



Briefing Paper - EOSC Federating Core v2.0

Status	Final
Version	0.3
Delivery Date	26/11/2019 18/02/2020
Dissemination Level	Public

About this paper

This Briefing Paper contains an update to the initial proposals for the organisation and composition of the EOSC Federating Core which were originally published in July 2019, based on feedback received from the research community. It also contains recommendations for an overall vision for the EOSC, its funding model and its value proposition.

~~This file is made available for consultation. Comments are welcome,~~

~~This document was made available for consultation during late November and December 2019. This version contains the feedback received, and responses to the comments made.~~

The consultation document itself is available to view at <http://tiny.cc/FedCoreBPv2>.

Formatted: Strikethrough

Formatted: Font: (Default) Calibri, Strikethrough

Formatted: Font: (Default) Calibri, Font color: Auto



Delivery slip

From	<p>Editor: Dale Robertson, Jisc</p> <p>Contributors and Reviewers: Tiziana Ferrari, EGI; Per Oster, EUDAT; Davide Salomoni, INDIGO-DataCloud</p>
-------------	--

COPYRIGHT NOTICE



This work by Parties of the EOSC-hub Consortium is licensed under a Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>). The EOSC-hub project is co-funded by the European Union Horizon 2020 programme under grant number 777536.

GLOSSARY

<https://wiki.eosc-hub.eu/display/EOSC/EOSC-hub+Glossary>

<i>Terminology/Acronym</i>	<i>Definition</i>
Compliance Framework	The Rules of Participation, the Interoperability Framework, the Service Management System and other policies and processes for the demand side and the supply side to engage with EOSC. The Compliance Framework constitutes the EOSC “ regulatory tier ”.
EGI Federation	A federation of computing and storage resource providers united by a mission to support research and innovation with advanced computing services. The federation is governed by the participants represented in the EGI Council and coordinated by the EGI Foundation.
EOSC	The European Open Science Cloud promoted by the European Commission to provide all researchers, innovators, companies and citizens with seamless access to an open-by-default, efficient and cross-disciplinary environment for storing, accessing, reusing data, tools,

	publications and any EOSC Resource for research, innovation and educational purposes ¹ .
EOSC Executive Board	Body of representatives from the research and e-infrastructures communities, appointed by the European Commission ²
EOSC Governance	Overall Governance Structure for EOSC, comprising EOSC Governance Board, EOSC Executive Board and Stakeholder Forum (latter not yet specified)
EOSC Governance Board	Also “EOSC board”: institutional group gathering the member states and the Commission to ensure effective supervision of the implementation
EOSC-hub	Project creating the integration and management system of the future European Open Science Cloud
EUDAT CDI	European e-infrastructure of integrated data services and resources to support research
FAIR	Guiding principles to make data Findable, Accessible, Interoperable, and Reusable
Federating Core	The activities, policies and resources required to facilitate, monitor and regulate as appropriate day-to-day transactions across the EOSC. The Federating Core is made up of the Hub Portfolio, the Compliance Framework and the Shared Resources.
GÉANT	Pan-European research and education network that interconnects Europe’s National Research and Education Networks (NRENs)
Horizon 2020	The European Union Framework Programme for Research and Innovation
Hub Portfolio	Functions which implement the federating tier of the EOSC: the activities and tools that are necessary to provide coordinated access to and management of resources (services and scientific products) 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 ”.

¹ <https://eosc-pilot.eu/eosc-glossary#overlay-context=eosc-glossary>

² <https://www.eoscsecretariat.eu/eosc-governance/eosc-executive-board>

INDIGO-DataCloud	Project developing a data/computing platform targeted at scientific communities, deployable on multiple hardware, and provisioned over hybrid (private or public) e-infrastructures
OpenAIRE-Advance	Project supporting Open Access/Open Data mandates in Europe
PRACE	Partnership for Advanced Computing in Europe
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 ".

Contents

Executive Summary	7
1. Introduction	10
2. Summary of Feedback Received	11
2.1 EOSC Requirements and Opportunities	11
2.2 Federating Core	14
2.3 Governance and Sustainability	14
3. Updated Proposals for the Organisation and Composition of the Federating Core	15
3.1 Introduction	15
3.2 Organisation of the EOSC	15
3.3 Recommendations for the Federating Core	16
3.3.1 Shared Resources	16
3.3.2 Hub Portfolio	18
EOSC Portal	18
EOSC Support Services	19
EOSC Data Transfer services	19
EOSC Security policies and security coordination functions	20
3.3.3 Compliance Framework	20
3.4 EOSC Service Portfolio	21
3.5 Summary	21
4. Recommendations for the EOSC	22
4.1 EOSC Opportunities	22
4.2 EOSC Funding Model	23
4.3 EOSC Added Value (Incentives)	24
5. Concluding Remarks	24
I.1 EOSC Requirements and Opportunities	26
I.2 Federating Core	26
I.2.1 Shared Resources	27
I.2.2 Hub Portfolio	27
I.2.3 Compliance Framework	28
I.3 Access, Governance and Sustainability	28
II.1 PaNOSC	30

II.2	EOSC-Life	31
II.3	ENVRI-FAIR	32
II.4	SSHOC	33
II.5	ESCAPE	34

Executive Summary

The EOSC-hub project published initial proposals for the organisation and composition of the EOSC Federating Core and its governance and sustainability in a Briefing Paper in July 2019. The proposals have now received feedback from the research community including FAIRsFAIR, FREYA, GÉANT, INCD and LIP, NI4OS-Europe and SSH Open Marketplace (DARIAH-ERIC), and from the five EOSC cluster projects ENVRI-FAIR, EOSC-LIFE, ESCAPE, SSHOC and PaNOSC. The cluster projects contributed input grounded in the experience and requirements of their respective user communities, to jointly define a clear and complete value proposition for EOSC (see use cases in boxes).

Based on the vision which has emerged, *the ability of the EOSC to provide quality assessment, combined with the ability for researchers to access and reuse data alongside Shared Resources through the same portal, represents the unique value-add of the EOSC and provides the incentive to supply to and use the EOSC.*

According to the provided use cases, EOSC emerges as the initiative which enables the reuse of research communities' FAIR data and data analytics tools, while sustaining the costs of providing the benefits of open data policies to a wider community of users, principally by widening access to data produced, curated and preserved by national and European research communities.

The EOSC should provide a universal and versatile discovery space for all data and tools, offer inclusive and transparent policies for access, and enable data and software sharing and exploitation tools, including general services for data management and processing. The coordinated provisioning of capabilities that are generally applicable to data lifecycle management in different disciplines would deliver economies of scale and avoid double funding if provided in a coordinated manner in EOSC. The EOSC needs to create the format to cover the costs of this ecosystem. In the interests of delivering opportunities and efficiencies, the EOSC needs to interoperate with other infrastructure initiatives in Europe and other regions, including EuroHPC³.

