EOSC Future

D3.1

Science Cases for Development of EOSC Architecture and Frameworks

The EOSC Future project is co-funded by the European
Union Horizon Programme call INFRAEOSC-03-2020,
Grant Agreement number 101017536

Version 1.0
September 2021

D3.1 / Science Cases for Development of EOSC Architecture and Frameworks

Lead by **EMBL**
Authored by Licia Florio (GÉANT)
Reviewed by Rudolf Dimper (ESRF-PaNOSC), Diego Scardaci (EGI), Klaas Wierenga (GÉANT), Jonathan Tedds (Elixir/EMBL), Christos Kanellopoulos (GÉANT), Ian Bird (CNRS-LAPP/ESCAPE), Carsten Thiel (CESSDA), Keith Jeffery (ENVRI-FAIR) & Athanasia Spiliotopoulou (JNP)

Dissemination Level of the Document

Public

Abstract

This deliverable gathers existing science cases from which it derives an initial baseline of technical requirements for the work in the 'Architecture and Interoperability' work package (WP3). The deliverable takes as a starting point the requirements identified by the science projects defined in the EOSC Future project. These initial requirements will be used to drive the EOSC Architecture and Interoperability Framework, which is being developed by WP3 (first version expected by the end of 2021).

New requirements will be considered by WP3 as they emerge in EOSC Future as well as other EOSC projects and Research Infrastructures.

Version History

Version	Date	Authors/Contributors	Description
V0.5	14/09/2021	Licia Florio (GÉANT) Jonathan Tedds (Elixir/EMBL), Rudolf Dimper (ESRF-PaNOSC), Keith Jeffery (ENVRI-FAIR), Ian Bird (CNRS-LAPP/ESCAPE), Carsten Thiel (CESSDA)	Version approved by the clusters representatives
V0.6	15/09/2021	Licia Florio (GÉANT), TCB members	Version sent to the TCB for review
V0.7	17/092021	Licia Florio (GÉANT)	Final version submitted to the project office
V0.8	19/09/2021	Licia Florio (GÉANT), Athanasia Spiliotopoulou (JNP)	Incorporation of Comments
V0.9	21/09/201	Licia Florio (GÉANT)	Version circulated to consortium
V1.0	24/09/2021	Licia Florio (GÉANT), Ron Dekker (TGB), Mike Chatzopoulos (ATHENA)	Final Version submitted to EC

Copyright Notice

Table of Contents

Table of Tables

List of Abbreviations

Acronym	Definition
AAI	Authentication and Authorisation Infrastructure
AARC BPA	Authenticatiotn and Authoorisation for Researcch Collaboration BluePrint Architecture
CSV	Comma Seperated Values
ENVRI-FAIR	European Environmental Research Infrastructures building for FAIR services for research, innovation and society
FITS	Flexible Image Transport System
NetCDF	Network Common Data Form
PANOSC	Photon and Neutron Open Science Cloud
ESCAPE	European Science Cluster of Astronomy & Particle physics
ExPaNDS	European Open Science Cloud Photon and Neutron Data Service
RI	Research Infrastructure
SRIA	Strategic Research and Innovation Agenda
SSHOC	Social Sciences and Humanities Open Cloud Science Cluster
VOTable	XML standard for the interchange of data represented as a set of tables
VRE	Virtual Research Environment
XML	eXtensible Markup Language

1 Executive Summary

This deliverable focuses on the technical requirements identified by the EOSC Future science projects, large cross-domain scientific projects driven by the science clusters. The science projects offer a good insight on the capabilities that the EOSC architecture should offer.

These requirements and the content of this deliverable form the starting point of a living document that will be periodically updated[1] (roughly every six (6) months) to review the existing requirements, identify gaps and capture new requirements that may emerge in the EOSC Future project, in other EOSC projects and in the Research Infrastructures.

The content of this deliverable (and its future versions) is used to drive the work on the EOSC Architecture and interoperability Framework that is carried out in the Work Package 3 (WP3).

Achieving a good level of interoperability within EOSC is essential to implement the EOSC vision of enabling secure access to data and services across disciplines and countries. As agreed by all EOSC stakeholders, an interoperability framework is needed to implement such a vision.

