February 2020: This version of the paper incorporates comments received during consultation (Nov-Dec 2019), and responses to them. It should still be regarded as a working draft.

The consultation document itself is available to view at http://tiny.cc/FedCorePPv1.

EOSC Federating Core

Community Position Paper v1.0

(Working Draft)

The EC's stated objective for the EOSC is for it to be a fundamental enabler of Open Science and of the digital transformation of science, offering every European researcher the possibility to access and reuse all publicly funded research data in Europe, across disciplines and borders, leveraging past investment in research data infrastructures to add value in terms of scale, interdisciplinarity, and faster innovation[1].

Based on the experience and knowledge of EOSC implementation projects and ESFRI projects and landmarks, and the needs of the user communities represented by the EOSC cluster projects[2], a set of requirements for the EOSC, and a definition and composition of the Federating Core and its value add for the EOSC are described below. These have been developed from initial proposals for the EOSC Federating Core produced by the EOSC-hub project[3], updated to incorporate feedback and input received from EOSC implementation projects and national research infrastructures, including ESFRI clusters and regional projects. The description of what the cluster projects require of the EOSC, and what they will offer to it, can be found in the Appendix.

Requirements of the EOSC

The EOSC should provide in a coordinated manner, across disciplines, the provisioning of capabilities that are generally applicable to data lifecycle management:

- discovery and reuse: provide a means for universal and versatile discovery and sharing of resources[4], through a portal, and with inclusive and transparent policies for access
- processing and analysis: provide for common user needs for generic storage and processing facilities for data management and analysis, such as high-performance and high-throughput

---

[1] https://ec.europa.eu/research/openscience/pdf/swd_2018_83_f1_staff_working_paper_en.pdf

[2] https://www.eosc-portal.eu/news/five-new-esfri-cluster-projects-eosc-panorama

[3] Briefing Paper - EOSC Federating Core Governance and Sustainability - https://www.eosc-hub.eu/sites/default/files/EOSC-hub%20Briefing%20Paper%20-%20EOSC%20Federating%20Core%20Governance%20and%20Sustainability%20Public.pdf

[4] Defined in the EOSC Portal glossary as *any asset made available (by means of the EOSC system and according to the EOSC Rules of Participation) to EOSC System Users to perform a process useful to deliver value in the context of the EOSC. EOSC Resources include services, datasets, software, support, training, consultancy or any other asset.* See https://www.eosc-portal.eu/glossary

distributed compute capabilities, for researchers to manipulate resources to which they have been afforded access via the EOSC

- <u>data management, curation, preservation</u>: complement what is provided by research institutions and communities, according to the subsidiarity principle, and provide researchers with a working space where they can use EOSC resources collaboratively. Research institutions, communities and infrastructures should remain the main custodians of research data, quality and FAIR policies
- <u>access/deposition and sharing</u>: widen access to data produced, curated and preserved by national and European research communities and enable the reuse of research communities' FAIR data and data analytics tools.

In the interests of delivering opportunities and efficiencies, the EOSC needs to interoperate with other infrastructure initiatives in Europe and other regions, including EuroHPC[5].

EOSC Federating Core

The Federating Core required to support this vision of the EOSC is defined below.

The **Federating Core** is a fundamental asset of the EOSC, composed of the technical, human, policy and resource elements required to facilitate, monitor and regulate as appropriate day-to-day transactions across the federation.

The Federating Core should deliver three capabilities:

(1) **Hub Portfolio**: The activities and tools necessary to provide coordinated access to and management of resources[6] provided in the EOSC Shared Resources or the Service Portfolio. EOSC resources are expected to be delivered at national and European level, together with the support and expertise necessary to address complex digital needs of the EOSC user communities. The Hub portfolio delivers the EOSC "**federating tier**".

(2) **Compliance Framework**: the Rules of Participation, the Interoperability Framework, the Service Management System and other policies and processes for suppliers and users to engage with the EOSC. The Compliance Framework constitutes the EOSC "**regulatory tier**".

(3) **Shared Resources**: Resources including scientific outputs (local copies of data; applications, software, pipelines etc) and the storage and compute hosting platforms needed to deposit, share and process them. The shared resources realise the EOSC "**resource tier**".

The Federating Core is complemented by the **EOSC Service Portfolio** which provides additional added-value services which exploit the Federating Core, are delivered by providers external to the

---

Commented [1]: The term "hub porfolio" is not included in the EOSC Glossary. This terminology is very confusing.

Commented [2]: Noted. Thank you for your comment. We are aware there are different terms circulating for similar concepts, and that this is confusing. We are open to reviewing the vocabulary used. Ongoing work in various projects and groups including the Glossary Interest Group will no doubt help to reduce this and converge on agreed terminology and definitions.

