A landscape overview of the EOSC Interoperability Framework Capabilities and Gaps

Authorship Community:

Diego Scardaci[1,3*] (0009-0007-7492-3616), Eva Sciacca[1,5] (0000-0002-5574-2787), Jean-Karim Hériché[2,4] (0000-0001-6867-9425), Mark Van De Sanden[2,6] (0000-0002-2718-8918), Klaas Wierenga[2,7](0000-0003-2297-0457), Paolo Manghi[2,8] (0000-0001-7291-3210), Damian Tamburri[2,9] (0000-0003-1230-8961), Jose Norberto Mazon[2,10] (0000-0001-7924-0880)

Editorial Board

Diego Scardaci[1,3] (0000-0002-2889-7695), Eva Sciacca[1,5] (0000-0002-5574-2787), Álvaro López García[2,11] (0000-0001-8267-3572), Wim Hugo[2,12] (0000-0002-0255-5101), Jerome Pansanel[2,13] (0000-0002-7067-5009), Lene Krøl Andersen[2,14] (0000-0002-4731-1338)

[1] Co-Chair, EOSC Task Force on Technical Interoperability of Data and Services
[2] Member, EOSC Task Force on Technical Interoperability of Data and Services
[3] EGI Foundation, Amsterdam
[4] European Molecular Biology Laboratory, Germany
[5] National Institute for Astrophysics (INAF), Italy
[6] SURF, Netherlands
[7] GEANT, Rijksuniversiteit Groningen, Netherlands
[8] Consiglio Nazionale delle Ricerche, Università degli Studi di Pisa, Italy
[9] Eindhoven University of Technology and the Jheronimus Academy of Data Science, Netherlands
[10] Universidad de Alicante, Spain
[11] Spanish National Research Council (CSIC), Spain
[12] Data Archiving and Networked Services - Royal Netherlands Academy of Arts and Sciences (KNAW), Netherlands
[13] University of Strasbourg, France
[14] DeiC, Denmark

* Corresponding author: diego.scardaci@egi.eu
** All TF members have had the opportunity to review and edit this document; invitations to be on the authorship list were open to all.

EOSC Association AISBL
Rue du Luxembourg 3, BE-1000 Brussels, Belgium
+32 2 537 73 18 | info@eosc.eu | www.eosc.eu
Reg. number: 0755 723 931 | VAT number: BE0755 723 931

Table of contents

Executive summary

This document is one of the deliverables of the EOSC-A Task Force on "Technical Interoperability of Data and Services". Its main aim is to collect the most up-to-date information regarding the EOSC Interoperability Framework, its main capabilities and implementation status.

The document summarises all the work that has been done on the EOSC Architecture and Interoperability Framework by different initiatives in the last years that allowed the convergence towards a common idea of what the EOSC Interoperability Framework should be. In line with the concept of *System of Systems*, the EOSC Interoperability Framework aims to put together, in a unique framework, the results and the decades of experience on interoperability of ESFRI/Thematic clusters, ERICs, general purpose e-Infrastructures (EGI, EUDAT, GEANT, OpenAIRE) and cross-domain initiatives like the Research Data Alliance, GO-FAIR and CODATA. The document presents the foundations of the EOSC Interoperability Framework, the EOSC Architecture and the Minimum Viable EOSC as they have been defined by the EOSC Working Groups, established during the previous EOSC Governance, and describes their evolution under EOSC Future, other relevant EOSC projects and via the different task forces of the EOSC Association.

Key technical design considerations to implement the EOSC IF, agreed between relevant EOSC stakeholders (research infrastructures, scientific clusters, e-infrastructures, etc.), are drawn. These include the bottom-up approach to build the EOSC IF and the concept of Interoperability Guidelines to promote the branding and adoption of standards and common best practices in EOSC as the basic elements of the EOSC IF. The EOSC Interoperability Guidelines describe the APIs, the metadata format of the exchanged data and all the processes required to really make EOSC resources interoperable. The need of a new EOSC Core component, the EOSC IF registry, to store and manage the library of interoperability guidelines has been identified together with how it should interact with the other EOSC Core components such as the EOSC Resource Catalogue. Technical considerations on how to achieve machine composability and data interoperability are also introduced.

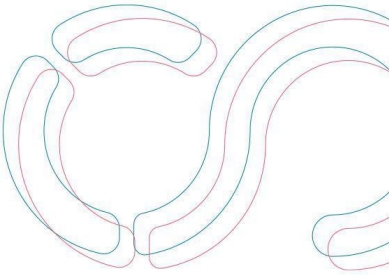
Furthermore, the deliverable depicts the ongoing EOSC Future work to implement the EOSC IF leveraging the design concepts previously discussed. A summary of the main achievements of EOSC Future in the areas of setting up an EOSC IF governance, developing the EOSC IF registry and populating the IF registry through the creation and registration of interoperability guidelines is described. In addition, Information on how this work is expected to be continued by future initiatives after the EOSC Future end is presented.

The document also reports an overview of the work on technical interoperability of other EOSC Projects and major scientific initiatives in Europe that are expected to feed the EOSC IF

in line with the bottom-up approach to populate the EOSC IF described earlier.

The results of an EOSC IF GAP analysis executed by the TF are reported. This exercise allowed the TF to identify areas that require further attention to fully realise the interoperability in EOSC. GAPs on both the EOSC IF Architecture components and the overall data and services interoperability have been identified and a list of recommendations has been provided.

Finally, the deliverable describes how the work on EOSC Interoperability relates and links to the same activity undertaken by other two major European initiatives such as the European Data Spaces and EuroHPC, and a list of recommendations is provided.



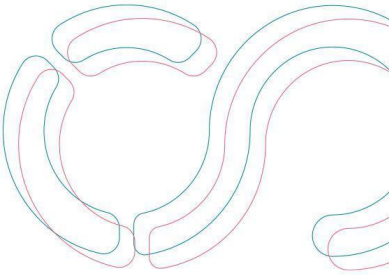
1. Introduction

The main goal of this deliverable is to present all the most up to date information regarding the EOSC Interoperability Framework (EOSC IF) with a focus on the Technical Interoperability of Data and Services. This document is elaborated by the EOSC Association Task Force on Technical Interoperability of Data and Services, within the Technical Challenges Advisory Group of the EOSC Association.

The intended audience for this document includes several EOSC stakeholders such as research communities willing to use EOSC, organisations and individuals involved in the implementation and expansion of EOSC, the Member States, the European Commission, etc.

This document is structured as follows:

- Section 2 provides background information on the EOSC Architecture and Interoperability framework summarising the work done by past initiatives (EOSC WGs, EOSC Projects, etc) and introducing key concepts and adopted terminology that are used in the rest of the deliverable.
- Section 3 introduces a series of key technical design concepts to implement the EOSC IF on which a consensus was found among several EOSC stakeholders such as the bottom-up approach to populate the framework and the concept of interoperability guideline. The chapter also includes a set of exemplar use cases that may benefit from the EOSC IF.
- The EOSC Future work to implement the EOSC IF is depicted in section 4.
- Section 5 presents an overview of the work on technical interoperability of other EOSC Projects and major scientific initiatives in Europe that are expected to feed the EOSC IF.
- The EOSC IF GAP Analysis accomplished by the TF is introduced in section 6 together with a set of recommendations.
- Work on interoperability from the European Data Spaces and EuroHPC and its relationship with EOSC interoperability are described in section 7.
- Section 8 summarises the main recommendations provided by this document.



2. The EOSC Architecture and Interoperability Framework - Current Status

As depicted in table 1.1 of the SRIA ("EOSC in its technological context"), the development of the Internet and the standards that shaped them, grew bottom-up, driven by the Internet Engineering Task Force (IETF). Exploiting the communication capabilities provided by the internet, development of the World Wide Web and the Semantic Web became possible, guided by the World Wide Web Consortium (W3C). The bottom-up approach also allowed for the development of community-based interoperability frameworks by ESFRI clusters, ERICs as well as through global and cross-domain initiatives like e-Infrastructures, the Research Data Alliance[1], AARC/AEGIS, GO-FAIR and CODATA. This bottom-up approach, while allowing for permissionless-innovation, and arguably being the main reason for the success of the Internet and its services, leads to many standards and issues with interoperability. The **EOSC Interoperability Framework** aims to put together, in a unique framework, the results of the aforementioned and other initiatives that are working on interoperability in the European Research Area, with the experience of general-purpose e-Infrastructures (EGI, EUDAT, GEANT, OpenAIRE) and of the thematic cluster communities to achieve interoperability in specific domains.

A lot of work has been done on the EOSC Architecture and Interoperability Framework by different initiatives in the last years. All these activities agreed on the basic principle that the EOSC Architecture and Interoperability Framework should be defined taking into account the decades of experience of Research Infrastructures, e-Infrastructures, and other user communities on operating services for researchers (data management, analytics, etc.) and wide distributed infrastructures as well as the established ways of working of the researchers. It should be particularly noted that the EOSC Architecture amends the Interoperability Framework, taking as a given that existing frameworks, services and ways of working will continue to exist, the architecture aims to bridge between those existing practices.

The foundations of the EOSC Interoperability Framework, the EOSC Architecture and the Minimum Viable EOSC have been defined by the EOSC Working Groups, established during the previous EOSC Governance. This work is advanced through work done in EOSC Future[2], and via the different task forces of the EOSC Association.

2.1. EOSC Interoperability Framework

The initial idea to establish the EOSC Interoperability Framework was formulated in the Iron

[1] The RDA is celebrating this year (i.e. 2023) it's 10-year anniversary
[2] EOSC Future Project web-site: https://eoscfuture.eu/

Lady report from the Sustainable WG on Solutions for a sustainable EOSC[3]. The concept of the EOSC Interoperability Framework has been further developed by the EOSC Interoperability Task Force of the EOSC Executive Board FAIR Working Group, with participation from the Architecture WG. The result of this work was published in a report[4] that identifies the general principles that should drive the creation of the EOSC Interoperability Framework (EOSC IF), and organises them into the four layers that are commonly considered in other interoperability frameworks (e.g., the European Interoperability Framework - EIF): technical, semantic, organisational and legal interoperability. The document finally contains a proposal for the management of FAIR Digital Objects in the context of EOSC and a reference architecture for the EOSC Interoperability Framework that is inspired by and extends the European Interoperability Reference Architecture (EIRA), identifying the main building blocks required.

In the same period the EOSC-hub project[5] dealt with the interoperability in EOSC in a more practical way, introducing the concept of **interoperability guidelines**: short documents promoting standards, well-known interfaces and community best practices within EOSC. EOSC-hub produced a first set of interoperability guidelines[6] with a focus on federation and common/general purpose services. Also at that time and before, the AARC projects and AEGIS brought together the research infrastructure and the e-Infrastructures to define an interoperable framework for authentication and authorisation: the AARC Blueprint Architecture and a set of accompanying guidelines.
Leveraging on the work of the EOSC Interoperability Task Force and of EOSC-hub, the EOSC Future project has been the first initiative that delivered an overall design[7] and a first implementation of the EOSC Interoperability Framework[8]. This is detailed in section 4.
After the end of EOSC Future, the EOSC Interoperability Framework will be delivered by the EOSC Platform operators that will be identified with the current procurement call[9]. The

[3] European Commission, Directorate-General for Research and Innovation, *Solutions for a sustainable EOSC : a FAIR Lady (olim Iron Lady) report from the EOSC Sustainability Working Group*, Publications Office, 2020, https://data.europa.eu/doi/10.2777/870770
[4] European Commission, Directorate-General for Research and Innovation, Corcho, O., Eriksson, M., Kurowski, K., et al., *EOSC interoperability framework : report from the EOSC Executive Board Working Groups FAIR and Architecture*, Publications Office, 2021, https://data.europa.eu/doi/10.2777/620649
[5] https://www.eosc-hub.eu/
[6] https://www.eosc-hub.eu/technical-documentation
[7] https://eoscfuture.eu/wp-content/uploads/2022/12/EOSC-Future-WP3-GEANT-D3.2a-EOSC-Architecture-and-Interoperability-Framework-2021-12-22.pdf
[8] https://eosc-portal.eu/eosc-interoperability-framework
[9] https://etendering.ted.europa.eu/cft/cft-display.html?cftId=12087

project that will be awarded in the call HORIZON-INFRA-2023-EOSC-01-05[10] will be in charge of setting up a stable governance of the framework while its innovation is expected to be delivered by the project that will be awarded in the call HORIZON-INFRA-2023-EOSC-01-04[11].

2.2. EOSC Architecture

In the report of the Sustainable WG on Solutions on a sustainable EOSC, the working group provided a high-level definition of the **Minimum Viable EOSC** (MVE). The MVE has been defined as including the EOSC Core and EOSC Exchange, that work with the FAIR datasets to be federated via EOSC.

On the basis of the work done in the Sustainability WG, the Architecture WG developed an architecture view on the MVE[12] defining the EOSC Core, EOSC Exchange and the EOSC Federation as the set of scientific services provided by RIs and Thematic Clusters to the respective communities.