Research communities and infrastructures are the main custodians of research data, quality and FAIR policies. The EOSC should complement what is provided by research institutions and communities according to the subsidiarity principle, and provide researchers with a personal space where they can use EOSC resources collaboratively.

The relationship of national infrastructures to the EOSC - their role in the EOSC, and how they contribute - is very important. The EOSC needs to consider and address the requirements of the national user communities.

The proposed definition of the Federating Core has been updated slightly with the benefit of the feedback received, and in line with the vision expressed by the EOSC cluster projects. The updated

³ <https://eurohpc-ju.europa.eu/>

definition is contained in the box below. It will remain under review as the further definition of the EOSC progresses, and as its proposed functions are defined in more detail in the coming months.

Together, the Federating Core and the data, tools and research outputs provided with the coordination of e-Infrastructures and ESFRI projects and landmarks, would constitute a rich ecosystem which would make a very significant contribution towards the “critical mass” notion recommended by the ~~first EOSC High Level Expert~~ first EOSC High Level Expert Group⁴. They could be regarded as constituting a significant part of the Minimum Viable Ecosystem (MVE) proposed by the second EOSC High Level Expert Group⁵, 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:

For researchers, the EOSC would provide visibility of and access to data and other research resources (software, tools) - many of which at present may be out of reach to them – presented in catalogues/marketplaces and “quality assessed” by the EOSC Rules of Participation, plus access to Shared Resources for common user needs such as storage and compute, as well as access to new communities of researchers.

For service providers, the EOSC would offer greater visibility of their services and the outputs they produce, presented in EOSC catalogues/marketplaces and “quality assessed” by the EOSC Rules of Participation. This would support their efforts to make their outputs FAIR.

For funders, it would deliver cost savings through economies of scale, and would avoid double-funding.

The **Federating Core** of the EOSC is a fundamental asset of the EOSC, composed of the activities, policies and resources 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 resources⁶ 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**”.

⁴ https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf

⁵ <https://publications.europa.eu/en/web/eu-law-and-publications/publication-detail/-/publication/5253a1afee10-11e8-b690-01aa75ed71a1>

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>

Commented [1]: delete

Commented [2]: Noted, thank you.

Commented [3]: Can we add that EOSC will also provide to researchers access to other networks? researchers? in order to develop their own network in Europe - collaboration platform for instance? And discovery tool to identify researchers...

Commented [4]: Thank you. This has been added, and a couple of corresponding small edits made to sections 4.1 and 4.3.

(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 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 Participation⁷, 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.

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).*

1. Introduction

In July 2019, the EOSC-hub project published initial proposals on the EOSC Federating Core in the context of its commitment within Task 2.3 (Governance and Sustainability) to develop and implement a Governance and Sustainability Roadmap for collaboration between EGI, EUDAT and INDIGO-DataCloud beyond the end of 2020. The proposals presented the capabilities the EOSC Federating Core should deliver and how they should be organised, as well as describing the wider EOSC Service Portfolio, and proposing how the Federating Core should be governed and sustained.

The Briefing Paper was widely advertised, and supported by a public webinar which had more than forty attendees. Feedback on its proposals has been gathered from an online consultation, and from discussing the proposals in several meetings and forums. The online consultation received five responses. Although the number of responses was disappointing, they were comprehensive and provided very helpful input and insights.

The proposals have also been discussed to date in the following meetings and forums:

- Discussion with FREYA project, 7 August
- Discussion with FAIRsFAIR project, 8 August
- Meeting with EC DG CNCT e-Infrastructures Unit, 9 August
- EC EOSC cross-project Concertation meeting, 9-10 September
- Science Business Cloud Consultation Group meeting, 11 September
- EOSC-hub WP7 (Thematic Services: Integration, Maintenance and Exploitation) meeting, 20 September
- EOSC-hub Strategy Board meeting, 25 September

In addition, the Sustainability Working Group of the EOSC Executive Board provided a response to the Briefing Paper in the form of their Strawman paper "Solutions for a Sustainable EOSC", published in early September 2019.

This paper summarises and discusses the feedback received on the initial proposals and presents recommendations as a result. Its focus is on the organisation and composition of the Federating Core rather than its governance and sustainability, although the governance and sustainability proposals from the Briefing Paper are briefly discussed in section 4.2. It is intended as a further contribution to the ongoing implementation of the EOSC and is addressed to the EOSC stakeholder community in general and in particular the EOSC implementation projects and the EOSC governance bodies.

In the following Chapters, Chapter 2 summarises the feedback received, Chapter 3 presents updated proposals for the definition and composition of the EOSC Federating Core, Chapter 4 makes recommendations for the vision, value-add and business model of the EOSC, and Chapter 5 provides some brief concluding remarks.

2. Summary of Feedback Received

This chapter provides a brief summary of the feedback received on the initial proposals in the Briefing Paper. The feedback is described in more detail in Appendix I. The cluster projects' views of the opportunities the EOSC represents, and their offers to and requirements of the EOSC, are also described in Appendix II.

Input was received from FAIRsFAIR, FREYA, GÉANT, INCD and LIP, NI4OS-Europe and SSH Open Marketplace (DARIAH-ERIC), and from the five EOSC cluster projects ENVRI-FAIR, EOSC-LIFE, ESCAPE, SSHOC and PaNOSC, as well as through discussion at the meetings listed in Chapter 1.

2.1 EOSC Requirements and Opportunities

The ESFRI cluster projects (ENVRI-FAIR, EOSC-LIFE, ESCAPE, PaNOSC, SSHOC) contributed input grounded in the experience and requirements of their respective user communities, to jointly define a clear and complete value proposition for EOSC (see use cases in boxes on the following pages).

According to the provided use cases, EOSC emerges as the initiative which enables the reuse of research communities' FAIR data and data analytics tools, while sustaining the costs of providing the benefits of open data policies to a wider community of users, principally by widening access to data produced, curated and preserved by national and European research communities.

This vision of the EOSC should provide a universal and versatile discovery space for all data and tools, offer inclusive and transparent policies for access, and enable data and software sharing and exploitation tools, including general services for data management and processing. The coordinated provisioning of capabilities that are generally applicable to the data lifecycle management in different disciplines would deliver economies of scale and avoid double funding if provided in a coordinated manner in EOSC. In the interests of delivering opportunities and efficiencies, the EOSC needs to interoperate with other infrastructure initiatives in Europe and other regions, including EuroHPC⁸.

The EOSC should deliver equality of access to data and other resources for all researchers, regardless of their institution or geographical location.

Research communities and infrastructures should remain the main custodians of research data, quality and FAIR policies. EOSC should complement what is provided by research communities according to the subsidiarity principle, and provide scientists with a personal space where they can create content (data analysis recipes, workflows, publications), store data temporarily and share their work with collaborators.