Chapter 2 sketches the context in which the work in the Architecture work package takes place.

Chapter 3 lists the EOSC science clusters' science projects from the researchers' point of view, and extracts the requirements for the relevant clusters and the architecture and interoperability framework to support those use cases.

Chapter 4 provides a high-level analysis of the science cases and identifies common requirements. These will inform the EOSC interoperability framework, the first iteraction to be released in November 2021 by the WP3 team.

Chapter 5 provides the conclusions and the next steps.

[1] The updates will be done on the WP3 EOSC Future wiki, https://wiki.eoscfuture.eu/pages/viewpage.action?pageId=1805501

2 Introduction

The Architechture and Interoperability Work Package (WP3) defines the architecture and interoperability guidelines and frameworks needed for the EOSC Core (which provides the minimum functionality that is required to enable the federation of data and services across domains and countries) and EOSC capabilities for integrating Research Infrastructures (RIs), providers (both research and commercial) and researchers. Achieving a good level of interoperability within EOSC is essential to implement the EOSC vision of enabling secure access to data and services across disciplines, countries and sectors. As agreed by all EOSC stakeholders, an interoperability framework is needed to implement such a vision.

Since the inception of EOSC there has been significant work carried out by the EOSC Science Clusters and by different EOSC projects, some of them preceding EOSC Future, others recently started. In addition the same communities have provided input and feedback to other relevant EOSC documents such as the SRIA [15]; valuable information is also presented in the reports produced by the EOSC Executive Board Working Groups[2] [6].

The approach taken for this deliverable and for WP3 in general is to collect existing knowledge, requirements and artifacts available at the start of the EOSC Future project,understand the current landscape, assess existing results and further develop them as needed. As existing science cases had emerged in previous projects, and as science projects were defined in the EOSC Future Description of Work, it was agreed that the starting point for this deliverable would be to collect these science cases, validate them with the communities in which they were generated and assess the requirements. This approach is meant to provide an accessible summary of the current developments and a basis to identify which requirements were not yet addressed in order to prioritise the work in WP3 concerning the Interoperability Framework.

This deliverable focuses on the science projects, large, cross-domain and mature projects driven by the Science Clusters: namely **ENVRI-FAIR** [4], **EOSC-LIFE** [8], **ESCAPE** [10], **PaNOSC** [13], **SSHOC** [16]. They are important stakeholders for the EOSC as a whole and encompass different disciplines and RIs. The Science Clusters are EU projects launched in 2019 to link ESFRI and other world-class Research Infrastructures (RIs) to EOSC. They bring together seventy two (72) world-class RIs from the ESFRI roadmap and beyond to work on FAIR data management and connecting their user communities to the EOSC.

This deliverable should be considered as the starting point of a living document that will be updated regularly (roughly every six (6) months to align with the project delivery cycles as defined in the EOSC Future Technical Annex), to review the identified requirements and to capture new ones allowing to further update the EOSC architecture and Interoperability Framework. To that end, WP3 with the support of WP10, will organise town hall meetings to present the work done and solicit feedback on that; and to (re)assess the requirements for the Science Clusters, other Research and Infrastructures and other EOSC projects.

Links to the science cases detailed in this document as well as to additional relevant information will be maintained in the dedicated section on the WP3 EOSC Future wiki [7].

[2] The Working Groups ended their mandate in December 2020.

3 Overview of existing science projects and their requirements in the EOSC Science Clusters

This section describes the science projects of the five (5) Science Clusters. Each of the Science Clusters connects a number of thematic Research Infrastructures and acts as the interface between their scientific communities, their Research Infrastructures and the EOSC.

It is important to note that all these clusters have matured significantly in recent months. The initial science projects mostly highlighted the needs for resources and capabilities within each cluster, whereas the final retained science projects also include cross-cluster collaboration and therefore will help address key interoperability concerns.

The next chapters summarise the requirements of the science projects from each Science Cluster providing also a short description how these requirements impact the EOSC architecture.

