, thus making them “cloud ready”. In this context “cloud ready” refers to making them interoperable, citable and findable and advertised in the Market Place, actionable via the SSHOC switch board and packaged for deployment in the EOSC.

SSHOC will also develop a new suit of tools and services for managing and processing SSH tools and services that are central to the SSH communities.

SSHOC Open-Source Software and Service Repository – the Market Place



Availability of an EU wide, easy to use SSH Open Market Place, where tools and data are openly available
The tools and data developed in SSHOC will be made widely available through the SSHOC Open Market Place¹ where scholars from the broader SSH domain can find solutions and resources for the digital aspects of their research.

It will adopt a platform which allows for contextualisation and interrelation of datasets, tools, and services offered, etc., with screenshots, tutorials and links to training material, user stories, showcases, and other related resources. It will also encompass community features and contain a rating and assessment feature, based on previous work in Humanities.

The SSHOC Market Place is a platform with (free and commercial) services, and tools for working on the SSH data cycle: from creation to reuse of data. These services include training and will also link existing data catalogues of the SSH infrastructures.

SSHOC Secured Environment for Data Analysis

EU wide availability of trusted and secure access mechanisms for SSH data, conforming to EU legal requirements

Research data in the social sciences is often connected to individuals, requiring special security and protection. Data anonymization can limit risk, but a loss of detail reduces utility and thereby limits research potential. To maximise data utility, research value, and policy impact, an interdisciplinary approach to secure data sharing is required. SSHOC will therefore develop services which will offer secure and trusted repositories for storing and accessing SSH data. This will be built on an open source software platform, customised to the needs of the European SSH community. The service will be developed in such a way as to ensure its sustainability after the end of the action. In this context, SSHOC will also address the legal issues related to open access and reusability of SSH research data, as well as issues related to legal and ethical implementation of the FAIR principles. The impact on SSH research data of the GDPR, Ownership and Intellectual Property Rights (IPR), the new European e-Privacy Regulation, as well as ethical issues will be analysed. One of the results will be the development of a common SSH GDPR Code of Conduct to support the realisation of the EOSC. To demonstrate that sensitive data can meet the standard of FAIR access to data, by being made “intelligently open”, SSHOC will provide a framework of confidentiality levels for SSH data (based on global ‘data tagging’ categorisation) and how it might be implemented to other data to meet FAIR principles.

SSHOC Engagement and Communication

Data sharing is accepted practice (the “new normal”) among the different SSH communities

Within the EOSC context, a successful implementation of the SSHOC requires a strong and sustained engagement with the user base and their communities.

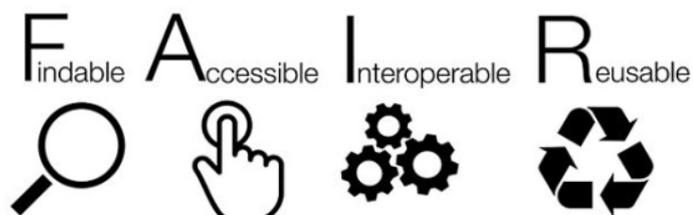
SSHOC will foster a data sharing culture according to FAIR principles by providing context-driven training around the SSHOC infrastructure. In this context SSHOC will harmonize existing training initiatives and expand the portfolio of training materials and actions from a SSH perspective in cooperation with other relevant EU funded projects in the area (e.g. FOSTER Plus, EOSC-pilot, EOSC-hub, OpenAIRE-Advance, FREYA).

¹[SSHOC D7.1 System Specification - SSH Open Marketplace DOI 10.5281/zenodo.3547649](https://doi.org/10.5281/zenodo.3547649)

State of the art advanced through dedicated SSH data pilots cluster projects

SSHOC will undertake several pilot studies on implementing FAIR principles in research communities:

- Increase Findability within Migration & Mobility by connecting with EC Projects on this subject;
- Improve Accessibility within Humanities and Language, also to transform data into information of interest to decision makers;
- Realise Interoperability by linking election data using semantic techniques by connecting with an EC Project on Historical Economic & Company Data;
- Encourage Reusability by developing training in digital heritage science and work on heritage science data transformation (text & data mining, machine learning, predictive modelling) to enable large heritage datasets to be interpreted.