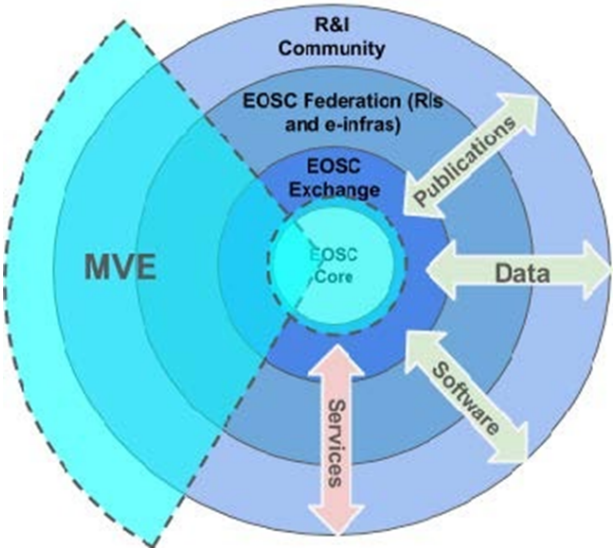


Figure 1: High-level diagram of the EOSC depicting the relationship between EOSC Core, EOSC Exchange, EOSC Federation and the MVE

[10] EOSC Architecture and Interoperability Framework: https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/horizon-infra-2023-eosc-01-05

[11] Next generation services for operational and sustainable EOSC Core Infrastructure: https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/horizon-infra-2023-eosc-01-04

[12] European Commission, Directorate-General for Research and Innovation, Sanden, M., Robertson, D., Appleton, O., et al., *EOSC architecture working group view on the minimum viable EOSC : Report from the EOSC Executive Board Working Group (WG) Architecture*, Publications Office, 2021, https://data.europa.eu/doi/10.2777/492370

The MVE is intended as a dynamic set of EOSC resources:

- The subset of EOSC resources necessary for forming the added-value and opportunities considered essential to be provided by the EOSC at a given moment in time, i.e., to allow essential services and research products (e.g., publications, datasets, software) to be discovered, composed, accessed and analysed via the EOSC, which could not be otherwise;
- The subset of EOSC Core components/services required to operate and deliver such resources.

The Architecture WG also developed an architecture diagram of the EOSC, providing a functional overview (see figure 2). It identifies the EOSC Users demand and Resource Providers supply side and the EOSC Core functions and capabilities. Furthermore, it introduces the EOSC Interoperability Framework with the interoperability guidelines, as a glue layer to support the integration and composability of resources across providers and for connecting resources to the EOSC Core functions.

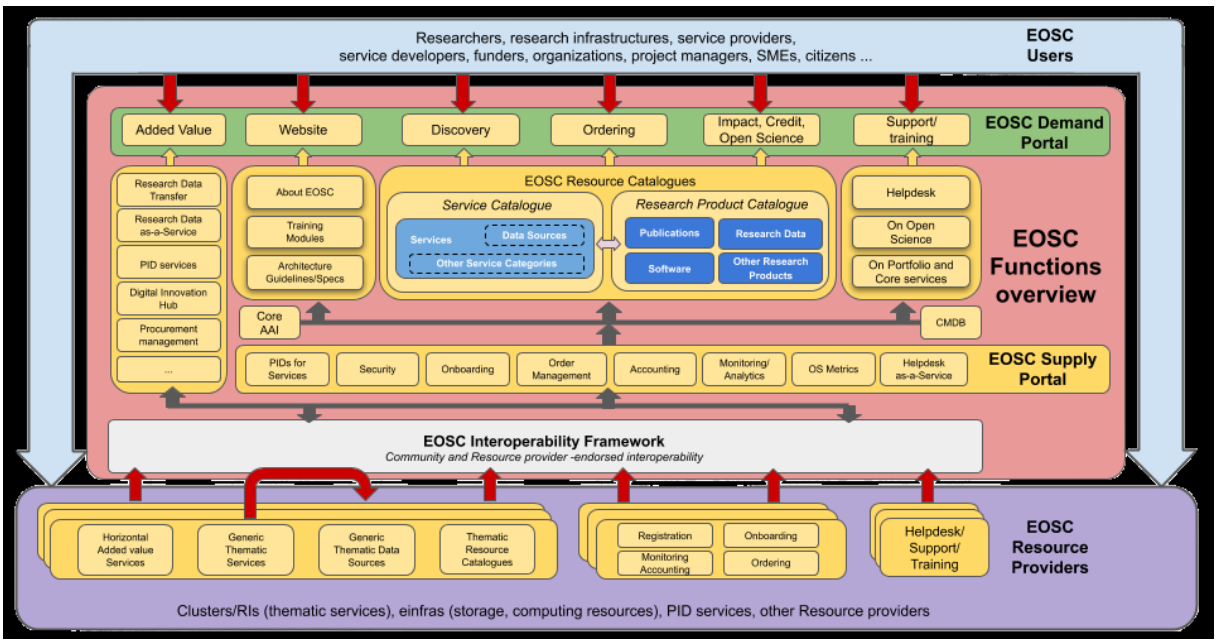


Figure 2: EOSC High-Level Architecture from EOSC Architecture Working Group

EOSC Future continued the work of the EOSC Architecture WG. The set of EOSC Core functions and capabilities defined by the Architecture WG were refined and extended considering the feedback from large communities such as the 5 Science Clusters (ENVRI-FAIR, ESCAPE, EOSC-Life, PaNOSC and SSHOC) as reported in D2.5a: Inventory of

Core Functions and Inclusion Criteria[13]. This work has been extended to include also EOSC Exchange capabilities in the D2.9: Co-designed Architecture Description[14] that has been used by the EC as the main reference source to develop the specifications of the EOSC Procurement. EOSC Future has also established a Service Portfolio Management process to evolve the EOSC Core components taking into account emerging requirements and new use cases from the user communities.

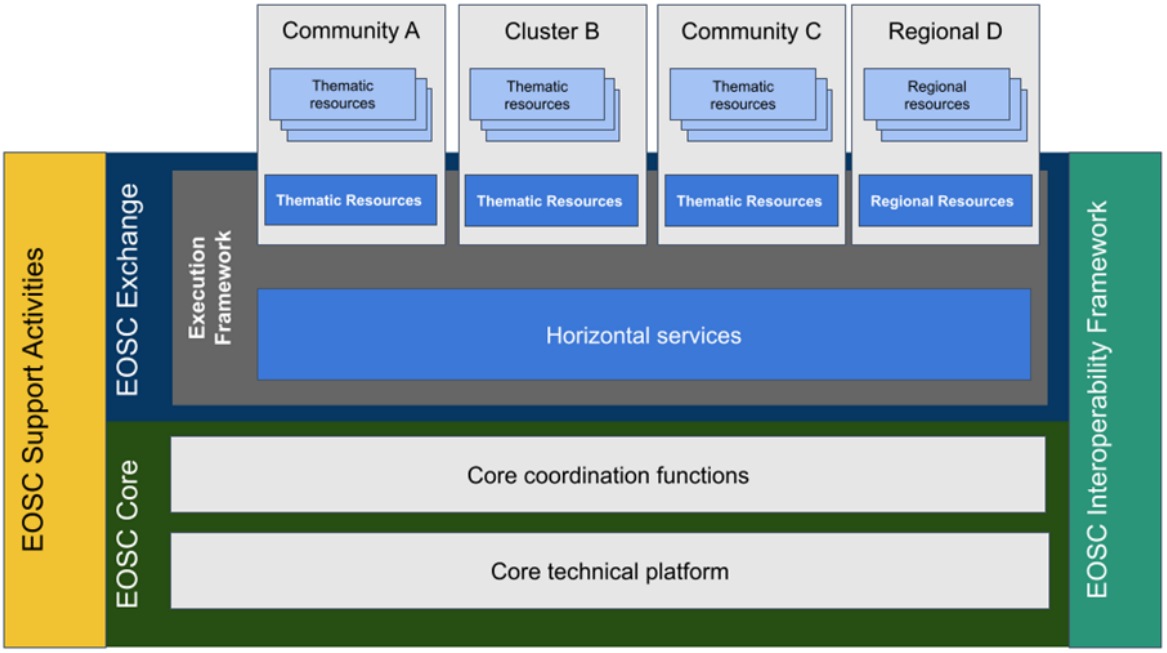


Figure 3: EOSC High-Level Architecture from EOSC Future

Furthermore, EOSC Future introduced the concept of **EOSC Platform** as an integrated operational environment that delivers the EOSC Core and a customisable portfolio of horizontal services. Exploiting EOSC Platform capabilities, research communities can add value to their own services in the Exchange and make them more useful and attractive to a wider range of science users. The EC has adopted the concept of EOSC Platform in the technical specifications of the EOSC Procurement to introduce the concept of EOSC Node, a deployment blueprint and operations guideline for future national, regional and/or institutional (even thematic) federated implementations of the EOSC Platform (see figure 3).

13 https://eoscfuture.eu/wp-content/uploads/2022/12/EOSC-Future-WP2-EGI-D2.5a-Inventory-of-Core-Functions-and-Inclusion-Criteria-2021-12-03.pdf

14 https://eoscfuture.eu/wp-content/uploads/2022/12/EOSC-Future-WP2-TGB-D2.9-Co-designed-Architecture-Description-2022-01-14.pdf

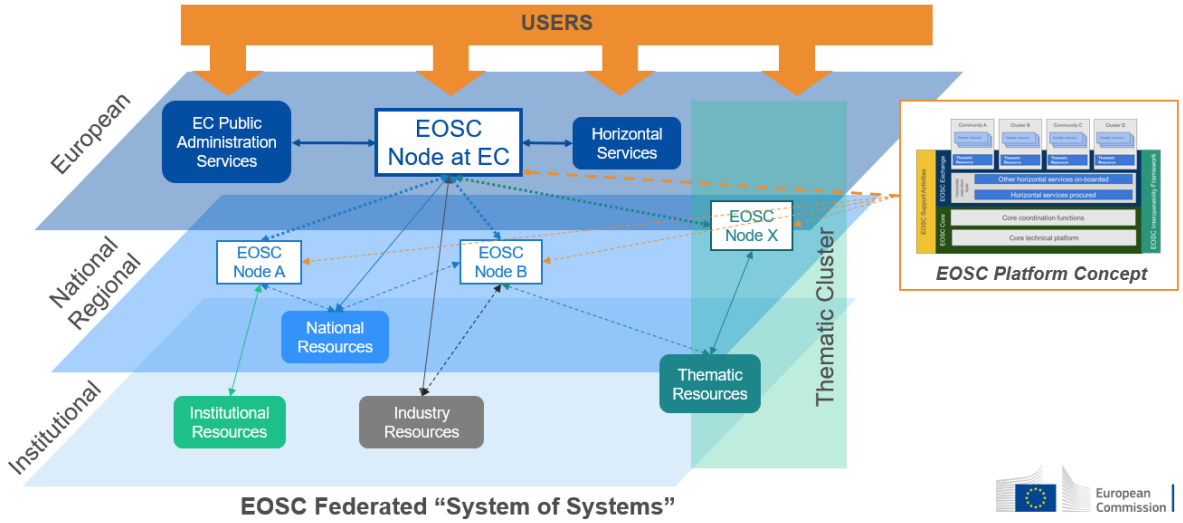


Figure 4: EOSC as a network of federated EOSC Nodes[15].

The architecture of the EOSC Platform[16], that is now operational, has been described using the C4 Model, a methodology to visualise software architecture, to achieve a better architecture definition and facilitate its maintenance and extension. The C4 Model[17] adopts an "abstraction-first" approach to diagramming software architecture, based upon abstractions in which the architecture starts at a high, conceptual level and subsequently drills down into more details. The EOSC Platform C4 diagrams show how the various components have been integrated in a unique platform and how they work together to enable key use cases such as registering resources (datasets, services, etc) in EOSC, discover and access resources, request support, move datasets, etc. detailing the high-level architecture defined at the start of the project.

[15] EC Call for Tenders - Managed Services for the European Open Science Cloud (EOSC) Platform. Descriptive Document: https://etendering.ted.europa.eu/cft/cft-documents.html?cftId=12087
[16] EOSC Future D3.3a Architecture and Interoperability Guidelines for Operational Services of the EOSC-Core: https://wiki.eoscfuture.eu/download/attachments/18940097/EOSCFU~1.PDF?version=1&modificationDate=1689754821263&api=v2
[17] https://c4model.com/

3. The EOSC Interoperability Framework (EOSC IF) - Technical design concepts

The experience gained with the different initiatives described in the previous section, which involved several different stakeholders, allowed the convergence towards a common idea of what the EOSC Interoperability Framework should be. This section summarises some EOSC IF technical design considerations that emerged from all these initiatives that are now agreed between many EOSC stakeholders (e.g. research infrastructures, scientific clusters, e-infrastructures, etc.).

3.1. Definition of EOSC Interoperability Framework

The EOSC Interoperability Framework (EOSC IF) provides a flexible framework of guidelines to support the interoperability and composability of resources (services, datasets and other research products) in the EOSC Core and EOSC Exchange. It will act as the glue to connect and orchestrate EOSC Resources (datasets, services and other research objects) shared by different communities and providers. The EOSC IF is defined as a Reference Architecture Framework which offers the freedom to providers to develop and operate provider specific implementations while conforming to the EOSC IF guidelines and standards.