The relationship of national infrastructures to the EOSC - their role in the EOSC, and how they contribute - is very important. The EOSC needs to consider and address the requirements of the national user communities.

⁸ <https://eurohpc-ju.europa.eu/>

Whilst for many research projects the EOSC may not be *essential* to achieving scientific goals, it could provide the framework to guarantee opening access to their outputs (datasets, software, publications) and results for a wider community, thus delivering additional value.

PaNOSC

PaNOSC is the INFRAEOSC-04 cluster project of ESFRI Photon and Neutron sources. It represents a plethora of scientific disciplines using photon and neutron sources to study materials on a wide range of scales - from angstroms to microns. A large user community of roughly 30,000 users annually profit from the Photon and Neutron sources in Europe (this includes the national and ESFRI sources).

Some of the PaNOSC Research Infrastructures have databases of data collected over the last 3 to 4 decades which are currently under-exploited due to lack of bandwidth to the repositories where they are stored, and due to lack of awareness of their existence. For example, palaeontology data in the repository at <https://paleo.esrf.fr> is an example of processed (curated) data which are not widely known or exploited yet. These data are ideal for cross-disciplinary applications and for linking up with data from museums and/or other scientific disciplines. The EOSC could improve the findability and accessibility, in particular, of data such as these to allow their potential to be tapped.

PaNOSC sees EOSC as an opportunity to generalise the adoption of FAIR data policies at all photon and neutron sources. Adopting FAIR data will enable data sharing across a wider community and the provisioning of services for remote data analysis.

EOSC-Life

EOSC-Life is the Life Sciences INFRAEOSC-04 cluster project. EOSC-Life brings together the thirteen Biological and Medical ESFRI Research Infrastructures to create an open, digital and collaborative space for biological and medical research.

EOSC-Life sees the EOSC as an opportunity to realise the benefits of open data policies and make access to data and resources equal for researchers no matter where they are located, and also as a route to access to the cloud for Life Sciences communities such as tool developers.

ENVRI-FAIR

ENVRI-FAIR is the connection of the Environmental Research Infrastructure community to the European Open Science Cloud, building FAIR services accessible for society, innovation and research. ENVRI-FAIR brings together all thirteen environmental domain Research Infrastructures from the ESFRI Roadmap but collaborates with the entire community of Environmental Research Infrastructures.

ENVRI-FAIR is working to create joint data service chains to allow the combined analysis of data and other resources from multiple sources. A major challenge is applying the FAIR principles to data and services. The EOSC has the potential to provide important support to the metadata and access preparation stages of ENVRI's work to make data and services FAIR, for example by providing templates for assigning provenance to data, and facilitating access to flexible, scalable research data and related services.

SSHOC

The Social Sciences and Humanities Open Cloud (SSHOC) cluster project unites twenty partner organisations and a further 27 associates to create the social sciences and humanities area of the European Open Science Cloud. It aims to effect the transition from the current data landscape with its disciplinary silos and separate facilities to an integrated, cloud-based network of leveraged and interconnected data infrastructures supported by the tools and training to allow scholars and researchers to access, process, analyse, enrich and compare data across the boundaries of individual repositories or institutions.

SSHOC aims to maximise the efficiency and effectiveness of data re-use by applying Open Science practices and FAIR principles to data management. It views EOSC as an opportunity to achieve this by providing visibility (findability) for resources, together with guidelines governing interoperability and FAIRness, supported by helpdesk, workflows, workspaces, training and tools to assist researchers to make use of the resources made available to them.

ESCAPE

The European Science Cluster of Astronomy and Particle physics ESFRI research infrastructures, ESCAPE, aims to support the implementation of the EOSC through the expertise of the astronomy, astroparticle and particle physics communities. The cluster contributes two major areas of expertise in data stewardship: the astronomy Virtual Observatory infrastructure, and the expertise of the particle physics community in large-scale distributed computing and big-data management. A major challenge faced by ESCAPE in

attempting to FAIRify data is how to make available the very large quantities of data produced by their community.

ESCAPE plans to implement Science Analysis Platforms for EOSC researchers to stage data collections, analyse them, access ESFRIs' software tools and bring their own custom workflows. It also plans to contribute to EOSC global resources federation by implementing a Data Lake concept to manage extremely large data volumes at the multi-Exabyte level. In addition, ESCAPE plans to expose scientific software through dedicated EOSC catalogues, and regards continuous shared software development and training as a significant opportunity for value-add for the EOSC.

2.2 Federating Core

There was not a complete consensus on the make-up or position of the **Shared Resources** in the EOSC from all of those providing feedback, but the majority opinion was that they should not contain research data. A clear view was expressed that data should continue to be owned, curated, preserved and funded through existing funding from research projects and infrastructures. The Shared Resources element of the Federating Core was seen as having the potential to add a great deal of value by enabling access to data, software and other resources by providing a resource tier for data hosting through storage and computing, and other generic capabilities and resources of broad interest. Further discussion about the position of the Shared Resources in the EOSC is in Chapter 3.

The initial proposed composition of the **Hub Portfolio and Compliance Framework** was confirmed by the feedback, but focussed attention on the need to define the value-add of the EOSC so as to clarify the incentive for users to consume resources through the EOSC rather than discovering them individually and consuming them "direct from source" as now. The composition of the Hub Portfolio and Compliance Framework has been added to based on feedback received but further functions are still expected to be added from other EOSC implementation projects.

2.3 Governance and Sustainability

There was broad agreement to the proposals in the Briefing Paper for the governance and sustainability of the EOSC, which were that the Federating Core should be governed and sustained by the EOSC, and that services in the EOSC Service Portfolio should continue to be independently owned by their respective providers. It was observed, however, that some services (e.g. PID assignment services) are fundamental to the EOSC but will not be governed or owned by EOSC. For a service to belong in the Federating Core is not just a question of being fundamental to the EOSC, but whether it is answerable to the EOSC and whether the EOSC should provide sustainability of the service.

Views on the governance and sustainability of the Shared Resources element of the Federating Core were less clear due to the lack of a clear definition of the Shared Resources in the Briefing Paper.

3. Updated Proposals for the Organisation and Composition of the Federating Core

3.1 Introduction

The initial proposals in the Briefing Paper were based on the EC's description of the EOSC model and analysis of recommendations from the EOSC High-Level Expert Group final report and from the EOSCpilot project Science Demonstrator recommendations and requirements. They drew on the operational experience of EOSC-hub project partners managing federating infrastructures. They were also derived from use cases emerging from research communities both within and outside of the EOSC-hub consortium including ESFRI projects and landmarks and international communities of practice, bringing input from a broad range of scientific disciplines. The proposals have now received feedback from the research community including the five EOSC cluster projects and several other EOSC implementation projects.