Commented [3]: Such as? Differences between Hub portfolio and shared resources can be tricky for some of the services. It would be useful to give precise examples of what will be inside the Hub.

Commented [4]: We are aware that the distinctions can be tricky to define and understand.
This position paper does not include such details, to keep it short, but instead references the two Briefing Papers (see references in footnote 8), the second of which also contains a list of examples of possible services which could comprise the Shared Resources, to try to better illustrate the Shared Resources and the difference between them and the Hub Portfolio or Service Portfolio functions/services.

---

[5] https://eurohpc-ju.europa.eu/

[6] Defined in the EOSC Portal glossary as *any asset made available (by means of the EOSC system and according to the EOSC Rules of Participation) to EOSC System Users to perform a process useful to deliver value in the context of the EOSC. EOSC Resources include services, datasets, software, support, training, consultancy or any other asset.* See https://www.eosc-portal.eu/glossary

> EOSC according to independent provider-specific business models, and are discoverable through the EOSC Portal.
>
> The composition of the Federating Core and the EOSC Service Portfolio will be driven by EOSC-defined Rules of Participation[7], technical and policy requirements that will define the EOSC conformance requirements for providers.
>
> The set of capabilities delivered by the Federating Core is defined by the EOSC governance, and the costs of its delivery shall be sustained by EOSC funding.

Initial proposals of the functions and resources which might comprise the three elements of the Federating Core are provided below[8]. Further resources or functions may still be added as definition of the EOSC progresses.

Shared Resources

The resources required from the EOSC Shared Resources differ from one discipline or community to another but possible capacities they could comprise include

- High bandwidth networking connectivity for high-performance access to EOSC data hosting nodes that provide storage and compute resources
- AAI services
- High-performance European distributed (federated) cloud storage environments for secure access, staging, downloading and deposition of large volumes of data across national, institutional and Research Infrastructure boundaries
- High-performance and high-throughput distributed (federated) compute capabilities for big data processing and analysis, including simulations
- A powerful search machine to support findability of scientific resources including data, tools, software and publications across many domains, and easily browsable federated dataset catalogues
- A repository of tools, services, software and workflows for data exploitation: simulation, analysis, enrichment and comparison of data from different national domains
- A catalogue of training materials and competence reference materials
- Open Science policy and practice recommendations for institutions and other EOSC stakeholders
- A code repository
- PID services
- Personalised workspaces for researchers, based on federated AAI.

Hub Portfolio

The list of functions proposed to constitute the Hub Portfolio includes

- EOSC Portal
- EOSC Support Services, including training, competence centres and knowledge bases

---

[7] Defined in the EOSC Portal glossary as *the principles defined by the EOSC Governance to drive the processes enacting an actor to play the role of EOSC System User (and any specialization of it)*.

[8] Descriptions of most of these can be found in the initial EOSC-hub Briefing Paper on the Federating Core (section 3.1) with information on more recent additions in the Briefing Paper v2.0 (section 3.3.2).

Commented [5]: these headings should be listed in the same way as they are abovie i.e. 1. hub portfolio 2. compliance framework 3. shared resources. otherwise it is confusing to the reader

Commented [6]: Thank you. This is noted for future versions.

Commented [7]: Regarding my comment above, does it mean that the Hub will be composed of very generic services dedicated to the use of the EOSC platform in itself and not for conducting research? What would be for instance the difference between EOSC data transfer services and other data transfer services which can be added in the Shared Resources? Just to be clear...

Commented [8]: The Hub Portfolio would consist of the "backend" functions required to deliver the federating tier. The researcher-facing, value-adding services would be in Shared Resources and Service Portfolio.

- EOSC AAI
- EOSC Data Transfer services
- EOSC Monitoring
- EOSC Accounting
- EOSC Configuration Management Database (CMDB)
- Collaboration Software
- Operations Portal
- EOSC Security policies and security coordination functions.

Compliance Framework

The Compliance Framework is proposed to comprise

- EOSC Rules of Participation
- EOSC Service Portfolio Management Tool
- EOSC Interoperability Framework
- EOSC Service Management System.

Federating Core Value and Cost Model

Together, the Federating Core and the resources and research outputs provided with the coordination of e-Infrastructures and ESFRI projects and landmarks, would constitute a rich ecosystem which represents a significant part of the Minimum Viable Ecosystem (MVE) proposed by the second EOSC High Level Expert Group[9], to be complemented by resources provided to the EOSC by national research infrastructures. This ecosystem has high potential value for users and other stakeholder groups.

The EOSC needs to sustain the costs of providing the benefits of open data policies to a wider community of users. It needs to create the financial vehicle to cover the costs of the Hub Portfolio and Compliance Framework, and of provision and consumption of the Shared Resources beyond their originating communities. Coordinated provisioning and funding of the Federating Core is expected to bring economies of scale by aligning investments from member states with the compensation of marginal costs associated with cross-border usage of depletable resources and services.