3.2. Leveraging well-known standards and experience

The EOSC IF incorporates well-known standards, best practices and interfaces developed by standardisation bodies, research infrastructures, e-Infrastructures, other user communities, etc. in an homogeneous framework with the aim to merge in a unique environment decades of experience on dealing with interoperability issues. According to this, the EOSC IF should not be seen as a new source of interoperability standards and best practices, indeed it acts as a collector of existing interoperability solutions that can co-exist and interoperate in line with the vision of EOSC as a System of Systems. This bottom-up approach also takes into account that there is no one-size-fits-all approach that can be applied to integration, interoperability, and composability of resources across disciplines allowing the coexistence of interoperability solutions specialised for different thematic areas. Furthermore, the collection of all these interoperability solutions in a homogeneous framework creates the conditions to converge towards a reduced set of frameworks and standards selected through a natural process that will reward those adopted by more communities and that will satisfy the highest number of common requirements.

Relevant examples of interoperability standards and good practices that can be registered in the EOSC IF are listed below:

- AARC Blueprint Architecture and AAI guidelines from the AARC project.
- Interoperability guidelines to expose metadata of publications, datasets, and software across communities from OpenAIRE-Advance & EOSC Future.

- Interoperability guidelines for EOSC Core (AAI, accounting, helpdesk, monitoring, etc.) from EOSC-hub & EOSC Future.
- Cloud/HTC/HPC computing environments, data management systems, and analytics from EOSC-hub & EOSC Future.
- Provider, resource, and data source profiles to register resources into EOSC from EOSC Enhance & EOSC Future.
- FAIR data policies and practices from FAIRsFAIR.
- Community-specific efforts on metadata interoperability crosswalks to expose metadata of datasets such as BioSchemas.org developed by Elixir and OpenAIRE-Advance through the EOSC Enhance and EOSC Future projects.

3.3. Library of Interoperability Guidelines

The EOSC IF should embed a library of **Interoperability Guidelines (EOSC IGs)** to promote the branding and adoption of standards and common best practices in EOSC. The EOSC Interoperability Guidelines are the basic elements of the EOSC IF and they specify all the information needed to interconnect EOSC resources. They are not a simple collection of references to standards and best practices, indeed they describe the APIs, the metadata format of the exchanged data (the **EOSC Profiles**[18]) and all the processes required to really make two resources interoperable.

EOSC IGs should allow: (i) providers to understand how to design and configure services to interoperate with a wider community of users; (ii) software developers to understand detailed technical requirements needed to implement code to interoperate with such services.

An EOSC resource (datasets, services, etc.) can **declare to be compliant with one or more interoperability guidelines**, this means that instructions on how to interoperate with this resource are present in such IF guidelines.

The EOSC Interoperability Guidelines should be recorded in an **EOSC IF registry/repository**, where they will be maintained and curated, to make them discoverable by users or machines. A **common metadata structure** should be adopted to describe IGs, this will make information on interoperability best practices available in a homogeneous way facilitating the discovery, sharing and reuse.

Interoperability guidelines can be classified in two main classes:

- **Interoperability Guidelines for the EOSC Core services:** to describe how to integrate EOSC Exchange resources with EOSC Core services (AAI Federation, Helpdesk, Monitoring, Accounting, Order Management, etc.). Interoperation between EOSC

[18] EOSC Profiles: metadata schemas for consistently describing EOSC Resources (services, research products, data sources, IGs, etc), so that they are accurately described and easily found in the EOSC Catalogue and Marketplace: https://eosc-portal.eu/eosc-providers-hub/what-are-eosc-profiles.

Exchange and EOSC Core services guarantees a homogeneous user experience over EOSC (Single sign-on, common support channels and quality metrics, common metrics for resource consumption, etc.).

- **Interoperability Guidelines for the EOSC Exchange resources:** to embrace the whole research data lifecycle (discover and reuse, processing and analysis, curation and preservation, access and sharing) covering both service/infrastructure interoperability and data interoperability. The interoperability guidelines for the EOSC Exchange are the enablers of the **EOSC composability** making possible the creation of new tools and solutions combining various EOSC resources. EOSC Exchange Interoperability Guidelines are further classified into:
    - **Horizontal IGs:** interoperability guidelines for *Horizontal Resources* that deliver functions useful to multiple scientific domains. IG for Horizontal Resources can be associated with metadata, data storage and processing, scientific publishing and discovery, and a number of other functions relevant to many scientific processes.
    - **Community IGs:** interoperability guidelines specific for a given thematic area or community.

3.4. The EOSC Interoperability Framework - Interaction with EOSC Core Components

To enable the composability of datasets, services and other research objects in EOSC, the EOSC IF needs to interact with other core components of the EOSC Architecture.
The main client of the EOSC IF is the EOSC Resource Catalogue, a registry of all the resources onboarded into the EOSC. The Resource Catalogue gathers information about IGs from the EOSC IF Registry and annotates resource descriptions with the guidelines they comply with. This mechanism enables an interoperability-driven overlay of EOSC resources across different disciplines and the discovery of resources based on interoperability features. Indeed, thanks to this extra information, users can navigate over the EOSC Resource Catalogue to identify services and research products that can be easily combined or to gather information on the interfaces to access and exploit a service.

3.5. Enabling machine composability via the EOSC Interoperability Framework

While the information about the resource compliance with one or more IGs is a first base to foster interoperability in EOSC, it is not sufficient to enable machine composability. The IGs are human-readable instructions that EOSC Providers need to implement to enable given functionality/interoperability between their EOSC services and/or EOSC research products that cannot be interpreted by other services.

Machine composability requires that IGs are annotated with *configuration templates,* structured metadata profiles through which providers can describe the actual access parameters of their services for a given guideline. In such a way, a provider can declare the actual parameters to access a given interface of its service together with the compatibility with a certain IG.

Through this capability, a service provider can create an application composing EOSC Exchange services by dynamically discovering their interfaces via the EOSC Resource Catalogue. For example, it can set up a workflow to solve a scientific problem with resources delivered by multiple EOSC providers.

3.6. Data Interoperability in the EOSC Interoperability Framework

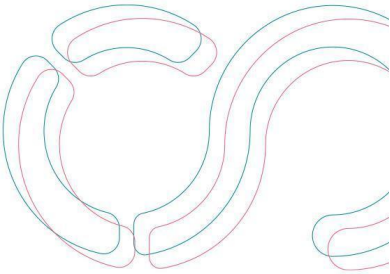
According to Intertrust Technologies Inc., data interoperability is defined as "the ability of systems and services that create, exchange and consume data to have clear, shared expectations for the contents, context, and meaning of that data"[19]. In the EOSC context, data interoperability therefore would allow data services and products to be stored, processed and forwarded for further use in different formats and locations to be fused together or otherwise (re-)combined. Such (dis-)aggregation and (re-)harmonisation is performed by so-called "data services", that is, compute entities which operate data transformations aimed at the aforementioned functionalities.
The EOSC IF will also include IGs for Data interoperability. These IGs should ensure that datasets can be exchanged between systems such that datasets can be understood, combined, processed and analysed by various EOSC services. This requires data to be machine-readable and available in suitable forms with enough contextual information to be interpretable and meaningfully composed with other data.
These requirements are addressed at four levels:

- **Data exchange protocols:** these are communication protocols that enable the sending and receiving of data between services. They are constructed on top of standard network protocols, most commonly in the form of application programming interfaces (APIs, including REST) over HTTP/HTTPs. However, files are also exchanged using FTP and the increase in data volume has led to the development and use of protocols over UDP (e.g. FASP, HTTP/3).
- **Data formats:** they define how information is structured and encoded in computer files. On the data consuming side, being able to parse a file format enables a service to make use of the data. On the data source side, the ability to emit data in widely supported formats promotes its uptake. The standardisation of formats for specific types of data varies by domain and community. For example, while the FITS image file format is a commonly used standard in astronomy, over 100 different image file formats, many of which are proprietary, are routinely used in the biomedical community.

[19] https://www.intertrust.com/blog/what-is-data-interoperability

- **Metadata standards:** metadata provides meaning and context to datasets and is essential to the interpretation and re-use of data as well as for data discovery. For example, metadata can record how a particular dataset was acquired or where a sample is coming from. To avoid ambiguities, metadata standards define common understandings of concepts and data elements and how this information should be represented. In some cases, metadata standards may also define the semantic relationships between the elements they represent, for example through the use of ontologies.
- **Governance and legal aspects:** these concern the responsibilities for data acquisition, processing and dissemination in compliance with applicable regulations as well as the allocation of suitable resources for making data available in EOSC. Adequate data management is a key component of interoperability by minimising information loss and creating an auditable trail ensuring data can be trusted. In addition, data interoperability in EOSC rests on the granting of permission to access and re-use available data. It is therefore essential that data is accompanied by clear licensing information and for the licences used to be as permissive as possible. However, regulatory and ethical requirements have to be met such that for example, privacy is protected or data is guaranteed to be complete. Harmonising and simplifying data-related legal frameworks across EOSC would promote interoperability by reducing frictions in data exchange.

Open specifications of data formats and protocols are the cornerstone of data interoperability by making data accessible without discrimination to all users and services and not tying data exchange and processing to specific hardware, software or groups. Open specifications offer guarantees against obsolescence as they can evolve with community needs and new implementations can be made available as technologies change. Open specifications can become de facto standards through wide adoption by a community and therefore don't need endorsement by a particular organisation. Of note, data interoperability can be greatly enhanced by data integration into data warehouses. However, in an environment of highly distributed data sources, interoperability of data from multiple sources requires data movement. To cope with increasing data volumes, EOSC services will need to support data exchange protocols that efficiently leverage the existing network infrastructure.
Although there are large variations between communities, technical elements required for data interoperability (such as data and metadata standards and exchange protocols) are generally available. Most often loss of interoperability arises from how data is managed during its lifecycle, for example through loss of metadata or locking of data in obsolete proprietary formats.



3.7. Different types of resources integration enabled by the EOSC IF

The EOSC Interoperability Framework with its IGs enables the **resource composability** that means the combined or integrated usage of two or more resources to provide researchers with added value or innovative solutions, facilitating and enabling research use cases and fostering collaboration between diverse research communities. Resources may be composed for different reasons; For example, a data analysis tool may be integrated with services from the EOSC Core (e.g. AAI, helpdesk, accounting, monitoring) to leverage well-established resources from EOSC. This would ensure a smooth integration of this analysis tool in EOSC, facilitating both its usage (e.g. integration with the EOSC AAI would allow users to access the tools with their usual credentials) and its operations, while integration with the EOSC Accounting would allow the providers to easily report back to the funders about the usage of the resource.

The same data analysis tool can also be combined with Horizontal Services from EOSC Exchange to handle recurrent patterns for scientific workflows such as publishing workflows, data transfer, data packaging, and container/VM deployment, etc. These Horizontal Services offer proven solutions to implement these functionalities in a distributed and diversified environment that can be simply imported by the scientific tool. As a result, the tool can be enriched with more features at low cost and can exploit the physical EOSC resources accessible via the horizontal resources without worrying about their heterogeneity.

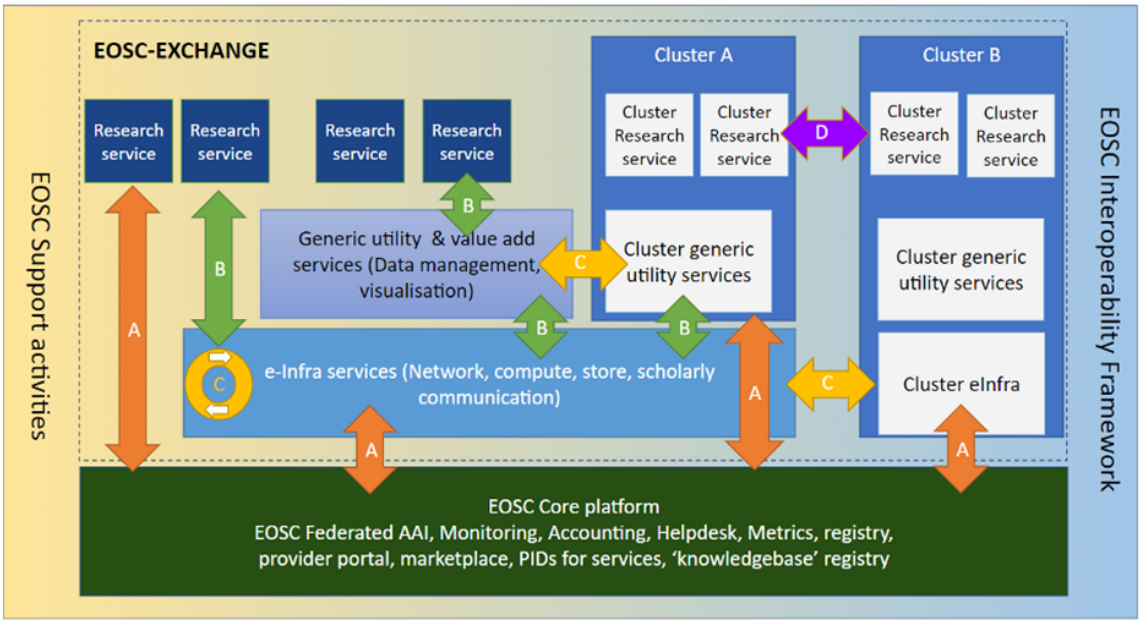


Figure 5: EOSC Composability and Integration typologies.

Another example is sharing data between different scientific communities. The same datasets can be used as input for analysis tools by both communities where there is a

common interest, such as for COVID-19, climate change, or other societal grand challenges, but they are all in different formats. A common data repository can be used to read and store the data in a common format. The user communities connect their tools to the shared data repository and datasets are converted on-the-fly from one format to another when needed.

It is evident that the nature of the above-described examples is not the same since the resource integration behind each of them is thought to achieve different objectives. However, resource composability covers all these cases.