As summarised in Chapter 2, overall the feedback received supported the organisation and content initially proposed for the Federating Core whilst helping to clarify some points about its composition. The feedback has been used to provide updated proposals which are described in this Chapter, although the original Briefing Paper⁹ should be consulted for a more detailed description of the EOSC components.

A clearer picture of the possible value-add of the EOSC and of its overall model also emerged from discussions. This is described in Chapter 4.

The focus of the proposals in this paper is on the Federating Core for the initial implementation of the EOSC. This corresponds to the Minimum Viable Product concept. Further functions may need to be added over time to cater for the needs of additional categories of services or users.

3.2 Organisation of the EOSC

The overall organisation of EOSC resources into the Federating Core and the Service Portfolio is unchanged. The Federating Core is proposed as a fundamental asset of the EOSC, providing the technical, human, policy and resource elements required to allow the research-targeted services to operate. It should consist of three elements: the Hub Portfolio, the Compliance Framework and the Shared Resources. The Shared Resources are still proposed to be part of the Federating Core due to their envisaged purpose and proposed funding and governance. The position of the Shared Resources in the EOSC is closely related to the value proposition of the EOSC. The value proposition is discussed in Chapter 4.

The Federating Core should be kept as lightweight as possible, but this must be balanced against the need for the EOSC to provide additional value compared to the current landscape. The Federating Core is still proposed to be governed and funded by the EOSC – that is, the EOSC governance should

⁹ Available from <https://www.eosc-hub.eu/sites/default/files/EOSC-hub%20Briefing%20Paper%20-%20EOSC%20Federating%20Core%20Governance%20and%20Sustainability%20Public.pdf>

be responsible for the definition and funding of the Federating Core. The precise distribution of responsibilities for this within the EOSC governance structures is not discussed in this paper.

To clarify the relationships between the different elements: the Service Portfolio would be subject to some, but not necessarily all, of the rules and policies from the Compliance Framework; the Hub Portfolio, Compliance Framework and Shared Resources are elements of the Federating Core but any one of these three is not a subset of any of the others. Figure 1 below attempts to demonstrate the elements more clearly than previously.

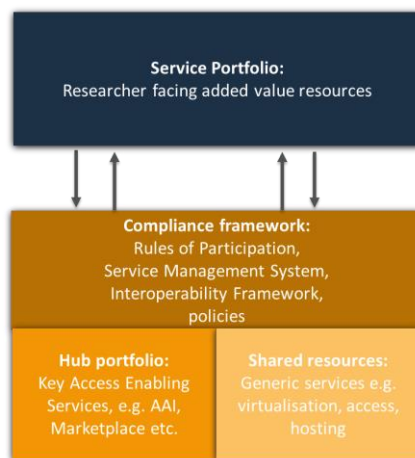


Fig. 1 – Updated representation of the EOSC Resources, organised into two portfolios: the EOSC Federating Core (yellow) and the EOSC Service Portfolio (blue)

3.3 Recommendations for the Federating Core

3.3.1 Shared Resources

In the initial proposals in the July Briefing Paper, the Shared Resources were envisaged to comprise resources of pan-European relevance which are developed by a given discipline but made available, by inclusion in the Shared Resources, to be used more broadly by external user communities and additional disciplines. It was proposed that the EOSC would provide access to such generic capabilities and resources in a more coordinated and cost-effective manner, avoiding duplication, than is possible at present, and that the Shared Resources would provide services such as those for data management and processing, hosted by a generic resource tier of storage and computing and scientific products. Several use cases were provided to illustrate how the Shared Resources may benefit EOSC users. The Shared Resources would be provided by existing Research Infrastructures and e-Infrastructures and other service providers.

Following the consultation, the original concept for the Shared Resources remains, with the notable clarification that the Shared Resources should not “own”, produce or guarantee the quality of data; the EOSC should provide access to data and should assess the quality of characteristics of datasets

such as metadata and PIDs (persistent identifiers), but data should continue to be owned, hosted and funded as now. The Shared Resources could however host a local copy of data downloaded by a researcher.

The Shared Resources have the potential to support researchers in data processing, analysis, enrichment and comparison. The services 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

The role of EuroHPC in the EOSC Shared Resources and/or Service Portfolio remains to be clarified. It is very likely that a closer relationship between the EOSC and the EuroHPC initiative would deliver greater benefit to users and may also provide cost efficiencies.

EOSC-hub continues to believe that the Shared Resources should be part of the EOSC Federating Core. The Briefing Paper's initial proposals were based on recommendations and use cases which demonstrate that communities wishing to offer large data archives and other resources to the EOSC see the need for broader and more equal access for researchers to generic services including data storage and compute as a means to realise the potential of the community-specific resources on offer. This testimony has since been further added to by the use cases of the five EOSC cluster projects (see Appendix II). Placing the Shared Resources inside the Federating Core positions them as a key part of the EOSC and provides the means for its resources to be governed and funded separately from resources provided as part of the EOSC Service Portfolio, opening the way for them to be made available and sustained in the best manner for the EOSC to deliver the most value to its users and providers.

3.3.2 Hub Portfolio

The purpose of the functions in the Hub Portfolio is to facilitate the combined exploitation of the Shared Resources and EOSC Service Portfolio; whether this is described as *enabling federation* or *providing an interoperable integrating layer* is perhaps a moot point: numerous use cases were presented in the Briefing Paper and these are supplemented here in Appendix II with the expressed desires and needs of the five EOSC cluster projects. Such use cases help to prove the demand for and value of the EOSC, and the Hub Portfolio should provide the necessary functions to support them.

The services and functions envisaged to make up the Hub Portfolio remain the same as those initially proposed, constituting “back office” functions necessary to enable access to services and support the federation of resources through the EOSC, although it should be clarified that the Hub Portfolio is intended to provide federation-enabling functions for services such as AAI, rather than providing actual AAI services.

The EOSC Portal is not proposed to be exclusive or to replace existing access routes to resources, but is intended to provide a universal entry point to access resources through the EOSC, boosting discoverability and accessibility. This may deliver an important part of the added value of the EOSC.

The EOSC AAI complements and supports the EOSC Portal to enable seamless access to research data and services in EOSC in a secure and user-friendly way. Research communities can leverage the EOSC AAI services for managing their users and their respective roles and other authorisation-related information. At the same time, the adoption of standards and open technologies, including SAML 2.0, OpenID Connect, OAuth 2.0 and X.509v3, facilitates interoperability and integration with the existing AAI of other e-Infrastructures and research communities.

The list of functions proposed to constitute the Hub Portfolio includes all those originally proposed but has been added to as a result of feedback received:

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

Functions which have been expanded and new functions which have been added to the proposed Hub Portfolio are described below.