3.1 PaNOSC/ExPaNDS

PaNOSC [13]/ ExPaNDS[11] is a collaboration between almost all Photon and Neutron sources in Europe; their work is funded via two dedicated projects. Their goal is to make FAIR data a reality in the partner RIs and connect the RIs to the EOSC. To do this, PaNOSC has developed a FAIR data policy framework together with the national PaN RIs.

The functional needs for the two PaNOSC/ExPaNDS science projects and EOSC are described in the table below.

Table 3-1 PaNOSC/ExPaNDS science project requirements

Science case(s) description from the users perspective	Requirements for PaNOSC/ExPaNDS	Requirements for EOSC Architecture
Researchers using analytical facilities to examine their samples are often using several facilities to probe their samples with different tools and methods. This is why: • a researcher from PaNOSC/ExPaNDS needs to seamlessly use compute and/or storage resources provided by the e-Infrastructures or other providers to analyse data from the PaNOSC Research Infrastructures using the PaNOSC (UmbrellaID) identity and without having to re-register across infrastructures. • users need to freely access, share, transfer and analyse large datasets of different types and sources. They also need to reuse and combine data for different research questions, generating new services that meet community standards.	PaNOSC/ExPaNDS needs to offer: • remote and authorised access to analytical facilities. • downloadable metadata & raw data. • software to browse, (pre)-visualise and analyse raw data. • long term open data archives • integrate services into EOSC for other science projects to use them and collect feedback.	EOSC needs to: • operate a reliable Federated AAI Interoperability Layer that allows for seamless integration of the PaNOSC/ExPaNDS AAI and supports the access to services and resources across Research Infrastructures. • offer a federated search capability for scientific data across a wide variety of domains. • offer a data transfer service for large data sets. • offer data storage and compute facilities.

3.2 EOSC-Life

The life science cluster, EOSC-Life [8] brings together 13 biological and medical Research Infrastructures (RIs) to create an open collaborative space for digital biology. It aims to publish FAIR life science data resources for cloud use creating an ecosystem of innovative tools in EOSC and through the Life Science Login (common AAI), enabling groundbreaking data-driven research in Europe by connecting life scientists to EOSC including for sensitive data.

The functional needs for the EOSC-Life science projects and EOSC are described in the table below. Additional information can be found in the EOSC-Life Report on requirements for regulatory compliance of sensitive health data and biological and medical research data [17].

Table 3-2 EOSC-Life science project requirements

Science case(s) description from the users perspective	Requirements for EOSC-Life	Science case(s) description from the users perspective
• Researchers working on different European or national research projects need to discover publications, find statistical data and data objects generated by different studies and access them in a privacy preserving manner. They need to have access to compute resources to analyse the data. When authentication to EOSC-Life is needed, it should be a user-friendly process. • In addition, researchers need a set of tools to increase the FAIRness, that is to ensure that more data generated in life-science research is Findable, Accessible, Interoperable and Reusable	EOSC-Life needs to enable European scientists to access advanced data resources and services in regulatory compliance with ethical and legal requirements. This requires from EOSC-Life: • user management and access services through a Life Science Login. • tools to increase FAIRness of data. • framework for open science policies. • open collaborative model to bring existing national cloud infrastructure together.	EOSC needs to: • provide core components (i.e. Federated AAI interoperability, cloud/ storage resources) that can be readily incorporated into the many national, regional and locally funded services within the life science ecosystem. • enable disciplines to provide rich discipline specific metadata description. • offer data storage and compute facilities. • federate distributed life science data resources and services using the FAIR principles.

3.3 ENVRI-FAIR

The ENVRI Science Cluster is composed of European Environmental Research Infrastructures that provide data and research key areas of the Earth system that encompasses, atmosphere, marine, solid sarth and biodiversity/terrestrial ecosystem. ENVRI-FAIR[4] aims to advance the findability, accessibility, interoperability, and reusability (FAIRness) of the data and services offered by the ENVRI Science Cluster Research Infrastructures.

The functional needs for the ENVRI-FAIR science projects and EOSC are described in the table below.