The diagram in Figure 4 shows the different types of compositions and integrations that can be encountered in the EOSC landscape. The diagram shows the elements of EOSC Core and EOSC Exchange, connected and supported by the EOSC Interoperability Frameworks and Support activities. Vertical arrows represent **vertical integrations** (integrating a resource with more basic and/or common resources and functions) while horizontal arrows represent **horizontal integrations** (connecting peer resources) to add value**.** These two categories can be further divided into subcategories represented with the letters A, B, C and D:

- **Vertical integrations:**
  - A: Compose a resource with resources from the EOSC Core to make the resources interoperable in EOSC (e.g. protein visualisation service integrates with EOSC AAI). Integrations of type A add significant value for users and providers, as they make the user experience more coherent, and for providers save them effort on developing functionality themselves. This also occurs within the thematic clusters, as they create 'stacks' of resources.
  - B: Compose a resource with horizontal added value services to enrich the resource with additional features and easy/elastic/on-demand access of EOSC resources (e.g. a materials science service from a Science Cluster is integrated with a horizontal cloud computing service from an e-Infrastructure).
- **Horizontal integrations:**
  - C: Make interoperable horizontal resources from e-Infrastructures and clusters (e.g. a horizontal data management service from an e-Infrastructure is integrated with data management functions and data from a cluster, or integration between e-Infrastructure services from different organisations). Composability of this type already occurs to a great extent within the thematic clusters, which try to collect their resources into coherent platforms.
  - D: Compose scientific resources (such as data and services) to create added value solutions to handle complex scientific problems (e.g. an epidemiological simulation service from one Science Cluster is composed with a rich data set on logistics and international trade from another Science Cluster to help track the spread of a global pandemic). This is perhaps the most challenging type of composability and, like others, already happens within the Science Clusters

and communities. For instance, inside EOSC-Life there is a significant diversity of research and supporting resources, but lateral connections between resources and datasets are possible and rational as all are in the same broad research domain. By connecting them and breaking down artificial or technical barriers, research is further supported and accelerated. Taking this to the larger EOSC scale is even more challenging. Composing and integrating e.g. research resources from a cluster on photon and neutron science and a cluster on social sciences, is possible but takes significant effort and not all such possible composability would be required.

3.8. Use cases

This section illustrates, with few exemplar use cases, the interoperability of EOSC services and data leading to their composability and integration for added value in other EOSC services.

An example of a service integrated with multiple EOSC Core services (vertical integration - type A) is the ARGOS platform for Data Management Planning. In order to facilitate FAIR and open science practices, it leverages multiple AAIs (including EduGAIN and OpenAIRE's one), connects with OpenAIRE MONITOR for tracking its KPI activities, utilises various PID providers for identifying its linking datasets and Data Management Plans, connects to Digital Object Repositories APIs to depositing Data Management Plans, consumes Semantic Knowledge Graphs, repositories APIs and on-line vocabularies to facilitate filling of information of DMPs and delivers DMPs in a standard, machine-readable form (RDA's maDMP).

The second level of service interoperability is demonstrated by platforms delivering computing services through the integration of generic EOSC resources. An example of this is the BAND, a virtual desktop platform from EOSC-Life that runs in the de.NBI cloud, uses the Life Science AAI for user authentication and connects to various data stores including S3-compatible storage, public data repositories and EUDAT's B2DROP service. Another example is the Galaxy Pulsar Network which is a job execution system that allows the distribution of computational workflows from any Galaxy server instance to multiple European computing centres.

Further examples include the NEANIAS thematic services for atmospheric, underwater and space scientific communities. All of them have been onboarded in the EOSC Resource Catalogue, users can access using EGI Check-in or the NEANIAS AAI for authentication and authorization. They are managed through the NEANIAS Service Management System and are monitored thanks to the EOSC Exchange Monitoring WebUI.

Interoperability at the data level is largely realised in the life sciences across quite diverse research domains and resources by the widespread use of PIDs, ontologies and open

metadata standards whose development and adoption is supported by the ELIXIR interoperability platform. This enables for example the development of resources such as the COVID-19 data portal which integrates data from the life sciences and from the social sciences and humanities.

In the same way, interoperability of data, tools and services in the astronomical sciences is supported by the Virtual Observatory, the European Virtual Observatory (Euro-VO) and the standards of the International Virtual Observatory Alliance (IVOA) whose developments are very intensive around the world since 2002 to enable the international usage of astronomical archives as an integrated and interoperating virtual observatory.

As part of the ELIXIR Compute platform activities, reference implementations for GA4GH standards concerning federated analysis of sensitive genomics data with container-based workflows are being implemented, in connection with the Life Science AAI and GH4GA Passports standard to allow dataset-specific clearance and access, as well as to protect and promote confidentiality, integrity, and availability of data and to safeguard the privacy of patients.

Other examples from the biomedical domain are:

- GA4GH's genomic data toolkit (https://www.ga4gh.org/genomic-data-toolkit/) with e.g. the Beacon (https://beacon-project.io/) and refget (https://samtools.github.io/hts-specs/refget.html) APIs.
- Interoperable ontologies in the life sciences (https://obofoundry.org/).
- Various data standards for biological data types (e.g. PSI-MI for molecular interactions, PSI-MS for mass spectrometry).
- Development of a (cloud-compatible) standard for bioimage data (https://ngff.openmicroscopy.org/latest/, https://www.biorxiv.org/content/10.1101/2023.02.17.528834v2.full.pdf).
- Use of persistent identifiers (e.g. guidelines for the life sciences: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5490878/) (https://identifiers.org).
- AR registry of data standards, databases / resources and policies, including recommendations aligned with organisational policies is available at FAIRsharing.org: https://fairsharing.org/search?fairsharingRegistry=Standard.
- Frictionless data, ecosystem of data packaging standards and open-source software enabling data management, data integration and data flows. See the quick introduction.

The following table shows for the use cases described in this section what are the relevant composability/interoperability types according to the notation presented in section 3.7.

Use cases	EOSC Composability Types
ARGOS	A
BAND	A, B, C
Galaxy Pulsar Network	A, B, C
Nenias Thematic Service	A
COVID-19 data portal	A, B, D
European Virtual Observatory (Euro-VO)	D
ELIXIR Compute platform	A

4. The EOSC Future Implementation of the EOSC Interoperability Framework

The EOSC Future project is the main initiative currently in charge of implementing the EOSC IF[20]. Leveraging the past results described in Section 2 and the design concepts described in Section 3, EOSC Future is working on both creating the EOSC governance and tools to support the EOSC IF and on populating it with a first set of IGs.
This section describes the current status of the EOSC IF implementation activities in EOSC Future and the short-term plan until the project ends (September 2023). After the end of EOSC Future, the work on the EOSC IF is expected to continue in the EOSC Procurement Lot 1 (operations and proactive maintenance of the EOSC IF supporting tools) and in the projects awarded in the calls HORIZON-INFRA-2023-EOSC-01-04 (EOSC Core service innovation) and HORIZON-INFRA-2023-EOSC-01-05 (EOSC IF governance).

4.1. EOSC IF Governance

EOSC Future designed and implemented a first project-based governance[21] in charge of populating the EOSC IF validating and curating the Interoperability Guidelines in the framework.
The project proposed an EOSC IF governance model taking a pragmatic approach with the aim to build on the existing structure of the EOSC Future project. The key aspects that have been taken into consideration are:

- The need to define an independent group (similar to an editorial board) that can assess that requests for inclusion into the EOSC IF are compliant with a minimum set of requirements, namely maturity, community uptake, the existence of a group that maintains the item that has been proposed for inclusions and some governing model that allows for its evolution. This group needs to have a variety of expertise, as the EOSC Interoperability Framework consists of a wide range of topics. This body makes recommendations for inclusion or exclusion.
- The need to have an overarching body that has the oversight and the responsibility for the EOSC IF, is formally responsible for endorsing new and/or deprecating guidelines into the EOSC IF. This body would also offer an escalation point.

In addition, a third group was also proposed to ensure that the EOSC IF guidelines for the EOSC Core can be well-supported. These bodies, alongside their proposed responsibilities are listed in the table below.

[20] https://eosc-portal.eu/eosc-interoperability-framework
[21] EOSC IF Governance: https://eosc-portal.eu/eosc-interoperability-framework/eiab-and-eiac-charter

Body	Responsibility	Interim body for duration of EOSC Future project
EOSC Interoperability Advisory Board (EIAB)	Responsible for overseeing the EOSC IF; it endorses/deprecates guidelines, based on the recommendations of the EIAC.	EOSC Future Technical Coordination Board.
EOSC Interoperability Area Chairs (EIAC)	Responsible to perform the initial assessment of the proposed standards and guidelines and to make recommendations for inclusion/exclusion to the EIAB.	EOSC Future WP3 task leads - they will call in experts to help with the review process as needed. The review process will consider impact, maturity, global interoperability, update, and any cross-thematic nature of the guideline and its impact.
EOSC Interoperability Core Guidelines Owners	Responsible for contributing Interoperability Guidelines relating to the EOSC Core, and providing input to impact analysis of proposed EOSC Interoperability Guidelines.	Service owners for core components.

This provisional governance is now operational and has started to work on populating the EOSC IF. At the end of the EOSC Future, the provisional governance set by the project is expected to move to a more stable governance part of the activities of the EOSC Association. The project that will be awarded in the call HORIZON-INFRA-2023-EOSC-01-05 will be responsible for this transition.

4.2. EOSC Core Tools supporting the EOSC IF

EOSC Future delivered the EOSC Interoperability Registry, a new EOSC Core component hosting the catalogue of EOSC promoted Interoperability Guidelines, as a part of the EOSC Provider Portal[22]. This registry offers to EOSC Providers and other scientific initiatives capabilities to propose IGs for onboarding in the EOSC IF. It is connected to the EOSC Resource Catalogue of the EOSC Platform allowing service providers to declare the compatibility of their services to a certain guideline. The EOSC Interoperability Registry implements the EOSC IG metadata schema defined by the project and is currently in beta.
The EOSC Marketplace (the EOSC Platform GUI) has been extended to show compliance of a resource (datasets, services or other research products) with one or more IGs. The

[22] EOSC Provider Portal - a tool supporting providers on registering their resources into EOSC: https://providers.eosc-portal.eu/home

Marketplace search engine now supports the discoverability of IGs for consultation by EOSC providers or users[23].
By September 2023, EOSC Future will also release a prototype of an extended version of the EOSC Interoperability Registry supporting machine composability implementing the design idea described in the previous section based on the *configuration templates*. This work will be continued by the project that will be awarded in the call HORIZON-INFRA-2023-EOSC-01-04.

4.3. Populating the EOSC IF with Interoperability Guidelines

This section presents the work that EOSC Future started to develop IGs and populate the EOSC IF. The project focused its effort on creating IGs for EOSC Core services and Horizontal resources that are now being registered in the EOSC IF following the processes and procedures setup by the provisional EOSC IF governance. For this aim, the project further enhanced the concept of EOSC Interoperability Guidelines introduced by EOSC-hub defining the EOSC IG metadata schema[24], which has become part of the EOSC Profiles[25], to describe IGs. In a later stage, EOSC Future will open the EOSC IF for registration of EOSC Community IGs.
In the following, details on EOSC Core and EOSC Exchange (Horizontal and Community) IGs are presented together with an initial list of guidelines that are being onboarded in the EOSC IF. Guidelines that have been already onboarded are available in the EOSC Marketplace search engine for IGs[23].

4.3.1. EOSC Core Interoperability Guidelines

EOSC Core guidelines provides context and description in order to deliver technical instructions to Providers that would like to integrate their services and/or resources with (or be interoperable with) one or more EOSC Core Services
The Interoperability Guidelines of the EOSC Core services include two main types of information, the description of the interfaces/APIs that should be used by other services to interact with the EOSC Core services, and the profiles that specify the data and the related format that should be used to interact with the EOSC Core services via the APIs.
These Interoperability Guidelines do not define a single interface for interoperation but varying levels of integration, the **integration options**, from which a provider (seeking to onboard services or research products in EOSC) can select the option that best fits its needs. Integration options can specify various ways to integrate resources to an EOSC Core service, from the tightest option, which allows providers to benefit from a greater number of EOSC Core service features but requires more integration effort, to the loosest option, which enables a more limited feature set but with a lower cost in terms of integration. This approach leaves each provider free to benefit from the added value functions delivered by

[23] EOSC Marketplace search engine for IGs:
https://search.marketplace.eosc-portal.eu/search/guideline?q=*
[24] EOSC Interoperability Guideline Data Model:
https://wiki.eoscfuture.eu/display/PUBLIC/EOSC+Interoperability+Guideline+Profile+-+Data+Model
[25] EOSC Profiles: https://eosc-portal.eu/eosc-providers-hub/what-are-eosc-profiles

EOSC Core at the level it prefers without raising the cost of the basic integration with EOSC and, without creating a barrier that may hinder some providers to onboard their resources.
As an example, a Provider onboarding Services into EOSC can decide to be integrated with the EOSC central Helpdesk choosing one of the following options: (a) full integration - the Provider decides to use the EOSC central Helpdesk as its own helpdesk (as-a-service), (b) integration through the helpdesk API - the Provider programmatically connects its helpdesk to the EOSC central Helpdesk so that a ticket created in the EOSC central Helpdesk is automatically forwarded to its own helpdesk, (c) integration through e-mail - the Provider is notified by the EOSC central Helpdesk via e-mail when a ticket for its Services is created in the EOSC central Helpdesk. While the option (a) addresses the needs of a community willing to use EOSC Helpdesk technology to implement its community helpdesk, the options (b) and (c) are targeted for communities already operating their own helpdesks willing to integrate them with the EOSC one. Option (b) offers a transparent user experience with user tickets that can be indifferently submitted and managed in the EOSC or community helpdesks but requires a certain integration cost (the development needed to interface the community helpdesk with the EOSC one through the API). Option c) offers a less optimal user experience but does not require development.