EOSC Portal

A detailed description of the Portal was provided in the initial Briefing Paper and is provided again here due to the significance of the Portal to the EOSC: the EOSC Portal provides a European-level

delivery channel connecting the demand-side (the EOSC Customers) and the supply-side (the EOSC Providers) to allow researchers to conduct their work in a collaborative, open and cost-efficient way for the benefit of society and the public at large. In particular it delivers the following functions:

- Enable different kinds of users, with different skills and interests, to discover, access, use and reuse a broad spectrum of EOSC Resources (services, datasets, software, support, training, consultancy, etc) for advanced data-driven research
- Support interdisciplinary research and facilitate Resource discovery and access at the institutional and inter-institutional level
- Allow researchers and institutions to focus on value creation through sharing and reuse as opposed to duplicating Resources and increase excellence of research and European competitiveness
- Improve the provisioning of access to integrated and composable products and services from the EOSC Catalogue
- Facilitate the composition of services and products to support multi-disciplinary science for example with high-level community-specific interfaces for running workflows involving EOSC services
- Help Providers gain additional insight into potential Users outside their traditional constituencies
- Give Providers the possibility to offer Resources under homogeneous terms of use, acceptable use policies, and in different configuration options, so that Users are guided in the choice.

During consultation and feedback, a clear desire was expressed for the EOSC to provide a catalogue of datasets, and a catalogue of APIs to enable remote data discovery and access. Discussion of PID services led to recognition that the PID Services Registry currently under development in the FREYA project has potential to add significant value to the EOSC and could be made accessible through the EOSC Portal. A library or repository of code to support semantic interoperability of FAIR could also perhaps be provided through the EOSC Portal.

EOSC Support Services

The EOSC Support Services were proposed in the original Briefing Paper as technical support, training and advice on specific policy and technical aspects and has been expanded here to mention competence centres and knowledge bases. Feedback shows that there is felt to be a widespread need for training and expertise to encourage and support EOSC users in their use of resources accessed through the EOSC. The definition and composition of the support services to be included in the Federating Core in particular will require to be further developed, and is very likely to evolve significantly over time.

EOSC Data Transfer services

Data Transfer services enable the movement of data files asynchronously between source and destination storage endpoints, including mechanisms to ensure automatic retry in case of failure and for optimisation of performance for large files or large numbers of files.

EOSC Security policies and security coordination functions

Central coordination of security activities ensures that Trust and Identity policies, operational security, and maintenance are compatible amongst all partners, providing monitoring services to check for security vulnerabilities and other security-related problems in the infrastructure: it guarantees that incidents are promptly and efficiently handled, that common policies are followed by providing services such as security monitoring, and by training and dissemination with the goal of improving the response to incidents. A harmonised EOSC Acceptable Use Policy (AUP) will be important and necessary for collaboration between Infrastructures as a single common baseline AUP, shared between Infrastructures, making user registration simpler¹⁰.

Further services or functions are expected to be added to the Hub Portfolio as definition of the EOSC progresses, notably as a result of collaboration with OpenAIRE to define a joint EOSC-hub-OpenAIRE service proposition, but also from contributions of other EOSC implementation projects. The proposed contents of the Hub Portfolio will also be reviewed as more detailed definition of its proposed functions is undertaken in the coming months.

3.3.3 Compliance Framework

The Compliance Framework is still envisaged to represent the policies and processes required to operate the Federating Core, the main vehicles of which are the Rules of Participation, the EOSC-hub Service Management System and the Interoperability Guidelines for thematic and common services. The Compliance Framework will also be responsible for standards compliance and quality assurance.

The proposed concept of the Compliance Framework has not been altered as a result of the feedback and discussion to date of the proposals, but additional components have been proposed for inclusion including certification of data repositories and datasets, a minimum metadata set for catalogues, and an Open Science Policy Monitor. It is expected that further detail will be specified and a number of further policy aspects will be added from other EOSC implementation projects.

The additional administrative burden imposed on service operators by EOSC accounting, configuration management database, service portfolio management tool and service management system will be kept to a minimum, however these functions are proposed in response to recommendations made by the EOSCpilot Science Demonstrators, and are required to support high levels of performance (e.g. availability and quality) for the seamless services to be offered through the EOSC, and to provide reporting on usage which is necessary information for EOSC users, providers and managers. The business processes regulated by the EOSC will be clearly defined to apply to services and resources that are part of the Federating Core. Providers of the EOSC Service Portfolio will be offered the possibility to opt for different integration scenarios depending on the level of service integration they would like to deliver in EOSC and conformance to its interoperability and quality best practices.

¹⁰ EOSC-hub is conducting ground-breaking work demonstrating the feasibility and advantages of a harmonised AUP (<https://www.eosc-hub.eu/training-event/it-security-management-ism-eosc-hub-policies-and-global-trust>).

The Compliance Framework is proposed to comprise the following functions, which were described in the initial Briefing Paper¹¹:

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

3.4 EOSC Service Portfolio

The EOSC Service Portfolio provides additional added-value services (common and thematic) which exploit the Federating Core and provide complementary capabilities to EOSC users. Services in the Service Portfolio are discoverable, selected, customised and instantiated through the EOSC Portal, to address the needs of specific user communities. They are envisaged to be typically focused on a specific capability in the data lifecycle management process, rather than being full general-purpose infrastructures as envisaged for the Shared Resources.

Services in the EOSC service portfolio should continue to be independently managed and governed, as they are at present.

3.5 Summary

The overall capabilities of the Federating Core remain similar to those proposed in the Briefing Paper in July 2019.

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:

- (4) **Hub Portfolio**: The activities and tools necessary to provide coordinated access to and management of resources¹² 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**”.
- (5) **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**”.
- (6) **Shared resources**: Resources including scientific outputs (local copies of data; applications, software, pipelines etc.) and the storage and compute hosting platforms

¹¹ <https://www.eosc-hub.eu/sites/default/files/EOSC-hub%20Briefing%20Paper%20-%20EOSC%20Federating%20Core%20Governance%20and%20Sustainability%20Public.pdf>

¹² 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>

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 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 Participation¹³, 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.

4. Recommendations for the EOSC

The vision for the EOSC put forward by the EC is that it should federate ~~the European data e-~~ ~~infrastructures and other resources~~ to support and deliver Open Science for European science outputs¹⁴. In attempting to define the Federating Core however, it is necessary to consider what problem(s) the EOSC should solve, how, and what are the incentives to supply resources to it and to consume resources through the EOSC rather than directly. The feedback provided showed that it is important to define the value-add of the EOSC, distinct from the value which can be provided by services already available to the research community.

The feedback provided on the Briefing Paper, particularly that from the five EOSC cluster projects, has helped to clarify views about the EOSC and articulate the opportunities it represents, how it should be funded and the value it adds to its three main stakeholder categories of users, service providers and funders. This information is included here due to its relevance to the organisation and definition of the Federating Core, and to the broader work of EOSC-hub Task 2.3. The proposals presented below are primarily, although not exclusively, based on discussion with the EOSC cluster projects.