Table 3-3 ENVRI-FAIR science project requirements

Science case(s) description from the users perspective	Requirements for ENVRI-FAIR	Requirements for EOSC Architecture
• A researcher is working to demonstrate the impact of climate change on the	ENVRI-FAIR should: • enable access to historical data in the ENVRI FAIR	EOSC should: • provide generic infrastructure services such as for

biosphere. A promising option is to investigate the rapid increase of non-Indigenous Invasive Species (NIS) in European ecosystems. Such research requires access to big datasets (from genomics to in-situ and satellite borne environmental data) and high computational power, especially for those models with iterative algorithms. • A researcher is investigating whether the relationship between earthquake swarms (increasing frequency locally) and volcanic eruptions indicate an imminent eruption of a volcano. If so, can the extent of the lava flows, ejected material clouds and possible pyroclastic flows (such as engulfed Pompei) be predicted? Given the above, what is the size of the threatened population, where are they located, what measures are available for evacuation by road, air, rail, sea? Are fire services adequate? Is water supply adequate? This requires access to SSHOC and (EOSC-Life/Excelerate/Elixir) and public health - it will be necessary to know of persons at health risk (e.g., asthmatics). • A similar scenario would be an Icelandic eruption and effects on air travel; requires analysis of effect of particles (and particle size) on blades of jet engine and other relevant information to make an assessment.	clusters as well as to other environmental domains and social science domains (SSHOC). • offer repositories and High-Throughput Computing (HTC)/High-Performance Computing (HPC) resources and data management services. • connect the analytical framework and federate access to relevant data infrastructures at the EOSC portal to mobilize and empower a larger community of researchers and potential data providers. • improve the FAIRness of the data gathered by the ENVRI-FAIR cluster. • offer data and research services all within an ecosystem of single sign on, consistent governance/access permissions, licensing allowing composition/orchestration of workflows at any RI node utilising assets at that RI and others. • Provide solutions that go beyond downloading dataset, but that can offer assisted workflow composition/orchestration using heterogeneous assets described (catalog metadata) homogeneously within a homogeneous governance / AAAI.	Federated AAI, PID, and provenance, for tailoring to specific Research Infrastructure needs and adoption by individual Research Infrastructures. • enable access to shared resources such as repositories, HPC, HTC and data management tools. • provide standard APIs to support remote data discovery, access, and sharing.

3.4 ESCAPE

The European Science Cluster of Astronomy & Particle physics ESFRI Research Infrastructures (ESCAPE) [10] brings together a large fraction of the European Research Infrastructures in Astronomy, Astrophysics, Particle and Nuclear Physics.

The ESCAPE project is building a Virtual Research Rnvironment (VRE) for the Astronomy, Astro-Particle, Particle, and Nuclear Physics communities, as a prototype of the European Open Science Cloud (EOSC).

Via this, ESCAPE aims to address the Open Science challenges shared by its partners and the community. These challenges are technical, operational, sociological and scientific. Open Science allows scientific information, data and outputs to be more widely accessible and harnessed.

The functional needs for the ESCAPE science projects and EOSC are described in the table below.

Table 3-4 ESCAPE science project requirements

Science case description from the users perspective	Requirements for ESCAPE	Requirements for EOSC Architecture
• Researchers in ESCAPE are carrying out experiments for dark matter research. He/she needs to connect results and potential discoveries from different experiments; this requires the engagement of all scientific communities involved - astrophysics, particle physics and nuclear physics. The researcher needs to collect all the digital objects related to those analyses (data, metadata and software) on a broad platform, to enable open sharing and analysis of the various data sets. • Researchers studying extreme phenomena in the Universe in ESCAPE RI's would like to build a sustainable platform for Multi-Messenger Astronomy (MMA) that allows combined analysis of astronomical instruments observing the same phenomena with different probes, for example a gravitational wave event triggering follow-up observations with different telescopes, over time periods from seconds to days and months.	ESCAPE needs to: • ensure that the ESCAPE AAI is connected to the EOSC AAI. • provide federated storage services to ensure that all of the data sets required are openly accessible to all participants. • offer a data transfer service for moving data around. • provide federated storage and computing services, some of which provided through EOSC provisioning, and integrated with ESCAPE Data Lake. • enable access to other HPC and other compute resources provided by other Research Infrastructures and projects (i.e. Fenix RI, EOSC-Future). • provide access to a VRE for large scale data analysis, based on a notebook service, with access to a scalable computing back-end, to process data on the federated data service.	EOSC should: • provide a Federated AAI that provides an interoperability layer across research community AAIs and with the EOSC Core Services. • Access to shared resources such as repositories, HPC, HTC and data management tools. • federate existing resources across national data centres, e-Infrastructures to access and reuse data produced by the ESFRI projects in astronomy and particle/nuclear physics. • offer a federated digital repository for data preservation. • include an open source repository of analysis, computing and storage services.