EOSC Core IGs	Short description
EOSC Profiles	Metadata schemas for consistently describing EOSC Resources (services, research products, data sources, IGs, etc), so that they are accurately described and easily found in the EOSC Catalogue and Marketplace.
Resource Catalogue	Describes the main integration and usage use cases for the EOSC Resource Catalogue, regarding resource publishing and management, Resource Catalogue onboarding and synchronisation, resource graph maintenance.
Helpdesk	Describe the three integration paths offered by the EOSC Helpdesk to providers and communities onboarding services in EOSC: (1) Full integration: full synchronisation between EOSC Helpdesk and community helpdesk; (2) Ticket redirection: EOSC helpdesk used only as a contact point to redirect the initial request to the provider's or community mailing list without further integration; (3) Direct usage: the EOSC helpdesk can be used as the ticketing system for the community and their onboarded services.
Monitoring	Describe the five integration paths offered by the EOSC Monitoring: (1) Monitor an Onboarded Service: covers

EOSC Core IGs	Short description
	the scenario to monitor a service Onboarded to EOSC; (2) Monitor an Infrastructure (community): enables the monitoring of a complex infrastructure via the EOSC Monitoring; (3) Integrate External Monitoring service: allows an external source to publish monitoring data in the EOSC Monitoring; (4) Combine monitoring results of multiple infrastructures: covers the scenarios where the monitoring results of multiple infrastructures need to be combined in a number of reports; (5) Third-party services exploiting EOSC Monitoring data: a customer can access the results of the EOSC Monitoring Service and publish them in an external service/dashboard.
Research Product Accounting	Describe how a data source can use the EOSC Research product Accounting to collect usage activity from events related to its research products.
Service Accounting	Describe how a service provider can use the EOSC Service Accounting to publish accounting information of its services.
Order Management	Describe how service providers can integrate their ordering and provisioning processes with EOSC in a federated system of systems architecture and, consequently, facilitating the access to their services. Two integration paths (via APIs) are presented: (1) Resource Offering API: exposes ordering configuration and technical parameters of EOSC Resources' offerings, (2) Ordering API: exposes information about orders and manages them through the API. Orders can be managed through an ordering management system (OMS) connected to the EOSC Marketplace via the Ordering API or using the catch-all order management system SOMBO[26].
Messaging	Describe how a service provider can integrate their services to the EOSC Messaging system to exchange messages with other connected services.
AAI	AARC guidelines[27] to make interoperable AAIs belonging

[26] https://opsportal.eosc-portal.eu/
[27] https://aarc-project.eu/guidelines/

EOSC Core IGs	Short description
	to different research/e-infrastructures.
PIDs	Describe how providers can associate PIDs to their resources.

Table 1: EOSC Core IGs from EOSC Future

4.3.2. EOSC Exchange Interoperability Guidelines

The aim of the EOSC Exchange Interoperability Guidelines is to facilitate the integration of services, datasets and other research objects into new, even multidisciplinary, tools and solutions for the research.
Currently, the EOSC Exchange IGs are human-readable instructions that EOSC Providers need to implement to enable given functionality/interoperability between their EOSC services and/or EOSC research products. These guidelines may refer to multiple existing standards, protocols and guidelines etc. (e.g. FAIRsharing DOIs), in such a way that: (i) Providers can understand how to design and configure services to interoperate with a wider community of users; (ii) Software development can understand detailed technical requirements needed to implement code to interoperate with such services.
This section details the initial work that has been undertaken by EOSC Future to start the population of the EOSC IF with Exchange IGs benefitting from the experience on interoperability of the 5 Science Clusters (ENVRI-FAIR, EOSC Life, ESCAPE, PaNOSC, SSHOC) and 4 pan-European e-Infrastructures (EGI, EUDAT, GEANT, OpenAIRE) that are part of the project consortium. However, in line with the bottom-up approach described in section 3, all the main scientific initiatives in Europe working on interoperability aspects are expected to register Exchange IGs in the EOSC IF. Section 5 presents a non-exhaustive list of interoperability activities that can be onboarded in the EOSC IF as IG.

4.3.3. Horizontal Interoperability Guidelines

The EOSC Horizontal IGs are related to generic EOSC resources that can help to solve common patterns in research applications regardless of the scientific domain.
After a scouting of relevant work done by e-Infrastructure and Science Clusters, EOSC Future has identified an initial set of Horizontal IGs in key technical areas that are currently under development. These are listed in the following table.

EOSC Horizontal IGs	Short description
Data Transfer	Guidelines for data ingestion and transfer for processing in hybrid cloud environments
Data Publishing and Open Data	Guidelines for deposition and access into data repository

EOSC Horizontal IGs	Short description
Compute Continuum	Extension of the resource profiles to describe Cloud/HTC/HPC resources
Machine Learning	Guidelines for Machine Learning/Deep Learning data analytics services

Table 2: EOSC Core IGs from EOSC Future

4.3.4. Community/Thematic Interoperability Guidelines

Community or Thematic IGs are also expected to be registered in the EOSC IF to facilitate the sharing of best practices and standards between research communities. This is in line with the bottom-up approach described in section 3 and to prevent the *reinventing the wheel problem*, minimising the risk of dealing again with already solved issues. EOSC Future will soon open the EOSC IF registry to European research communities willing to share their standards and best practices to make them easily discoverable through the EOSC Platform. Examples of community IGs are the EOSC-Life roadmap to make tools and workflows from the biomedical sciences interoperable in EOSC[28] or the work on interoperability for data and services done by ESCAPE[29] [30].

[28] https://zenodo.org/record/7217294
[29] https://projectescape.eu/sites/default/files/WP2%20D2.1%20.pdf
[30] https://projectescape.eu/sites/default/files/D5.2%20Detailed%20project%20plan%20for%20WP5.pdf

5. Building the EOSC Interoperability Framework – A community-based approach

As stated in the previous sections, Interoperability Guidelines for the EOSC IF will come mainly from the existing experiences from RIs, Clusters, e-infrastructures and other relevant initiatives. The key objective of the EOSC IF is integrating in an homogeneous framework all the past, current and future work on interoperability in the European Research Area.

This section presents the work on technical interoperability of some of the main past, current and future EOSC initiatives (section 5.1) and of the major European scientific communities (section 5.2). The lists presented below are not exhaustive. They mainly present the initiatives where members of this task force have been directly involved.

5.1. Current and Future EOSC Initiatives on Interoperability

Initiative	Short description	Interoperability aspects
Main EOSC Implementation Project		
EOSC Future Apr 21 - Sep 23	The vision of EOSC Future is to demonstrate at the end of the project an operational EOSC Platform ('System of Systems') with an integrated execution environment consisting of data, professionally provided services, and open research products and infrastructure that will be accessed and used by the European researchers who will be engaged, facilitated, trained and supported to utilise the EOSC resources and solutions. EOSC Future vision is centred around three key tenets: • Realisation of EOSC Core and EOSC Exchange with interoperable data and resources. • Integration of data and resources from the Science Cluster communities into the	EOSC Future is making the first steps in establishing EOSC Interoperability Framework by implementing all the EOSC Core components to support the EOSC IF, establishing an initial governance to manage the onboarding of guidelines into the EOSC IF and starting the population of the EOSC IF with interoperability guidelines..

Initiative	Short description	Interoperability aspects
	EOSC Platform. ● Direct involvement of users in the co-design and implementation of the EOSC Platform.	
EOSC projects started in 2022		
FAIR-IMPACT Jun 22 - May 25	The overall objective of FAIR-IMPACT is to realise a FAIR EOSC, that is an EOSC of FAIR data and services, by supporting the implementation of FAIR-enabling practices across scientific communities and research outputs at a European, national, and international level.	The typical outputs from the project will be recommendations and best practices in respect of several interoperability topics, as well as pointers to well-functioning, FAIR-compliant implementations that may be useful across domains. A significant focus on aligning practices across European initiatives and infrastructures.
FAIRCORE4EOSC Jun 22 - May 25	The FAIRCORE4EOSC project focuses on the development and realisation of core components for the European Open Science Cloud (EOSC) supporting a FAIR EOSC and addressing gaps identified in the Strategic Research and Innovation Agenda (SRIA). Leveraging existing technologies and services, the project will develop nine new EOSC-Core components aimed to improve the discoverability and interoperability of an increased amount of research outputs. The 9 core components are: ● EOSC Research Discovery Graph; ● EOSC PID Graph; ● EOSC Metadata Schema and Crosswalk registry;	By establishing these new EOSC Core components, FAIRCORE4EOSC will develop guidelines on how to make use or connect to these components. These guidelines will be onboarded as EOSC Core guidelines in the EOSC Interoperability Framework.

Initiative	Short description	Interoperability aspects
	• EOSC PID Meta Resolver; • EOSC Compliance Assessment Toolkit; • EOSC Research Activity Identifier Service; • EOSC Research Software APIs and Connectors; • EOSC Software Heritage Mirror.	
EuroScienceGateway Sept 22 -Aug 25	EuroScienceGateway will deliver a robust, scalable, seamlessly integrated open infrastructure, based on the Galaxy framework, for data-driven research, contributing an innovative and customizable service for EOSC that enables operational open and FAIR data and data processing, empowering European researchers to embrace the new digital age of science.	The application of FAIR principles to workflows and adoption of FAIR Digital Objects to stimulate reusable and reproducible research and enable the EOSC Interoperability Framework. EuroScienceGateway will leverage a distributed computing network across 13 European countries, accessible via 6 national, user-friendly web portals, facilitating access to compute and storage infrastructures across Europe as well as to data, tools, workflows and services that can be customised to suit researchers' needs. At the heart of the proposal workflows will integrate with the EOSC Core. Adoption, development and implementation of technologies to interoperate across services, will allow researchers to produce high-quality FAIR data, available to all in EOSC. Communities across disciplines -- Life Sciences, Climate and Biodiversity, Astrophysics, Materials science -- will demonstrate the bridge from

Initiative	Short description	Interoperability aspects
		EOSC's technical services to scientific analysis.
EOSC4Cancer Sept 22 - Feb 25	EOSC4Cancer will make diverse types of cancer data accessible: genomics, imaging, medical, clinical, environmental and socio-economic. It will use and enhance federated and interoperable systems for securely identifying, sharing, processing and reusing FAIR data across borders and offer them via community-driven analysis environments.	EOSC4Cancer's well curated data sets will be essential input for reproducible and robust analytics and computational methods – including machine learning and artificial intelligence. EOSC4Cancer's five use cases will cover the patient journey from cancer prevention over diagnosis to treatment, laying the foundation of data trajectories and workflows for future European Cancer Mission projects.
BY-COVID Oct 21 - Sept 24	BY-COVID provides access to data relevant for infectious diseases from life sciences, public health and social sciences. RO-crate, a FAIR Digital Object implementation, is developed in this project with a specific focus on making workflow run metadata available in a standardised way. We also develop and collect workflows for analysis of data across the different domains, which are shared through WorkflowHub and annotated according to the defined best practices.	The design of the BY-COVID project rests on this foundation, aiming to connect well-established data resources and deliver access to heterogeneous, yet interlinked and organised data, across domains and jurisdictions via the COVID-19 Data Platform components. Over the course of the project, emerging national data portals will be linked to the COVID-19 Data Platform, thus establishing a federated digital space for infectious disease data.
EOSC Resource Provisioning (INFRAEOSC-07-2020)		
EGI-ACE Jan 21 - Jun 23	EGI-ACE is a 30-month project (Jan 2021 - June 2023) with a mission to empower researchers from all disciplines to collaborate in data- and compute-intensive research	The EGI-ACE project is addressing interoperability through actions distributed in several activities. These include the alignment of the services

Initiative	Short description	Interoperability aspects
	through free-at-point-of-use services. EGI-ACE delivers the 'EOSC Compute Platform', an integrated compute environment that federates compute and storage facilities with various platforms and access layers. The project also contributes to the EOSC Data Commons through the setup and provisioning of 'Data Spaces' that integrate scientific datasets and data analytics tools on top of the Compute Platform and deliver them as 'Thematic services' in EOSC. EGI-ACE services are made available for users via the EOSC Portal, and via the continuously open EGI-ACE Call for Use Cases.	operated by EGI-ACE with the policies and the technical and interoperability guidelines promoted by EOSC and the development of the EOSC Interoperability Guidelines to access compute and storage resources and for Data Spaces. Outcomes of the project will include several recommendations concerning the interoperability aspects.
DICE Jan 21 - June 23	DICE is enabling a European storage and data management infrastructure for EOSC, providing generic services to store, find, access, and process data in a consistent and persistent way while enriching the services to better support sensitive data and long-term preservation. 18 providers from 11 European countries are offering 14 state-of-the-art data management services together with more than 50 Petabytes of storage capacity. Over the duration of the DICE project, the resources are provided free at the point of use for its users.	DICE is working in four areas on the interoperability of services. In storage and compute DICE is working on the integration of EUDAT B2 services with S3 and HPC storages and the FENIX infrastructure. For PIDs DICE delivered the integrity check for PIDs and is working on the interoperability with the data type registry and for ePIC handles and DOIs using PID graph resources. For long term archives (LTAs) DICE produced a policy template that supports the interoperability of policies for data services, e.g. B2SHARE, and outsourcing the corresponding LTA, and on the implementation of signposting for B2SHARE to increase the FAIRness of the data published