4.1 EOSC Opportunities

The EOSC is an opportunity to generalise the adoption of FAIR data policies for the benefit of a wide community of users, principally by widening access to data produced and stored by national and European research communities.

The EOSC should provide the framework to guarantee opening access to research outputs (datasets, software, publications) for a wider community. It should provide a universal and versatile discovery space for resources, offer inclusive and transparent policies for access for all researchers regardless of

¹³ 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).

¹⁴ The EC Staff Working Document “Implementation Roadmap for the European Open Science Cloud”, defined the EOSC model as “a pan-European federation of data infrastructures built around a federating core and providing access to a wide range of publicly funded services supplied at national, regional and institutional levels, and to complementary commercial services” (see Section 2.1).

Commented [5]: and Research Infrastructures (ESFRIs)

Commented [6]: Thank you. I have inserted a small edit.

their institution or geographical location, and provide for common user needs for generic storage and processing facilities for data management and analysis, such as high-performance and high-throughput distributed compute capabilities, for researchers to manipulate resources to which they have been afforded access via the EOSC. Researchers would thus also extend their networks of collaboration with other researchers.

The EOSC should add value to what is already available, complementing 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.

4.2 EOSC Funding Model

The Federating Core as proposed above supports the objectives of the five EOSC cluster projects to make their data and other outputs available through the EOSC. Together, the Federating Core and the data, tools and other research resources provided with the coordination of e-Infrastructures and ESFRI projects and landmarks would constitute a rich ecosystem which would make a very significant contribution towards the “critical mass” notion recommended by the first EOSC High Level Expert Group¹⁵. They could be regarded as constituting a significant part of the Minimum Viable Ecosystem (MVE) proposed by the second EOSC High Level Expert Group¹⁶ - to be complemented by resources provided to the EOSC by national research infrastructures. The services required from the EOSC Shared Resources differ from one discipline or community to another. The EOSC needs to create the format to cover the costs of this ecosystem. Some initial proposals for how this may be achieved are included below, but funding and charging mechanisms still need to be elaborated in more detail. Comparison with the EuroHPC model may be informative.

The EOSC Federating Core is proposed to be governed and funded by the EOSC governing bodies. The EOSC should fund the costs of providing the benefits of open data policies to a wider community of users. Ownership and use of data and resources should be separated.

The Research Infrastructures are already funded, often with co-funding from national and European funds. The co-funding model could continue to be used as it is now for national and European research communities. Research Infrastructures and others should ensure their project funding includes the cost of storage and processing of data produced by their users

Allowing access to data through the EOSC would require charging on a cost recovery basis: payment could be made by the EOSC for access to and re-use of resources from outside of users’ funded projects. The generic storage and processing facilities of the Shared Resources could be used by researchers to manipulate resources to which they have been afforded access via the EOSC; the cost of longer-term storage should be borne by the researcher’s home institution.

This approach would deliver economies of scale and avoid double funding, providing significant cost savings for funders. To support it, the EOSC Rules of Participation should include a Memorandum of

¹⁵ https://ec.europa.eu/research/openscience/pdf/realising_the_european_open_science_cloud_2016.pdf

¹⁶ <https://publications.europa.eu/en/web/eu-law-and-publications/publication-detail/-/publication/5253a1af-ee10-11e8-b690-01aa75ed71a1>

Understanding from the EOSC Governing Board which commits participants of all legal entities in a project consortium to EOSC eligibility criteria, which define access policies to be met by user communities, and cost recuperation conditions and rules for providers. The mechanism to recover the costs paid by the EOSC needs to be determined.

4.3 EOSC Added Value (Incentives)

The MVE - or part of it - described above has high potential value for users and other stakeholder groups.

For researchers, the EOSC would provide visibility of and access to data and other research resources (software, tools) - many of which at present may be out of reach to them – presented in catalogues/marketplaces and “quality assessed” by the EOSC Rules of Participation, plus access to Shared Resources for common user needs such as storage and compute, as well as access to new communities of researchers.

For service providers, the EOSC would offer greater visibility of their services and the outputs they produce, presented in EOSC catalogues/marketplaces and “quality assessed” by the EOSC Rules of Participation. This would support their efforts to make their outputs FAIR.

For funders, it would deliver cost savings through economies of scale, and would avoid double-funding.

The quality assessment aspect, plus the ability to access and reuse data alongside Shared Resources through the same portal, which represents the unique value-add of the EOSC and provides the incentive to supply to and use the EOSC.

5. Concluding Remarks

The initial proposals presented in the Briefing Paper in July 2019 were based on the EC’s proposed model for the EOSC and analysis of recommendations from the EOSC High-Level Expert Group final report and from the EOSCpilot project recommendations and requirements. They drew on the operational experience of EOSC-hub project partners managing federating infrastructures and were also derived from use cases emerging from research communities both within and outside of the EOSC-hub consortium including ESFRI projects and landmarks and international communities of practice, bringing input from a broad range of scientific disciplines.

The Briefing Paper was widely advertised and feedback received from the community since July, from the consultation and discussions, has helped to broaden the range of input on which the proposals are based, although the voices of national academic institutions are not directly represented in the feedback received, and the proposals here perhaps reflect the perspective of the European research infrastructures and big data users and producers more than that of national infrastructures and institutions. The feedback served to confirm the initial proposals for the overall organisation and composition of the Federating Core in three elements (Hub Portfolio, Compliance Framework and Shared Resources), and helped to clarify and refine the proposals for the composition of each of these three elements particularly the Shared Resources. Discussion also however highlighted the need to

define and clearly articulate the purpose of the EOSC and the value it will add compared to the current state of play. This influences and is influenced by definition of the Federating Core and its governance and funding. This update to the initial proposals has therefore described a proposed vision for the EOSC and possible value statements, which emerged from some of the discussions and are relevant to the wider context within which the work on the Federating Core is taking place.

These proposals are presented as a further contribution to the ongoing work to develop the EOSC Federating Core proposition and definition. The proposals will continue to be developed, for example to address aspects such as procurement and business models, the role of HPC in the Federating Core, certification, user support and the Interoperability Guidelines.

EOSC-hub will continue to gather feedback with the aim of agreeing a community position paper on the Federating Core.

Appendix I. **Feedback Received to the Initial Proposals in the July Briefing Paper**

This Appendix summarises the input received in response to the Briefing Paper's initial proposals. The proposals for the Federating Core are updated in Chapter 3 in response to this input.

I.1 EOSC Requirements and Opportunities

Different visions were expressed in the feedback received, of the purpose of the EOSC, or the needs it should address. The clearest and most complete overall vision which emerged was of the EOSC as an opportunity to generalise the adoption of FAIR data policies for the benefit of a wider community of users, principally by widening access to data produced and stored by national and European research communities. The EOSC could provide a universe for all data, with equality of access to data, and be an enabler for common user needs for compute and storage resources. This would deliver economies of scale and avoid double funding, providing significant cost savings for funders. Nearly all the feedback received was compatible with this version of the EOSC although it didn't always articulate a complete vision. One set of feedback advocated a minimal EOSC although the functions it recommended for the Hub Portfolio and Compliance Framework were similar to those proposed in the Briefing Paper. The view was also expressed that a harmonised approach should be adopted to the EOSC and the EuroHPC initiative¹⁷, in the interests of delivering opportunities and efficiencies.