3.5 SSHOC

The Social Sciences and Humanities Open Cloud Science Cluster (SSHOC [16]) brings together a large fraction of the European Research Infrastructures in Sociology, Psychology, Economics, Political Sciences, Anthropology, History, Languages, Arts, Cultural Heritage and other SSH disciplines.

The SSHOC cluster project is working to transform the current social sciences and humanities data landscape with its disciplinary silos and separate facilities into an integrated, cloud-based network of interconnected data infrastructures. Within the EOSC-Future project, two science projects will investigate specific use cases and build appropriate solutions to address them.

The functional needs for the SSHOC science use cases and EOSC are described in the table below.

Table 3-5 SSHOC science project requirements

Science cases description from the users perspective	Requirements for SSHOC	Science cases description from the users perspective
• A researcher wanting to access data or instruments found in a data catalogue, either local, domain-specific or EOSC-wide, wants to have a single point of access and a unified application process to apply for this access. A common platform with defined interfaces and standards is required to facilitate domain independent application processes.	SSHOC should: • work with other clusters, in this case ENVRI-FAIR and EOSC-LIFE to align various metadata standards and controlled vocabularies (DDI-CDI as node). • ensure AAI interoperability. • define extended AAI user profile to include user specific training and certification • define processes and standards for researchers to apply for access to restricted data and instruments. • integrate catalogue and data services • develop automatic harvesting, transformation, merging and processing. • enable easy access to research results for other collaborations.	EOSC should provide: • a Federated AAI that enables the research communities AAIs to interoperate and supports researcher qualification profiles. • environments for storage, sharing, accessing and using data, coupled to compute resources for the (re)analysis of data. • a definition of a common standard/controlled vocabularies of access restrictions. • allow access to restricted datasets in a common EOSC storage solution. • a definition and adoption of common open standards for interoperability. • a marketplace ecosystem for services and data available to all researchers.

4 Analysis of the current projects requirements

Looking at the tables above it is easy to identify common requirements across the Science Clusters. The common requirements can be summarised as follows:

1. provide a common EOSC AAI for all researchers,
2. define a common standard for FAIR data across communities,
3. provide a powerful search engine for data and services,
4. provide access to high performance storage, computing, archiving, simulation and analysis services,
5. define common standards for data and metadata to federate different catalogue of services,
6. make cluster community services available to the scientific community via EOSC and across clusters.

The rest of this chapter will elaborate on the first two (2) requirements (EOSC AAI and FAIRdata) as they are common to all science clusters, mostly relevant to WP3 in this phase of the project and are needed to create the foundation of EOSC. The other requirements will be analysed in the next revisions of the document.

As already stated above, this requirement list should not be considered exhaustive. It will be extended during the project lifetime gathering further requirements from Research Infrastructures, other EOSC projects and relevant initiatives.

4.1 Common EOSC Authentication and Authorisation Infrastructure

The need for a common EOSC AAI emerges in all discussions around EOSC and its implementation. The purpose of an EOSC AAI is to establish a common global ecosystem for identity and access control to services of the EOSC.

Prior to the start of the EOSC Future project the EOSC Executive Board had established several Working Groups, one of which focused on the EOSC architecture (Architecture WG). This WG, which finished its mandate in December 2020, had created an Authentication and Authorisation Task Force (AAI Task Force [3]) to establish a common global ecosystem for identity and access control infrastructures for the EOSC.

EOSC Future builds on the output and recommendations contained in the report [3] produced by the EOSC Authentication and Authorisation Task Force. The report states that the goal of the EOSC AAI is not to define a new AAI architecture, but rather to provide an AAI design that follows the AARC BluePrint Architecture (BPA) and the AARC Interoperability Guidelines [1] and to work with the international community through AARC and AEGIS for shaping the upcoming versions of the Blueprint Architecture to meet the evolving needs of EOSC. In this way it will be interoperable with existing and evolving AARC compatible AAI systems in use across the clusters.