Initiative	Short description	Interoperability aspects
		in B2SHARE. The fourth area is sensitive data. DICE is working on the interoperability of Secure B2SHARE, Laniakea (Galaxy Portal), and the Sensitive Data Archive (SDA).
OpenAIRE-Nexus Jan 21 - Jun 23	OpenAIRE is a European e-Infrastructure with the mission to establish, maintain and operate an open scholarly communication infrastructure. It provides services, resources and network to support a common European e-science environment. Maintaining a network of 37 National Open Access Desks (NOADs), OpenAIRE provides a forum for the alignment of policies and best practices on Open Access to publications, interoperability of repositories, Research Data Management and training to researchers, content providers, research managers and policy makers. Operating since 2009, OpenAIRE is an integral part and a leading force behind the European Open Science Cloud developments. OpenAIRE AMKE is a Non-Profit Partnership established in September 2018 with 47 members.	OpenAIRE in the context of Nexus project tackles with several interoperability challenges such as: depositing of Digital Objects in repository (Zenodo), issuing of PIDs (Zenodo), consumption of Data Management Plans (Argos), access to Semantic Knowledge Graph for 3rd party services (OpenAIRE Catalogue), exchange of services KPIs (MONITOR), Discovery of Datasets and other research products (OpenAIRE research graph, Zenodo) etc.
C-SCALE Jan 21 - Jun 23	C-SCALE federates existing European Earth Observation (EO) service providers, cloud resources and computing centres to empower the European EO research community to discover more easily, access, process, analyse and share Copernicus	The project delivered a blueprint, setting up an interaction model between service providers to facilitate interoperability between commercial (e.g. DIAS-es) and public cloud infrastructures to support Earth Observation services.

Initiative	Short description	Interoperability aspects
	data, tools, resources and services. The C-SCALE Data and Compute federation will ensure interoperability between distributed data catalogues, computational tooling and infrastructure. By making such a scalable Big Copernicus Data Analytics federated services available through the EOSC Portal, the project aims to support the EO sector in its development and to enable the integration of EO data into other existing and future domains within EOSC.	
Regional (INFRAEOSC-05-2018-2019)		
EOSC Nordic Sept 19 - Nov 22	EOSC-Nordic's overall objective is to foster and advance the take-up of EOSC at Nordic level by coordinating relevant initiatives taking place in Finland, Sweden, Norway, Denmark, Iceland, Estonia, Latvia and Lithuania; bringing them to bear in the context of the EOSC. The project aims to foster and coordinate all EOSC-relevant initiatives within the Nordic & Baltic countries and exploit synergies to achieve greater harmonisation at policy and service-provisioning level not only across the Nordic and Baltic region, but also with other countries, in compliance with EOSC agreed standards and practices.	EOSC-Nordic Service Interoperability Framework[31] (EOSC-Nordic IF) is based on the European Interoperability Framework (EIF) and the EOSC Interoperability Framework (EOSC IF). The recommendations from both frameworks were taken into account and analysed from the similarities and differences point of view. Recommendations which were considered to be applicable to services and that would enable more alignment in achieving interoperability of EOSC-related services, were selected. Six service providers, all possible EOSC services, were selected to analyse their

[31] https://eosc-nordic.eu/eosc-nordic-service-interoperability-framework/

Initiative	Short description	Interoperability aspects
		services from an interoperability perspective[32]. The compliance to the interoperability issues differs significantly from service to service. Nevertheless, the study indicates that the services achieve overall medium-level adherence to the interoperability principles. The service providers demonstrated a high degree of awareness, insight, and responsiveness of the interoperability issue.
EOSC Pillar Jul 19 - Dec 22	EOSC-Pillar gathers representatives of the fast-growing national initiatives for coordinating data infrastructures and services in Austria, Belgium, France, Germany, and Italy, to establish an agile and efficient federation model for open science services covering the full spectrum of European research communities.	At the technical level, EOSC-Pillar helped national actors (users and providers) to join EOSC by: - integrating different national services to deliver new functionalities; - supporting the integration of their services with the EOSC Core services; - supporting the onboarding of services into one of the EOSC recognized catalogues; - delivering a prototype of a national registry; - delivering federated AAI based on INDIGO-IAM; In particular the project also aimed to further develop the Laniakea cloud stack[33] with the possibility to import data from the Dataverse installation of EOSC Pillar partners.

[32] Lukkarinen, Ari, Vitlacil, Dejan, Kjeldgaard, Morten, & Riungu-Kalliosaari, Leah. (2020). "D3.3 Service Interoperability Framework". Zenodo. https://doi.org/10.5281/zenodo.4433556
[33] https://laniakea-elixir-it.github.io/

Initiative	Short description	Interoperability aspects
EOSC Synergy Sept 19 - Oct 22	EOSC-Synergy worked to expand the capacity and capabilities of the EOSC by leveraging the experience, effort and resources of national publicly funded digital infrastructures in a coherent way, therefore acting also as an incentive for national resource providers. EOSC-Synergy extended the EOSC coordination to participating countries by harmonising policies and federating relevant national research e-Infrastructures, scientific data and thematic services, bridging the gap between national initiatives and EOSC. The participating countries are: Spain, Portugal, UK, Czech Republic, Slovakia, Poland, the Netherlands, and Germany.	Within the project an analysis[34] has been conducted on the adaptations made in ten thematic services grouped into four categories: Earth Observation, Environment, Biomedicine, and Astrophysics. The analysis led to the identification of commonalities, best practices and common requirements, regardless of the thematic area of the service. EOSC-Synergy has a strong focus on software and service Quality in EOSC, with the development of the SQAaaS and FAIR evaluator.
NI4OS Sept 19 - Feb 23	NI4OS-Europe has been a core contributor to the EOSC service portfolio, committed to EOSC governance and ensuring inclusiveness at the European level for enabling global open science. The project supported the development and inclusion of the national Open Science initiatives in 15 Member States and Associated Countries in the EOSC governance. NI4OS-Europe instilled within the community the EOSC philosophy and FAIR principles. It also provided technical and policy support for the on-boarding of	NI4OS aimed to provide technical support to service providers from the region, ensuring interoperability between national and EOSC services, for their federation and integration into EOSC. Specific efforts were devoted to ensuring interoperability with EOSC-hub, OpenAIRE, and other core initiatives, as well as training. The initial test-cases of services for on-boarding in EOSC also include usage of services from the current EOSC offering, to test and resolve interoperability

[34] Calatrava, Amanda, et al. "A survey of the European Open Science Cloud services for expanding the capacity and capabilities of multidisciplinary scientific applications." arXiv preprint arXiv:2211.07738 (2022).

Initiative	Short description	Interoperability aspects
	service providers into EOSC.	issues.
ExPaNDS Sept 19 - Feb 23	ExPaNDS is the EOSC Photon and Neutron (PaN) Data Service. PaN Research Infrastructures are service providers with a high relevance for the success of EOSC. Through ExPaNDS for national PaN RIs and PaNOSC for PaN ESFRIs, coherent FAIR data services will be enabled to the scientific users of all European PaN facilities, universities and even industry.	The project delivered standardised, interoperable, and integrated data sources and data analysis services for Photon and Neutron facilities.
National initiatives		
NFDI (Germany)	The aim of the national research data infrastructure (NFDI) is to systematically manage scientific and research data, provide long-term data storage, backup and accessibility, and network the data both nationally and internationally. The NFDI will bring multiple stakeholders together in a coordinated network of consortia tasked with providing science-driven data services to research communities. NFDI is a German funding stream guaranteeing up to 90 million euro per annum for 10 years. The NFDI's programme aims for consortia include: ● Development of reliable and interoperable data management measures and	On the level of the interoperability NFDI's programme aims at: ● Establishment of data handling standards, procedures and guidelines in close collaboration with the community of interest. ● Development of cross-disciplinary metadata standards. ● Involvement in developing and establishing generic, cross-consortia services and standards in research data management together with other consortia.

Initiative	Short description	Interoperability aspects
	services tailored to the needs of the community of interest. ● Increased reusability of existing data, also beyond subject boundaries. ● Improved networking and collaboration with partners outside the German academic research system with expertise in research data management.	
HOSA	Within the Netherlands an architecture framework (HOSA) has been established for Higher Education and Research. The HOSA aims to define an architecture for sector facilities that are important for strategic collaborations between higher education institutions, sector partners and market parties. The basis of the HOSA is therefore an optimal articulation of the sector's demand regarding sector facilities. The HOSA also offers a facilitating framework for interoperability between institutions and providers of common ICT facilities. HOSA will make current and new initiatives in the field of sector facilities faster, more future-oriented, and more future-proof. Sector partners and market parties of ICT facilities can respond well to this with their service portfolio.	The HOSA has established domain specific architectures for: ● Flexible Education and Lifelong Learning ● Research Data Management ● Identity & Access Management
FRDN	The Flemish Research Data Network (FRDN) develops the preconditions necessary to motivate and enable researchers	Facilitating & driving discussions at Flemish level to increase interoperability across institutes to make research data

Initiative	Short description	Interoperability aspects
	from Flemish research performing organisations to exchange and reuse (FAIR) research (meta) data. This contributes to the integrity, quality, and efficiency of research and thus to accelerating innovation in society. The FRDN is a network of 36 Flemish research performing organisations who work together on Open and FAIR data. Important partners are the Flemish funding agencies and the regional nodes of European Research Infrastructures (ESFRI).	available. Building on existing formats such as CERIF to achieve semantic interoperability. Agreements across institutes on metadata for datasets generated in research supported by Flemish public funding.

5.2. Scientific Community Interoperability frameworks and standards

This section presents an inventory of scientific community interoperability frameworks and horizontal standards. These are the initiatives analysed in the context of the task force, however the list should not be considered exhaustive.

Initiative	Short description	Interoperability aspects
ESFRI Thematic clusters (INFRAEOSC-04-2018)		
ESCAPE Feb 19 - Jan 23	ESCAPE establishes a single collaborative cluster of next-generation ESFRI facilities in the area of astronomy- and accelerator-based particle physics in order to implement a functional link between the concerned ESFRI projects and EOSC.	WPs related to interoperability of data is WP4 CEVO (Connecting ESFRI projects to EOSC through VO framework) while interoperability of services is built from WP5 ESAP (ESFRI Science Analysis Platform) and from WP2 DIOS (Data Infrastructure for Open Science). Applicable deliverables: D4.4, D5.2 and D2.1.
PaNOSC Dec 18 - Nov 22	PaNOSC brings together strategic European research infrastructures in the area of Photon and Neutron science with the goal of contributing to the construction and development of the EOSC. The mission is to contribute to the realisation of a data common for Neutron and Photon science.	The PaNOSC data policy framework[35] includes the definition and adoption of common open standards for interoperability.
EOSC-Life Mar 19 - Aug 23	EOSC-Life brings together the Biological and Medical ESFRI to create an open, digital and collaborative space for biological and medical research. The project will publish FAIR data and a catalogue of services provided by participating RIs for the management, storage, and reuse of data in the EOSC.	WP2 Tools Collaboratory https://www.eosc-life.eu/tools-workflows/: • Implementation of RO-Crate • Containerisation of tools and workflows • WorkflowHub as registry for workflows, development of

[35] https://www.panosc.eu/data/panosc-data-policy-framework/

Initiative	Short description	Interoperability aspects
		BioSchema standards for workflows WP6 Provenance: • ISO Common Provenance Model • Fair Evaluation Services, Citation • EOSC-Life Report on data standards for observational and interventional studies, and interoperability between healthcare and research data[36].
SSHOC Jan 19 - Apr 22	SSHOC develops the social sciences and humanities area of EOSC, transforming the current data landscape with its disciplinary silos and separate facilities into an integrated, cloud-based network of interconnected data infrastructures.	The SSHOC project has worked with metadata and data format interoperability issues and built an interoperability hub consisting of a portal (called Conversion Hub) and selected metadata conversion solutions[37].
ENVRI-FAIR Jan 19 - Jun 23	ENVRI-FAIR connects the Environmental Research Infrastructure (ENVRI) community to the EOSC.	The overarching goal of ENVRI-FAIR is that all participating research infrastructures (RIs) will provide a set of interoperable FAIR data services that enhance the efficiency and productivity of researchers, support innovation, enable data-and knowledge-based decisions and connect the ENVRI cluster to the

[36] https://zenodo.org/record/5810612#.YtqHTXZBw2x
[37] SSHOC Deliverable - D3.6 Report on SSHOC format interoperability solution services, including new software: https://www.sshopencloud.eu/d36-report-sshoc-format-interoperability-solution-services-including-new-software

Initiative	Short description	Interoperability aspects
		EOSC[38].
Prototyping new innovative services (INFRAEOSC-02-2019)		
NEANIAS Nov 19 - Oct 22	NEANIAS delivered to EOSC thematic services in three scientific disciplines: underwater, atmospheric and space. Developed cross thematic services (core and delivery services) to enable cloud exploitation, service access and monitoring and common functionalities (e.g. data visualization or AI integration).	Apart from the thematic services, in WP6 "Core Services", WP7 "Delivery Services" and WP8 "EOSC Integration" NEANIAS developed vertical cross-thematic services to enable the deployment of the services to the Cloud and to onboard the services in EOSC making them interoperable within EOSC and across the different thematic communities. Applicable deliverables: D6.8, D8.6, D7.7
CS3MESH4EOSC Jan 20 - Jun 23	The project CS3MESH4EOSC develops ScienceMesh, an interoperable, pan-European mesh of data and higher-level services to allow friction-free collaboration between all European researchers. ScienceMesh federates existing EFSS storage nodes (Enterprise File Sync and Share) and facilitates access and sharing of data between them. ScienceMesh interconnects storage nodes with other services for Open Science: Data Science Environments, Digital Repositories and Open Data, Collaborative Editing and Large-Scale Data Transfers. Science Mesh transcends research disciplines and national and institutional boundaries.	The project builds on de-facto interoperability standards developed by the CS3 Community and beyond: OCM (REST API) for federated sharing between EFSS nodes supported by Owncloud, Nextcloud, Seafile and REVA middleware; CS3APIs for connecting data and applications, service and capability discovery; RO-CRATE as a metadata packaging format; WebDAV for data access. Interoperability with ScienceMesh is referenced as a requirement for Managed Services for the European Open Science Cloud Platform (CNECT/LUX/2022/CD/0023).