The EOSC should deliver equality of access to data and other resources for all researchers, regardless of their institution or geographical location.

More than just a data lake, the EOSC should develop to provide scientists with a personal space where they can create content (data analysis recipes, workflows, publications), store data and share their work with collaborators via a versioning system.

That which can be done in the community or Research Infrastructure, should be done there; the EOSC should add additional value to what is already available (subsidiarity principle).

Whilst for many research projects the EOSC may not be *essential* to achieving scientific goals, it could provide the framework to guarantee opening access to their outputs (datasets, software, publications) and results for a wider community.

I.2 Federating Core

There was general agreement to the proposed organisation of the Federating Core into the three elements of Hub Portfolio, Compliance Framework and Shared Resources, although the question was raised of whether the Shared Resources should be inside or outside of the Federating Core. It should be clarified that the Federating Core should be kept as lightweight as possible, to avoid imposing a substantial new burden of cost or effort on providers or users, but that it may evolve over time as technologies and practices change, as the range of resources provided through the EOSC extends, and also as the range of users expands. Also to be clarified is that the Portal is envisaged as a Universal

¹⁷ <https://eurohpc-ju.europa.eu/>

entry point to access resources through the EOSC, but it is not intended to be exclusive or to replace existing access routes to resources.

I.2.1 Shared Resources

The Shared Resources were the focus of the most comments about the Federating Core. The definition proposed in the Briefing Paper was acknowledged to be incomplete, but the Shared Resources also represent the highest direct potential value to users of the three elements of the Federating Core.

In terms of the composition of the Shared Resources, a query was raised about whether *services* should be in the Shared Resources, but the majority opinion was clearly that the Shared Resources should contain services. On the other hand, a clear view was expressed that *data* should continue to be owned, hosted and funded as it is now, so responsibility for data curation, preservation and quality should remain as it is now: the governance and regulation of data is complex, the archives and repositories are already there, and it is not consistent with an organisation's open data policy to place its data in the EOSC shared resources and charge for access to it. The Shared Resources could however include a local copy of data. The EOSC is envisaged as providing a thin layer on top of curated, preserved data, to facilitate discoverability, access and interoperability.

In terms of the position of the Shared Resources in the organisation of the EOSC, the Sustainability Working Group Strawman paper envisages the Shared Resources as belonging outside of the Federating Core¹⁸. Other feedback commented that the distinction between the Shared Resources and the Service Portfolio was insufficiently clear. Input from other sources including the EOSC cluster projects provided the view that the EOSC could add a great deal of value by enabling access to data, software and other resources such as storage and computing, and other generic capabilities and resources of broad interest that can be delivered through EOSC in a more coordinated and cost-effective manner than at present, avoiding duplication. Virtualisation, access and hosting services should therefore belong in the EOSC Shared Resources. The EOSC should provide dataset findability, data exploitation and workflow execution to researchers. AAI is key; adequate network bandwidth to the sites where data is hosted is also crucial. The Shared Resources could also include recommendations to institutions and other stakeholders, for policies and practices for FAIRness of repositories and other resources; tools for finding and selecting certified repositories; and resources to assist with developing competences, such as data stewardship training, FAIR competencies good practice training, or a FAIR competencies adoption handbook.

I.2.2 Hub Portfolio

There were differing views of the purpose of the Hub Portfolio. In one view, the Hub Portfolio should be an *interoperable integrating layer* facilitating the combined exploitation of the Shared Resources such as infrastructure-wide orchestration, scheduling, workflows, data discovery and management

¹⁸ The Strawman paper uses the term EOSC Minimum Viable Product (MVP) which it regards as the same as the Federating Core in the initial implementation of the EOSC, before extension of the EOSC to comprise wider public sector data and commercial users

and other functions facilitating seamless access. Another view was that it should be minimal and should focus on enabling *federation* rather than on integrating services.

It is widely agreed that PID services, specifically PID creation, assignment, resolution, description and search, should have a place in the Federating Core, but it was pointed out that the governance proposals in the Briefing Paper don't take account of the place of existing infrastructures/services with established governance mechanisms and possibly a global reach. Care should be taken to avoid duplicating existing services just to have them in the EOSC and subject to it.

The FREYA project is developing a PID Services Registry, a searchable catalogue of PID-related services, which could be accessible through the EOSC Portal. FREYA is also developing a PID Knowledge Hub which could potentially become part of the Federating Core in future.

There was a suggestion that collaboration software should include a code repository. Other suggested resources to include in the Hub Portfolio included competence centres with knowledge bases, and train-the-trainer resources - although these would need to be scoped in more detail - and FAIR semantic interoperability tools.

The functions planned for the EOSC Portal were regarded as being ambitious and provoked the question of why users should not simply continue to go direct to services, as now, focussing attention on the "value-add" of the EOSC and the urgent need not only to define it, but also for it to be distinct from the value which can be provided by services already available to the research community.

1.2.3 Compliance Framework

There was agreement that the Compliance Framework should be governed by the EOSC. Exclusivity should be avoided, and established practices and tools of communities should be acknowledged. Quality Assurance was seen as important.

The overall purpose of the Compliance Framework was accepted, but - as was recognised in the original Briefing Paper - its contents can be enhanced by additional components provided by other projects or organisations, such as policies/certification of data repositories and datasets, FAIR (and other) requirements for persistence and interoperability and minimum metadata set for catalogues as contributions to the Interoperability Framework; an Open Science Policy Monitor.

The necessity and value of the service management platform was queried. The SMS and also the accounting and CMDB functions in the Hub Portfolio were recognised as imposing an additional layer of administrative effort for service providers, providing a possible source of friction with established service management processes in operation within research communities. The business processes regulated by the EOSC should be clearly defined to apply to services and resources that are part of the Federating Core in order to meet high quality standards.

A query was raised about whether a single Acceptable Use Policy (AUP) for all resources accessible through EOSC was a realistic aim, with a recommendation to follow the principle of subsidiarity.

1.3 Access, Governance and Sustainability

It was considered important to look at the incentives for suppliers to make their resources available through the EOSC, and also to look at the costs involved for potential suppliers, particularly

commercial cloud providers, to integrate their current mechanisms for advertising and accessing their offerings.

There was agreement with the proposal to make the Hub Portfolio and Compliance Framework elements of the Federating Core accessible through the Wide Access mode to all suppliers participating in the EOSC federation, with the caveat that there should be no charge for access at the point of use.