The AARC Blueprint Architecture (BPA) provides a set of building blocks for software architects and technical decision makers who are designing and implementing access management solutions for international research collaborations. To date it has become a de-facto standard for research communities that want to build their own AAI in an interoperable way. The latest version of the AARC BPA distinguishes between two types of AAI services: one focuses on infrastructure management, while the other focuses on community management. Both types of AAI services may comprise the same interfaces (e.g., a proxy), but their functionality and their organisational purposes differ.

The AARC-BPA introduced the logical separation between the Community AAIs and the Infrastructure Proxies with the assumption that the number of Community AAIs can grow to a high number, but the number of the Infrastructure Proxies would remain low. This assumption is being challenged as in EOSC we are expecting that many Services will be made available from the national components of EOSC and probably, National Research and Education Networks and National/European/Global Research Infrastructures and will be operating their own Infrastructure Proxies. As an example, most of the clusters mentioned in this report operate an AARC-compliant community AAI; equally the three major e-Infrastructures, EGI, EUDAT and GÉANT implement an AARC-compliant AAI infrastructure.

As indicated in the AAI Task Force report, the interoperability across the research community AAI and the EOSC Services is addressed through the introduction of the EOSC AAI Federation, that encompasses the EOSC Services to the EOSC; Infrastructure Proxies that aggregate other service providers and Research Community AAIs that allow communities to manage their user membership and the access rights to shared services and resources. Acceptance to the EOSC AAI federation will be conditional on compliance with the EOSC AAI Federation Participation Policy which is under preparation; the criteria for such a policy are described in the AAI Task Force report.

This implementation of the EOSC federation is currently being addressed in the EOSC Service Planning and Delivery work package (WP7) of EOSC Future. WP3 and WP7 collaborate to promote guidelines and best practices. The deployment of the EOSC federation will highlight the needs for additional features and will identify possible operational and architectural challenges that need to be reflected in the EOSC Architecture.

A new AAI Task Force [2] has been created by the EOSC Association and it is expected to start shortly; this will be an open group in which WP3 will participate to address topics that are in the scope of the work package. The task force will focus on the evolution of the EOSC AAI Federation.

4.2 Common standard for FAIR data across communities

When talking about EOSC it is impossible not to mention FAIR data, which refers to data that are Findable, Accessible, Interoperable and Reusable. The majority of reports and studies on FAIR practice focus on research data, but there is increasing consensus that FAIR principles can and should be applied to other research objects (i.e. software, scientific workflows, etc.).

As the science projects above highlighted, researchers are increasingly asking to be able to reuse data from previous research and this can only be done if the data is FAIR.

Other projects and initiatives paved the way to EOSC Future, including the FAIR Practice Task Force, set up as one of the four task forces of the EOSC Executive Board FAIR Working Group.

The FAIR Practice Working Group, taking into account previous projects results and global initiatives, identified key obstacles and recommendations towards the implementation of FAIR practices and therefore to interoperability in EOSC, as documented in the Interoperability Framework [12] and Six Recommendations for Implementing FAIR Practice [14] reports.

The reports highlighted that although there are many generic and many data-type or discipline-specific repositories, some fields lack specific repositories (e.g. earth sciences) or repositories that can deal with complex outputs ('complex digital objects') (humanities) or insufficient infrastructure for transferring and archiving large data to/from repositories. Also reported is a lack of sufficiently flexible and secure infrastructure for archiving data (this becomes even more critical for medical data). The reports stated that 'There is a need for a minimum metadata application profile for the EOSC context to allow users to discover and deal seamlessly with data available in multiple generic or community-based formats' and that 'When searching for research data (or other research objects) that may be reusable across communities, such data may need to be discovered at different levels of granularity: high level / coarse-grained (e.g., look for data about DNA sequences or land-use) or low level / fine-grained (inside data collections, e.g., look for a specific DNA sequence or land-use in a given town)'.