[38] ENVRI-FAIR - Interoperable Environmental FAIR Data and Services for Society, Innovation and Research: https://ieeexplore.ieee.org/document/9041704

Initiative	Short description	Interoperability aspects
Community Interoperability Frameworks and Standards analysed		
TDWG/ GBIF	Biodiversity Informatics has a 20+ years long history of defining data standards, vocabularies, exchange formats, identifiers and citation mechanisms. This vibrant community is global with numerous links with related European projects. See TDWG and GBIF for more details on DarwinCore and ABCD standards and related services.	The GBIF informatics architecture provides an open platform to connect and access biodiversity databases around the world. The distributed infrastructure spans across the hundreds of institutions participating in GBIF, enabling users to discover, access, integrate and help curate the growing content shared on the network. The GBIF architecture encompasses well-known community-developed data standards and protocols enabling interoperability at global scale. As an open infrastructure, a growing number of tools and workflows are able to connect and participate in the GBIF network.
GEO	The Group on Earth Observations has been in operation since 2005, and has published Data Management Principles and its implementation guidelines in 2015, with revisions and updates.	The guidelines address aspects of interoperability of both data and metadata, from the perspective of services (syntactic), schema (encodings), and semantics.

Initiative	Short description	Interoperability aspects
IVOA	The Virtual Observatory (VO) is the vision that astronomical datasets and other resources should work as a seamless whole. Many projects and data centres worldwide are working towards this goal. The International Virtual Observatory Alliance (IVOA) is an organisation that debates and agrees the technical standards that are needed to make the VO possible. It also acts as a focus for VO aspirations, a framework for discussing and sharing VO ideas and technology, and body for promoting and publicising the VO.	Astronomy has dedicated efforts on the creation of the Virtual Observatory, not only as a technical platform for sharing and exchanging data, but also as a set of specifications and standards for the definition of data sources that can be used by researchers, with a clear governance model.
LS AAI	LS AAI (LS Login) incorporates all the life science infrastructures operating in EOSC-Life into a common AAI system.	interoperability/incorporation into EOSC AAI.
FAIRsharing	A curated, informative and educational resource on data and metadata standards, inter-related to databases and data policies.	Guide consumers to discover, select and use these resources with confidence, and producers to make their resource more discoverable, more widely adopted and cited.
GA4GH	The Global Alliance for Genomics and Health (GA4GH) is an international, nonprofit alliance formed in 2013 to accelerate the potential of research and medicine to advance human health. Bringing together leading organisations working in healthcare, research, patient advocacy, life science, and information technology, the GA4GH community is working together to create frameworks and standards to enable the responsible, voluntary,	Creates standards for interoperability between human health and research organisations, both field-specific (e.g. genomic beacon API) and generic (e.g. GA4GH passports, Workflow Execution Service (WES) and Task Execution Service (TES)).

Initiative	Short description	Interoperability aspects
	and secure sharing of genomic and health-related data.	
Horizontal interoperability standards analysed		
RO-Crate	Interoperable lightweight research packaging format[39] [40] [41].	Implementation of the FAIR Digital Object approach.
Frictionless Data	Ecosystem of data packaging standards and open-source software enabling data management, data integration and data flows. See quick introduction	Frictionless Standards (aka Specifications) help to describe data. The core specification is called a Data Package, which is a simple container format used to describe and package a collection of data files. The format provides a contract for data interoperability that supports frictionless delivery, installation and management of data.
OASIS TOSCA	OASIS Topology and Orchestration Specification for Cloud Applications	TOSCA templates (blueprints) are platform-agnostic and

[39] Stian Soiland-Reyes, Peter Sefton, Mercè Crosas, Leyla Jael Castro, Frederik Coppens, José M. Fernández, Daniel Garijo, Björn Grüning, Marco La Rosa, Simone Leo, Eoghan Ó Carragáin, Marc Portier, Ana Trisovic, RO-Crate Community, Paul Groth, Carole Goble (2022): Packaging research artefacts with RO-Crate.Data Science 5(2)
https://doi.org/10.3233/DS-210053
[40] https://zenodo.org/record/4011999#.ZD1zVnbP02y
[41] Peter Sefton, Eoghan Ó Carragáin, Stian Soiland-Reyes, et al. (2020): RO-Crate Metadata Specification researchobject.org / Zenodo https://doi.org/10.5281/zenodo.3406497

Initiative	Short description	Interoperability aspects
	(TOSCA)[42] is a standard that uses the concept of service templates to describe the topology cloud workloads as an actionable document which is used by cloud orchestrators to instantiate, configure and adapt virtual infrastructures on top of cloud offerings.	provide reproducibility on the infrastructures. TOSCA is supported by many cloud orchestrators and platforms.

42 https://docs.oasis-open.org/tosca/TOSCA-Simple-Profile-YAML/v1.3/os/TOSCA-Simple-Profile-YAML-v1.3-os.pdf

6. EOSC Interoperability Framework GAPS

This section reports the results of an EOSC IF GAP Analysis accomplished by the TF. It has identified GAPs for both the EOSC IF Architecture components and the overall data and service interoperability. A list of recommendations to achieve full interoperability in EOSC is provided at the end.

6.1. EOSC IF Architecture Components

As described in section 2, the concepts of the EOSC Architecture and Interoperability Framework were initially defined by the EOSC Working Groups and further detailed in EOSC projects such as EOSC Future that defined the architecture of the EOSC Platform, a project based governance structure to manage the onboarding of guidelines in the EOSC IF registry and is populating the registry with a first set of IGs (see section 4).
After the initial work conducted in the EOSC Working Groups and through different EOSC related projects, the following gaps are still to be addressed:

Sustainable independent governance structure to maintain the EOSC Architecture and Interoperability Framework
On the basis of the initial work done within the EOSC Working Groups, the EOSC Interoperability Framework is to be further developed within the EOSC Future project. EOSC Future is a project which will finish in September 2023. While developing an initial governance structure within a project (e.g. EOSC Future[43]) is valid, it is not sustainable because of the temporary nature of a project. The aim of EOSC IF is not to build the biggest collection of standards out there, but rather the collection of standards to be promoted through EOSC. The choice of which standards to be promoted must be based on a rough consensus among stakeholders. Therefore, a sustainable, independent governance structure is required. In the Horizon Europe Work Program, a call has been included to create this sustainable and independent governance structure for the EOSC Architecture and Interoperability Framework[44].

Populating the EOSC IF registry with Interoperability Guidelines
At time of this writing, the EOSC IF contains a limited set of IGs. While the EOSC Future project released the first version of the EOSC IF Registry, it only becomes valuable to researchers and researcher communities when populated with a rich set of guidelines, including from existing interoperability frameworks from research- and e-infrastructures.

Addressing gaps
The aim of the EOSC IF is not to develop new interoperability standards and guidelines, but rather to promote existing guidelines and interoperability frameworks. Supporting Open

[43] https://eosc-portal.eu/eosc-interoperability-framework/eiab-and-eiac-charter
[44] https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/horizon-infra-2023-eosc-01-05

Science practices require technical and cultural changes. To support researchers and research communities in this evolving ecosystem, EOSC should promote the development of new interoperability guidelines and standards addressing gaps, for example supporting semantic and cross domain and cross infrastructure interoperability.

6.2. Services and data interoperability

This section offers (a) gaps identification and (b) challenges concerning all the above, stemming from discussions in the EOSC Technical Interoperability of Data and Services task force; other lessons learned from the EOSC IF open comments[45] procedures are included in the table below.

6.2.1. Gaps

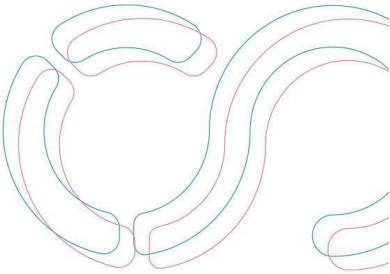
The table below offers details over the service and data interoperability gaps specifying:

- a name (Col. 1),
- their description (Col. 2),
- a risk-level, arranged according to a three-fold classification: (L)ow, (M)edium, (H)igh,
- an impact (Col. 3) in which the impact concern is reported as underlined and finally,
- a severity, also reported in a three-fold classification.

Name	Description	Risk Level	Impact	Severity
Domain focus	The interoperability of services and data was defined primarily by practitioners of domain-specific disciplines and might lack feedback from more disciplines (e.g., Engineering or Computer Science) with higher or more specific demands for data and services interoperability.	H	This gap might severely impact the applicability of the EOSC approach in a wider context. For example, specific concepts and needs might be under-represented or altogether absent.	M

45 https://www.eoscsecretariat.eu/eosc-liaison-platform/post/eosc-interoperability-framework-out-comment

Name	Description	Risk Level	Impact	Severity
Standardisation of the definitions	To address technical interoperability of services and data, it is required to define which are the high-level research objects considered (services, scientific workflows, software, hybrids objects, etc.). This standardisation of the definitions is so far latent or missing.	M	This gap might severely impact the applicability of the EOSC approach in a wider context. For example, specific concepts might suffer from concept drift and therefore might be defined in different ways.	M
Objects classification	A classification of research objects, as well as how data and services are composed to create more complex research objects is missing (e.g., considering interoperability to enable composability of scientific workflow steps).	H	This gap might severely impact the usability of the EOSC approach under-defining the entities upon which it provides specifications.	M
Resource federation	Large scale compute options exist which are under the effect of the EOSC considerations addressed in this document, yet Policy based federation of computing resources is under-defined and under-specified. For example, reserving slices of an SDN network, handling Cloud bursting per cloud compute zone, or more concretely defining an OSG federation with containerized SLURM (slurmd) options is neither defined nor addressed at the current stage.	H	This gap severely impacts the very goal of the EOSC specification and approach, namely the interoperability among services and data. It limits the ability for parties to act on their own orchestration exercises. Conversely, research communities may profit from EOSC by accessing multiple resources from multiple providers which is currently	M



Name	Description	Risk Level	Impact	Severity
			difficult to achieve.	
Security restriction	Authentication and authorisation often needs to be performed separately for each community/service.	M	This impacts the performance of EOSC-consistent services and data, which itself limits the usability.	L
Multilingual support	Multilingual data and services interoperability is required by multiple disciplines and specific research questions cannot be addressed via the current interoperability framework until the gap is addressed.	H	This impacts the language-agnosticity of the framework, which itself limits its larger scale usability.	M
Multi-level interoperability	Several disciplines reflected requirements to make data usable through both (i) categorised service with the focus on data and their usage, and (ii) searching services for discovering and accessing data at different levels of granularity and considering different formats.	M	This impacts the data usability at a larger scale. Discovering and processing of data at multiple levels would be required to promote the use and interoperability of the framework.	M
Data provenance	Usage-tracking and proper attribution for data and services remains an unresolved issue (data provenance to be considered as first-class citizens).	M	Some core-periphery service arrangements are needed to make the framework usable (e.g., format transformation services) and a framework to specify granularity levels across domains is	L

Name	Description	Risk Level	Impact	Severity
			required (beyond time and space dimensions).	
Data connectivity	Mechanisms to connect users' private data to cloud computational services remain unattainable.	L	Computational resources connected to data processing are required to ease their processing connection operations and further processing of data within specific boundaries, as specified by computational policies.	L

6.2.2. Challenges from EOSC initiatives

Stemming from the overview and discussions of interoperability initiatives described in Sec. 5, an overview of the main challenges to achieve interoperability dealt by some key EOSC initiatives and their relations with the gaps described in the previous table is presented.

EOSC related projects often focus on one or several specific domains. There are initiatives that are very topic-specific, such as EOSC4Cancer (that describes five use cases on cancer prevention) and BY-COVID (focused on COVID-19 data). Other initiatives, such as EOSC Synergy or FAIRCORE4EOSC, cover a wider range of domains, from environment to social sciences and humanities. EOSC Future is working on that, by defining a common standard for FAIR data, services and products across communities. Although these initiatives aim to transfer interoperability experiences and best practices to other domains, they generally lack explicit feedback from Computer Science and related disciplines to detect commonalities for a successful knowledge transfer, thus making such data and services interoperable from different domains. Also, as the FAIRCORE4EOSC project states, data may have different levels of granularity and may miss contextual metadata and data provenance, hampering the interdisciplinary re-use.