Source: <https://dit.libguides.com/c.php?g=670487&p=4760339>

The Social Sciences and Humanities are seamlessly integrated in EOSC

For SSHOC to reach its stated objective of pulling down traditional silos and build an integrated cloud infrastructure, it needs to seamlessly interact with other Data Clusters within EOSC. For this purpose, SSHOC will develop and implement a common governance model for the project results as part of EOSC. It will include common policies on FAIR principles, data stewardship and harmonization, as well as quality assessment (including certifications), legal and ethical issues and fostering strong collaboration with different EOSC thematic clusters. SSHOC will also assist (inter)disciplinary user communities in overcoming the challenges that they encounter when attempting to contribute to SSHOC and EOSC² by implementing principles, procedures, tools and services developed in specific research communities and other projects.

What added value will EOSC bring to the research?

Increase the efficiency and productivity of researchers

By providing a full-fledged social sciences and humanities Cloud where data, open data tools and services, are offered as part of infrastructure with easy and seamless discovery, access, and re-use are available for users of social, humanities and cultural heritage data.

Contribute to the creation of a cross-border and multi-disciplinary open innovation environment

By fostering the innovation of infrastructural support for digital scholarship, stimulating multi-disciplinarity and collaboration across the various subfields of social sciences and humanities and with other science domains.

New discoveries

From the Davos-2020 speech of EC President Von der Leyen: There are “hidden treasures and untapped opportunities in the data we generate. ...Every researcher will be able to better use not only their own data, but also those of others. They will thus come to new insights, new findings and new solutions”. If we want to solve societal challenges – e.g. the Sustainable Development Goals, EC Mission – scientists

²SSHOC D9.1 Challenges that user communities face when attempting to contribute to SSHOC DOI: [10.5281/zenodo.3569844](https://doi.org/10.5281/zenodo.3569844)

must be able to cooperate beyond disciplinary barriers and be able to use each other's data. The EC Mission on 'climate-neutral and smart cities' will require experts from social behaviour, economics, urban planning, biology, geography, chemistry, medicine, biology, environmental science, computer science, physics, etc.



Source: https://www.eltis.org/sites/default/files/news/shutterstock_563320966_0.jpg

Which main issues should be considered by EOSC executive and governing boards?

- Address and coordinate on key elements of a platform or Research Data Commons³
- Persistent Identifiers – for (meta)data, publications, services – and possibly organisations and researchers .
- Metadata standards for the respective domains.
- SSH will use metadata templates, controlled vocabularies and data models used in well-curated datasets. But especially controlled vocabularies need support and require agreement over domains and on a global scale.
- User Interfaces – for Service Catalogues, Market Place, and other key parts of EOSC.

Foster the Interoperability and Reuse of Research Data

- Secured Data Environments for storage, management, access, computation and analysis of research data.
- Instruments to describe connections & interoperability between data sets.
- For example, SSH will use DDI (XML) as a rich schema to support extensive variable metadata and references to other data⁴.
- A key issue for use of platforms is trust.
- This implies quality assurance for the data, the services, the (software) tools that are in EOSC.

Ensure and foster the composability of EOSC

- Breaking down silo's – between data, computing power, storage and networks – is a salient value added of EOSC. End-users should be able to compose combinations of services themselves in an easy and straightforward way.
- Encourage EOSC as a working space for researchers – to work in secured environments, to combine data from various domains, to use tools & services from the Market Place.

Provide stability and sustainability in providing EOSC

- Partners – either commercial or non-commercial – will only invest in EOSC if there is a guarantee that EOSC will be ensured for a long-term period.
- Implement an operational, scalable and sustainable EOSC federation, allowing seamless alignment and convergence with data infrastructures.
- Provide clear rules of participation with defined funding responsibilities and transparent and ongoing financial commitment by all irrelevant stakeholders.
- Take advice and align efforts with ongoing (especially domain specific) initiatives and H2020 projects.
- Provide and maintain communication channels and close connection to the research community.

³cf. NIH Commons <https://nihdatacommons.us/>

⁴D3.1 SSHOC D3.1 Report on SSHOC (meta)data interoperability problems DOI: [10.5281/zenodo.3569868](https://doi.org/10.5281/zenodo.3569868)