Projects which have contributed to this working draft to date:

EOSC-hub   ENVRI-Fair   EOSC-Life   ESCAPE   FAIRsFAIR   FREYA   NI4OS   PaNOSC   SSHOC

---

[9]   https://publications.europa.eu/en/web/eu-law-and-publications/publication-detail/-/publication/5253a1af-ee10-11e8-b690-01aa75ed71a1

Commented [9]: I'd be happy to know more about this.

Commented [10]: Some more description is included in the first Briefing Paper (reference to which is in footnote 8).

Commented [11]: Is it between researchers or service providers? both?

Commented [12]: Envisaged as support for internal processes: issue management and documentation co-development and sharing. So mainly between service providers. Further detail still to be developed.

Summary of What the EOSC Cluster Projects Requirements Of, and Will Offer To, the EOSC

PaNOSC

What PaNOSC Needs From the EOSC
a common way of identifying, authenticating, and authorising users (AAI) across Europe
a free service for downloading data efficiently (distributed and high bandwidth)
a (commercial or free) solution for long term archiving of large quantities of open data (petabytes) coupled to (commercial or free) high-performance storage and compute resources for the (re)analysis of this open data
a search machine for searching and finding scientific data in a wide variety of domains
a catalogue of (free and commercial) services for analysing data ranging from generic services like Jupyter notebooks to specific applications per scientific domain
What PaNOSC Can Offer to EOSC
petabytes of raw and processed data in a wide variety of scientific domains
tools for generic and specific data simulation and data analysis
recipes and expertise for reducing and analysing data
training material for understanding photon and neutron science

EOSC-Life

What EOSC-Life Needs From the EOSC
AAI
interoperable European clouds: computational resources, including secure, federated cloud computing environments that offer secure access across national boundaries to raw data and interoperable results
common quality management for resources (data and services)
common application programming interfaces (APIs) to enable remote data discovery and access
a repository of tools and services, including workflows used to analyse deposited data while enabling these analysis workflows to cover data across national borders
What EOSC-Life Can Offer to EOSC

publish FAIR life science data in EOSC (subject to suitable data management policies for sensitive data)

an ecosystem of innovative life-science tools in EOSC (tools collaboratory)

ENVRI-FAIR

What ENVRI-FAIR Needs From the EOSC
Generic infrastructuredata and metadata services such as for AAI, PID, and provenance, for tailoring to specific Research Infrastructure needs and adoption by individual RIs
Generic workflow management tools and services, for tailoring to specific Research Infrastructure needs and adoption by individual RIs
Access to shared resources such as repositories, HPC, HTC and data management tools
Common Standard APIs to support remote data discovery, access, and sharing
Provision of notebook-based environments which allow to access and integrate data services for the community

What ENVRI-FAIR Can Offer to EOSC
Collective domain-specific knowledge and competencies that underlie all the data and other services provided by the European ENVRIs
FAIR-based tools and resources for easy and seamless access to environmental data and services provided by the European ENVRIs
ENVRI-hub – a virtual, federated machine-to-machine interface to access environmental data and services provided by the contributing ENVRIs

SSHOC

What SSHOC Needs From the EOSC
A common way of identifying, authenticating and authorising users (AAI) across Europe
Personalised workspaces based on federated AAI: provision of secured environments for temporary storage, accessing and using data, coupled to compute resources for the (re)analysis of these data
A platform (digital marketplace) to host the SSHOC marketplace and integrate it in EOSC
EOSC Helpdesk, training resources and tools to support data processing and analysis
Workflows to support contribution of resources to the SSH/EOSC Marketplace
EOSC Rules of Participation, Interoperability and FAIR Guidelines to govern quality and standards

What SSHOC Can Offer to EOSC
SSH Open Marketplace, offering seamless access to high-quality, free and commercial, data, tools, services (including training), repositories and other resources from the Social Sciences and Humanities community, integrated through the EOSC Marketplace

ESCAPE

What ESCAPE Needs From the EOSC
AAI
EOSC Rules of Participation and Interoperability Guidelines
Common APIs to support remote data discovery and access
A hosting environment for data and community-specific virtual research environments providing specific data analytics capabilities and data products
What ESCAPE Can Offer to EOSC
Publish astronomy, astroparticle and particle physics data in EOSC as part of a Data Lake concept including multi-Exabyte datasets
Science Analysis Platforms for EOSC researchers to stage data collections, analyse them, access ESFRIs' software tools and bring their own custom workflows
Scientific software for data analysis, enrichment and comparison
Contribution to continuous shared software development and training