Apart from specific domains, there are initiatives tailored to specific territories. For example, the EuroScienceGateway project focuses on facilitating access to data and services across several European countries. FAIRCORE4EOSC highlights that research data come from national research information systems and should be connected beyond the institutional

repositories. As stated in the NI4OS initiative, technical support to service providers from the region must be achieved and interoperability must be ensured between national and EOSC services, for their federation and integration into EOSC. Therefore, cross-national interoperability of data and services are required to mitigate multilingual issues when accessing research data and services.

Standards must be adequately classified regarding the high-level research objects considered. For example, the RELIANCE project proposes a service ecosystem based on the 3 complementary and interconnected technologies: Research Objects, Data Cubes and AI-based Text Mining. To do so, GEO initiative proposes guidelines to address aspects of interoperability of both data and metadata in research objects. Also, as detected by FAIR-IMPACT, within the scope of the data and services interoperability, developers are limited to adhere to the standards due to their implementation complexity. According to the EOSC Nordic project, the challenge with existing interoperability frameworks is that they either focus on research and data outputs (i.e., EOSC Interoperability Framework) or generic IT services (i.e., European Interoperability Framework). However, an efficient research data lifecycle is supported by services, software, and infrastructure. Therefore, it is essential to focus on research outputs and the underlying enabling aspects and mechanisms, thus it is crucial to reduce the complexity of EOSC from the researchers' point of view.

A classification of research objects is also required. As a first step, CS3APIs within the CS3MESH4EOSC initiative enables the creation of easily accessible and integrated science environments, facilitating cross-institutional research activities.

EOSC Future aims to bridge the multilevel interoperability gap, since it aims to provide a powerful search engine for data and services. EOSC Future also considers the resource federation gap since it defines common standards for data and metadata to federate different catalogues of data and services. Lessons learned by FAIR-IMPACT shows us that resources tend to be siloed and alignment among catalogues is not achieved.

The comprehensive Service Interoperability analysis from EOSC Nordic suggested that the challenge with existing interoperability frameworks is that they either focus on research and data outputs or generic IT services. However, an efficient research data lifecycle is supported by services, software, and infrastructure. Therefore, it is essential to focus on research outputs and the underlying enabling aspects and mechanisms, thus it is crucial to reduce the complexity of EOSC from the researchers' point of view. The results of this analysis led to a series of recommendations for potential improvements and further work[46].

Finally, regarding data connectivity gap, the Science Mesh defined in CS3MESH4EOSC consists of an infrastructure that enables interoperability across different cloud service providers. Also related to this data connectivity gap, the EGI-ACE initiative highlights the

[46] EOSC-Nordic Service Interoperability Framework: https://eosc-nordic.eu/eosc-nordic-service-interoperability-framework/

importance of connecting EOSC with the Common European Data Spaces for enabling a Business2Researcher scenario.

7. EOSC Interoperability with other major existing systems or industry-related activities

7.1. European Data Spaces and SIMPL

Multiple pan-European Data Spaces are being designed and developed to implement the EU's "European Strategy for Data"[47], coordinated by the Data Space Support Centre (DSSC)[48]. The European Strategy for Data identified EOSC as a key sectoral data space for research data, expected to interoperate with other sectoral data spaces (e.g. in health, Green Deal and agriculture). EOSC initiatives (EOSC-A, EOSC related projects, etc.) have to coordinate with the DSSC, and through the DSSC with other Data Spaces, to ensure that such interoperation is possible through the technical, governance, access and use, and semantic interoperability mechanisms that EOSC will implement.
Many Data Spaces are expected to eventually operate on top of a new SIMPL Smart Middleware Platform for Cloud-to-Edge Federations and Data Spaces[49], that enables cloud-to-edge federations. While the consortium that will design and operate SIMPL has yet to be identified, the EC has committed that SIMPL will be developed through an open source community, following a published architectural approach which shares many functional requirements with EOSC. EOSC initiatives should be part of that open source community and work with the SIMPL project through a co-design approach to ensure compatibility and interoperability.

7.2. PRACE and EuroHPC

PRACE and EuroHPC are two international collaborations that aim to develop the European HPC-ecosystem in a complementary way. While PRACE is a science-driven infrastructure developed as a bottom-up activity, EuroHPC operates from top to bottom with high-level funding of pre-exascale supercomputers. These ecosystems are based on a federated, secure and hyper-connected infrastructure, including supercomputing, quantum computing, services and data. As such, they are playing a key role in several scientific domains and enable deeper scientific understanding and breakthroughs: they are important partners for the adoption and the outreach of EOSC.

Achieving interoperability between EuroHPC, PRACE and EOSC is essential in order to remove barriers between these projects and to facilitate the movement of codes and data across the infrastructures. World-class supercomputing systems, such as those provided by the EuroHPC initiative, have as a main objective to provide optimised tools to exploit these systems as efficiently as possible. In order to maintain the most efficient use of these systems, it may be necessary to develop gateways to ensure the interconnection with EOSC,

[47] https://digital-strategy.ec.europa.eu/en/policies/strategy-data
[48] Data Space Support Centre - project web-site: https://dssc.eu/
[49] https://digital-strategy.ec.europa.eu/en/news/simpl-cloud-edge-federations-and-data-spaces-made-simple

and services integrating the EOSC interoperability framework. Thus, to ensure excellent collaboration between all actors, it is necessary to develop the appropriate communication channel between PRACE, EuroHPC and EOSC.

Interoperability has already been studied by PRACE, and outcomes will probably be used by EuroHPC. The study is realised along the four interoperability layers, the same as the one used and described in the EOSC Interoperability Framework document[50]. Interoperability standards used by PRACE can be considered by EOSC to foster the collaboration between EOSC and PRACE, and hopefully with EuroHPC.

According to the analysis that have been done, the main issues to guarantee interoperability between EOSC and HPC centers are the following:

- AAI: HPC machines are usually accessed using local authentication/authorisation systems based on the username and password. There is no support for federated identities,
- Network: several network restrictions, varying site by site, due to strict security policies prevent easy access to HPC machines.
- Resource allocation: very often HPC resources are allocated to projects according to strict time scheduling and access on demand is not available.
- Software dependencies: special versions of software libraries are needed to process software modules to specific accelerated HW platforms (e.g. special CPUs or GPGPUs). Also, installing new libraries or software modules can require strict procedures (e.g. formal requests to the sys-administrators are needed).
- Data Access: stage-in/stage-out of data can be extremely complex due to the same restrictions mentioned above (e.g. AAI, Network, etc.).

For more information on some of the efforts that are underway in building bridges between HPC centres and the EOSC platform please see this Session[51] available on YouTube. It particularly zooms in on hot issues like resource allocation and persisting interoperability challenges.

[50] European Commission, Directorate-General for Research and Innovation, Corcho, O., Eriksson, M., Kurowski, K., et al., *EOSC interoperability framework : report from the EOSC Executive Board Working Groups FAIR and Architecture*, Publications Office, 2021, https://data.europa.eu/doi/10.2777/620649

[51] HPC centres as EOSC providers

8. Recommendations

This section summarises the main recommendations provided by this document.

Area	Recommendations
Establishment of the EOSC Interoperability Framework	The EOSC Interoperability Framework has to be established adopting a **bottom-up approach** aiming to put together, in a unique framework, the results of the most relevant initiatives that are working on interoperability in the ERA, such as general-purpose e-Infrastructures (EGI, EUDAT, GEANT, OpenAIRE), thematic clusters and other scientific communities, as well as through global and cross-domain initiatives like the Research Data Alliance, AARC/AEGIS, GO-FAIR and CODATA
Establishment of the EOSC Interoperability Framework	The EOSC IF should embed a library of **Interoperability Guidelines (EOSC IGs)** to promote the branding and adoption of standards and common best practices in EOSC. The EOSC Interoperability Guidelines should specify the APIs, the metadata format of the exchanged data (the **EOSC Profiles**[52]) and all the processes required to really make two resources interoperable. The EOSC IF should include IGs for EOSC Core and EOSC Exchange (horizontal and thematic) services.
Implementation of the EOSC Interoperability Framework	The EOSC Interoperability Guidelines should be recorded in a curated **EOSC IF registry/repository**. A common metadata structure should be adopted to describe IGs, this will make information on interoperability best practices available in a homogeneous way facilitating the discovery, sharing and reuse.
Implementation of the EOSC Interoperability Framework	The **EOSC Resource Catalogue** should gather information about IGs from the EOSC IF Registry and annotate resource descriptions with the guidelines they comply with. This mechanism enables an interoperability-driven overlay of EOSC resources across different disciplines and the discovery of resources based on interoperability features.

[52] EOSC Profiles: metadata schemas for consistently describing EOSC Resources (services, research products, data sources, IGs, etc), so that they are accurately described and easily found in the EOSC Catalogue and Marketplace: https://eosc-portal.eu/eosc-providers-hub/what-are-eosc-profiles.

Area	Recommendations
Enabling machine composability via the EOSC Interoperability Framework	**Machine composability** requires that IGs are annotated with configuration templates, structured metadata profiles through which providers can describe the actual access parameters of their services for a given guideline. In such a way, a provider can declare the actual parameters to access a given interface of its service together with the compatibility with a certain IG. For example an IG can describe how to make a data repository OAI-PMH compliant, then a data repository provider can declare to be compliant with this guideline and specify, as a configuration, the actual parameters to allow third-parties to access the repository.
Enabling data interoperability via the EOSC Interoperability Framework	The EOSC IF will also include **IGs for Data interoperability**. These IGs should ensure that datasets can be exchanged between systems such that datasets can be understood, combined, processed and analysed by various EOSC services. This requires data to be machine-readable and available in suitable forms with enough contextual information to be interpretable and meaningfully composed with other data.
Interoperability with Data Spaces	EOSC initiatives (EOSC-A, EOSC related projects, etc.) have to coordinate with the **Data Space Support Centre (DSSC)**, and through the DSSC with other Data Spaces, to ensure that such interoperation is possible through the technical, governance, access and use, and semantic interoperability mechanisms that EOSC will implement.
Interoperability with Data Spaces	EOSC initiatives should join the **SIMPL open source community** to work with the SIMPL project through a co-design approach to ensure that the development of SIMPL takes into account the experiences in interoperability of EOSC and to ensure the interoperability of the Data Spaces with EOSC and vice-versa.
Interoperability with EuroHPC	It may be necessary to develop gateways to ensure the **interconnection between world-class supercomputing systems** from EuroHPC and EOSC services to serve research communities leveraging both infrastructures.

9. Conclusions

In summary, the following key points are presented:

- Past work on EOSC Architecture and Interoperability Framework has been summarised introducing key concepts and terminology such as the Minimum Viable EOSC (MVE), the EOSC Core and the EOSC Exchange and a common definition of EOSC Interoperability Framework.
- A series of key technical design concepts, agreed among relevant EOSC stakeholders, to implement the EOSC IF have been introduced such as the bottom-up approach to build the EOSC IF leveraging the experience of thematic clusters, ESFRIs and e-infrastructures, and the concept of interoperability guideline as the basic elements of the EOSC IF.
- The status of the EOSC IF implementation from EOSC Future has been depicted including details about the definition of a provisional EOSC IF governance framework, the creation of the EOSC IF registry and its population with EOSC Core and EOSC Exchange (horizontal and thematic) interoperability guidelines. After the end of EOSC Future, EOSC IF implementation work will continue both in the context of the EOSC Procurement and of follow-up EOSC projects.
- Work on technical interoperability from EOSC projects and other relevant scientific initiatives has been introduced.
- The results of an EOSC IF Gap Analysis accomplished by the TF have been presented together with a list of recommendations to achieve fully interoperability in EOSC.
- The work on interoperability from two other major European initiatives, the European Data Spaces and EuroHPC, have been introduced together with their relationship with the activities to create the EOSC IF.
- A list of recommendations to establish the EOSC IF has been provided.

Acronyms and Abbreviations

Acronym	Name
AAI	Authentication and Authorisation Infrastructure
API	Application Programming Interface
DOI	Digital Object Identifier
EC	European Commission
EIAB	EOSC Interoperability Advisory Board
EIAC	EOSC Interoperability Area Chairs
EIF	European Interoperability Framework
EIRA	European Interoperability Reference Architecture
EOSC	European Open Science Cloud
EOSC IF	EOSC Interoperability Framework
EOSC IG	EOSC Interoperability Guideline
ERIC	European Research Infrastructure Consortium
ESFRI	European Strategy Forum on Research Infrastructures
EU	European Union
FAIR	Findability, Accessibility, Interoperability, and Reusability
GUI	Graphical User Interface
HTC	High Throughput Computing
HPC	High Performance Computing
IETF	Internet Engineering Task Force
ISO	International Standardisation Organisation
IT	Information Technology
IVOA	International Virtual Observatory Alliance
MVE	Minimum Viable EOSC
PID	Persistent Identifier
RDA	Research Data Alliance
REST	Representational State Transfer
SIMPL	Smart Middleware Platform
SRIA	Strategic Research and Innovation Agenda
SSO	Single Sign On
URL	Universal Resource Locator
W3C	World Wide Web Consortium

Acronym	Name
WG	Working Group