Whilst there were no objections to the proposals for access modes applicable to the Shared Resources, there was, unsurprisingly, some restraint in agreeing to them due to the varying views expressed of the content and position of the Shared Resources, and also the need to understand funding and compensation models and to ensure suitable governance proposals are developed to cater for services (e.g. PID services) which should be included but which are already operational (and in some cases international in scope) and have their own established governance.

There was agreement to the proposal that services in the EOSC service portfolio are independently owned by their respective providers.

It was pointed out that where the EOSC intends to provide services (i.e. as part of the Shared Resources), significant operational requirements and costs will be involved, including supply market analysis and marketing, procurement and contracting. This aspect of the EOSC was not addressed in the original Briefing Paper and is also not substantively addressed in the present paper, but it needs to be borne in mind as part of the consideration of business models and procurement processes and requirements.

Appendix II. Cluster Projects' Vision For and Requirements of EOSC

The EOSC-hub Strategy Board includes representatives of the five EOSC cluster projects PaNOSC, EOSC-Life, ENVRI-FAIR, SSHOC and ESCAPE. Strategy Board members discussed the initial Federating Core proposals and provided input on how the EOSC could add value and what they could offer to the EOSC.

The cluster projects aim to prepare for EOSC several research communities which are significant data producers. Their vision for EOSC, their requirements of it and the contribution they envisage making towards it are significant indicators of the functions EOSC will need to deliver to support their needs, and of the resources which may be available to a wider community through the EOSC. This input is summarised below.

II.1 PaNOSC

PaNOSC is the INFRAEOSC-04 cluster project of ESFRI Photon and Neutron sources. It represents a plethora of scientific disciplines using photon and neutron sources to study materials on a wide range of scales - from angstroms to microns. A large user community of roughly 30,000 users annually profit from the Photon and Neutron sources in Europe (this includes the national and ESFRI sources).

Some of the PaNOSC Research Infrastructures have databases of data collected over the last 3 to 4 decades which are currently under-exploited due to lack of bandwidth to the repositories where they are stored, and due to lack of awareness of their existence. For example, palaeontology data in the repository at <https://paleo.esrf.fr> is an example of processed (curated) data which are not widely known or exploited yet. These data are ideal for cross-disciplinary applications and for linking up with data from museums and/or other scientific disciplines. The EOSC could improve the findability and accessibility, in particular, of data such as these to allow their potential to be tapped.

PaNOSC sees EOSC as an opportunity to generalise the adoption of FAIR data policies at all photon and neutron sources. Adopting FAIR data will enable data sharing across a wider community and the provisioning of services for remote data analysis. The table below summarises what PaNOSC requires from the EOSC, and what it will provide to the EOSC, to realise these objectives.

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

II.2 EOSC-Life

EOSC-Life is the Life Sciences INFRAEOSC-04 cluster project. EOSC-Life brings together the thirteen Biological and Medical ESFRI Research Infrastructures to create an open, digital and collaborative space for biological and medical research.

EOSC-Life sees the EOSC as an opportunity to realise the benefits of open data policies and make access to data and resources equal for researchers no matter where they are located, and also as a route to access to the cloud for Life Sciences communities such as tool developers. The table below summarises what EOSC-Life requires from the EOSC, and what it will provide to the EOSC, to realise these objectives.

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)

II.3 ENVRI-FAIR

ENVRI-FAIR is the connection of the Environmental Research Infrastructure community to the European Open Science Cloud, building FAIR services accessible for society, innovation and research. ENVRI-FAIR brings together all thirteen environmental domain Research Infrastructures from the ESFRI Roadmap but collaborates with the entire community of Environmental Research Infrastructures.

ENVRI-FAIR is working to create joint data service chains to allow the combined analysis of data and other resources from multiple sources. A major challenge is applying the FAIR principles to data and services. The EOSC has the potential to provide important support to the metadata and access preparation stages of ENVRI's work to make data and services FAIR, for example by providing templates for assigning provenance to data, and facilitating access to flexible, scalable research data and related services. The table below summarises what ENVRI-FAIR requires from the EOSC, and what it will provide to the EOSC, to realise its objectives.

What ENVRI-FAIR Needs From the EOSC

Generic ~~infrastructuredata and metadata~~ services such as for AAI, PID ~~and~~, provenance ~~and workflow management~~, 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 resources, to foster the FAIRification process, particularly for early stage RIs~~

Commented [7]: Entries in this table have been updated to make them consistent with edits made by ENVRI-FAIR to the equivalent table in the Community Position Paper v1.0 working draft document.

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 ENVRI

FAIR-based tools and resources for easy and seamless access to environmental data and services provided by the European ENVRI

ENVRI-hub – a ~~virtual~~ federated machine-to-machine interface to access environmental data and services provided by the contributing ENVRI

~~FAIR-based tools and resources for easy and seamless access to environmental data and services provided by ENVRI~~

II.4 SSHOC

The Social Sciences and Humanities Open Cloud (SSHOC) cluster project unites twenty partner organisations and a further 27 associates to create the social sciences and humanities area of the European Open Science Cloud. It aims to effect the transition from the current data landscape with its disciplinary silos and separate facilities to an integrated, cloud-based network of leveraged and interconnected data infrastructures supported by the tools and training to allow scholars and researchers to access, process, analyse, enrich and compare data across the boundaries of individual repositories or institutions.

SSHOC aims to maximise the efficiency and effectiveness of data re-use by applying Open Science practices and FAIR principles to data management. It views EOSC as an opportunity to achieve this by providing visibility (findability) for resources, together with guidelines governing interoperability and FAIRness, supported by helpdesk, workflows, workspaces, training and tools to assist researchers to make use of the resources made available to them. The table below summarises what SSHOC requires from the EOSC, and what it will provide to the EOSC, to realise these objectives.

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 storage, sharing, 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

II.5 ESCAPE

The European Science Cluster of Astronomy and Particle physics ESFRI research infrastructures, ESCAPE, aims to support the implementation of the EOSC through the expertise of the astronomy, astroparticle and particle physics communities. The cluster contributes two major areas of expertise in data stewardship: the astronomy Virtual Observatory infrastructure, and the expertise of the particle physics community in large-scale distributed computing and big-data management. A major challenge faced by ESCAPE in attempting to FAIRify data is how to make available the very large quantities of data produced by their community.

ESCAPE plans to implement Science Analysis Platforms for EOSC researchers to stage data collections, analyse them, access ESFRIs' software tools and bring their own custom workflows. It also plans to contribute to EOSC global resources federation by implementing a Data Lake concept to manage extremely large data volumes at the multi-Exabyte level. In addition, ESCAPE plans to expose scientific software through dedicated EOSC catalogues, and regards continuous shared software development and training as a significant opportunity for value-add for the EOSC. The table below summarises what SSHOC requires from the EOSC, and what it will provide to the EOSC, to realise these objectives.

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