Beside the aspect of discovering data, interoperability came up as the biggest challenge. This is due to the format in which the data are expressed (for instance in multiple general-purpose formats -CSV, Excel, XML, etc - or community-based models - Darwin Core (Standard maintained by the Darwin Core maintenance group to facilitate the sharing of information), FITS, VOTable (XML standard for the interchange of data represented as a set of tables), VOResources (XML Encoding Schema for resource metadata), NetCDF, the metadata associated with the data, the vocabularies used and the (lack of) qualified references to other meta(data).

There is a clear need for principled approaches and tools for ontologies and metadata schema creation, maintenance and governance and use across disciplines.

In addition, there are insufficient ways of automatically collecting, updating and preserving metadata.

FAIR cannot be achieved without globally unique identifiers (i.e. the same identifier cannot be reused/reassigned without referring to the original data) and persistent identifiers (i.e. it can be resolved in the future); as indicated in the FAIR Working group reports 'there is a need to have a common and well-understood PID policy across communities'.

Work in this area continues in the EOSC Association, via dedicated Task Forces that will be launched shortly as well as in WP3, via dedicated Working Groups that will be starting soon, to address aspects that are of immediate relevance for the interoperability framework and that complement work done in the EOSC Association.

5 Next Steps and Conclusions

This document provides a snapshot of the EOSC Future science projects and summarises their requirements. It also describes in some detail two of these requirements for which work is already underway.

The next revision of this document (in about six (6) months) will highlight the progress made and will present the work done to federate different service catalogues and to make services discoverable across communities, a clear need that emerged from the analysis of the science projects requirements. Some of this work is ongoing in the EOSC Enhance project[3] and some has started in EOSC Future. A transition team is being created to migrate EOSC Enhance work to the EOSC Future project.

This deliverable (and any subsequent update) is one of the instruments to gather, document, and assess requirements to drive the evolution of the EOSC interoperability framework; more information on the process to manage the interoperability framework will be provided in the dedicated deliverables that will be released throughout the EOSC Future project duration (first version is expected by the end of 2021).

[3] The EOSC Enhance Project (https://eosc-portal.eu/enhance) ends in November 2021, hence the need to transition results.

References

[1] AARC Blueprint Architecture (BPA)
https://aarc-community.org/architecture/

[2] AAI Task Force (EOSC Association)
https://www.eosc.eu/sites/default/files/tfcharters/eosca_tfaaiarchitecture_draftcharter_20210614.pdf

[3] AAI Task Force Report (EOSC Executive Board)
https://op.europa.eu/en/publication-detail/-/publication/d1bc3702-61e5-11eb-aeb5-01aa75ed71a1/language-en/format-PDF/source-188566729

[4] ENVRI-FAIR
https://envri.eu/home-envri-fair/

[5] EOSC Executive Board AAI Task Force
https://op.europa.eu/en/publication-detail/-/publication/d1bc3702-61e5-11eb-aeb5-01aa75ed71a1/language-en/format-PDF/source-188566729

[6] EOSC Executive Board Working Groups
https://eoscsecretariat.eu/eosc-governance/eosc-executive-board-outputs

[7] EOSC Future Wiki
https://wiki.eoscfuture.eu/pages/viewpage.action?pageId=1805501

[8] EOSC-Life
http://www.eosc-life.eu/

[9] EOSC-Life Report for regulatory compliance
https://zenodo.org/record/3820390#.YRu6vtNKg-Q

[10] ESCAPE
https://projectescape.eu/

[11] ExPaNDS
https://expands.eu

[12] Interoperability Framework (Report from EOSC Execute Board Working Groups FAIR and Architecture)
https://op.europa.eu/s/somA

[13] PaNOSC
https://panosc.eu

[14] Six Recommendations for Implementing FAIR Practices
https://ec.europa.eu/info/publications/six-recommendations-implementation-fair-practice_en

[15] SRIA
https://eosc.eu/sites/default/files/EOSC-SRIA-V1.0_15Feb2021.pdf

[16] SSHOC
https://www.sshopencloud.eu/project

[17] EOSC-LIFE Report on requirements for regulatory compliance of sensitive health data and biological and medical research data.
https://zenodo.org/record/3820390#.YUsbnbgzZRV