DG CNECT
Unit C.1

Project Charter

European Open Science Cloud (EOSC) Platform

Date: 11/03/2024
Doc. Version: 1.1
Template Version: 3.2.1

This template is based on PM² V3.0

For the latest version of this template please visit the PM² Wiki

Document Control Information

Settings	Value
Document Title:	Project Charter
Project Title:	[redacted]
Document Author:	[redacted]
Project Owner:	[redacted]
Business Manager:	[redacted]
Solution Provider:	[redacted]
Project Manager:	[redacted]
Doc. Version:	1.1
Confidentiality:	[redacted]
Date:	

Document Approver(s) and Reviewer(s):

NOTE: All Approvers are required. Records of each approver must be maintained. All Reviewers in the list are considered required unless explicitly listed as Optional.

Name	Role	Action	Date
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]

Document history:

The Document Author is authorised to make the following types of changes to the document without requiring that the document be re-approved:

- Editorial, formatting, and spelling
- Clarification

To request a change to this document, contact the Document Author or Owner.
Changes to this document are summarized in the following table in reverse chronological order (latest version first).

Revision	Date	Created by	Short Description of Changes
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]

Configuration Management: Document Location

[redacted]

TABLE OF CONTENTS

1 EXECUTIVE SUMMARY 5

2 CONSIDERATIONS ON THE BUSINESS CASE 5

2.1 Analysis of Possible Alternatives 6
2.2 Interrelations and Interdependencies 6
2.3 Expected Outcomes 10

3 PROJECT DESCRIPTION 10

3.1 Scope 10
3.1.1 Includes ("IN" Scope) 10
3.1.2 Excludes ("OUT" Scope) 12
3.1.3 Scope Statement 13
3.2 Success Criteria 14
3.3 Stakeholder and User Needs 14
3.4 Deliverables 16
3.5 Features 17
3.6 Alignment with EC Digital Strategy (ECDS) 18
3.7 Constraints 19
3.8 Assumptions 20
3.9 Risks 21

4 COST, TIMING AND RESOURCES 23

4.1 Cost 23
4.2 Timing and Milestones 26
4.3 Planned Resources 28

5 APPROACH 30

5.1 Methodology 30
5.2 Change Management 31
5.2.1 Project Change 31
5.2.2 Configuration Management 32
5.2.3 Organisational Change 32

6 GOVERNANCE AND STAKEHOLDERS 34

6.1 Structure 34
6.2 Roles and Responsibilities 34
6.3 Other Stakeholders 35

APPENDIX 1: REFERENCES AND RELATED DOCUMENTS 36

APPENDIX 2: SYSTEM ELEMENTS – ARCHITECTURE DETAILS 37

1 HIGH LEVEL ARCHITECTURE DIAGRAM 37

2 COMPONENTS DESCRIPTION 38

2.1 LOT1 - Core Federating Services 38
2.1.1 Web Portal Front Office 38
2.1.2 Multifaceted Resource Catalogues and Registry Services 38
2.1.3 Application Workflow Management 39
2.1.4 Identity Management (Single Sign On) 40
2.1.5 Monitoring and Accounting 40
2.1.6 Service Management System 41
2.2 LOT2 - Exchange Infrastructure Services 42
2.2.1 Container Platform Service 42

2.2.2 Compute Service 43
2.2.3 Bulk Data Transfer Service 44
2.3 LOT3 - Exchange Application Services 45
2.3.1 EFSS – Enterprise File Synchronisation and Sharing Service 45
2.3.2 Interactive Notebooks 46
2.3.3 Large File Transfer Service 47

1 EXECUTIVE SUMMARY

The European Commission intends to unlock the reuse potential of different types of data and facilitate its free flow across borders to achieve a European digital single market, for the benefit of the economy and society as part of the European Common Data Spaces strategy. The European Open Science Cloud (EOSC) is a fundamental enabler of open science and of the digital transformation of science, being the de-facto data space for Research and Innovation. It offers the possibility to access and reuse all publicly funded research data and other research objects (e.g., publications, software, services, tools, etc.) in Europe, across scientific disciplines and countries. By federating existing research data infrastructures, EOSC leverages national investments and adds value in terms of scale, interdisciplinarity and faster innovation.

To design and deploy the first reference service instance of the EOSC ecosystem, hereafter referred as the EOSC EU Node, and to roll that out into production operated 24/7, the Commission procured the "Managed Services for the European Open Science Cloud (EOSC) Platform" (outsources model). In the context of this Project Charter, the awarded third-party contractors are going to implement, operate, maintain, and support the EOSC EU Node services for 36 months. The service owner is represented by the European Commission, DG CNECT, Unit C.1.

[redacted]

[redacted]

2 CONSIDERATIONS ON THE BUSINESS CASE

[redacted]

Key findings of the Business Case remain valid to the EOSC EU Node platform services. The major development since the submission of the Business Case has been the successful completion of the competitive dialogue type of public procurement action that resulted in three awarded consortia providing the managed services for the EOSC EU Node platform officially started on 14 February 2024.

Due to the competitive dialogue type of procedure, the detailed technical specifications of the tender cannot be shared publicly therefore they are not referenced to this Project Charter. However, the Descriptive Document of the tender gives a high-level description of the scope and requirements of the EOSC EU Node managed services that is publicly available in TED at: https://etendering.ted.europa.eu/cft/cft-documents.html?cftId=12087

The European Open Science Cloud (EOSC) champions research data management and application to guarantee scientists' access to data-driven science. Supporting the EU's Policy of Open Science, EOSC aims to give the EU a global lead in research data management and ensure that European scientists enjoy the full benefits of data-driven science. EOSC is also part of the European Data Strategy, aiming to make the EU a leader in a data-driven society. It will provide seamless access and reliable re-use of research data to European researchers, innovators, companies and citizens through a trusted and open distributed data environment and related interoperable services.

The priorities for building EOSC are set jointly by the European Commission and the EOSC Association through the co-programmed European Partnership (the EOSC Partnership) in the Strategic Research and Innovation Agenda (SRIA). Horizon Europe continues to support building EOSC on basis of the baseline that started in the Horizon 2020 programme.

The managed operational services for the first reference EOSC EU Node have been procured for the initial contract duration of 36 months (including the managed EOSC Core Platform and the horizontal EOSC Exchange services available 24/7) and owned by the Contracting Authority (DG CNECT C.1 is the service/business owner) made available to all the EU Member States and Associated Countries.

2.1 Analysis of Possible Alternatives

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

[redacted]

2.2 Interrelations and Interdependencies

[redacted]

It's been identified in the Business Case that the procured EOSC EU Node platform is going to operate in the cross-section of the Scientific Research and Academic domain and the Public Administration sector.

- On the one hand, it should fulfil the expected outcome described in the Horizon Europe Work Programme serving the needs of the scientific research community (outside of the organization);
- On the other hand, it must comply with the Commission's internal rules, regulations and policies (IT, Security, Data Protection, Visual appearance, etc.) as well as technically connect to the existing internal systems relevant to EOSC (e.g., EU login, Cellar/CORDIS, data.europa.eu) and promote the use of others (inside of the organization).

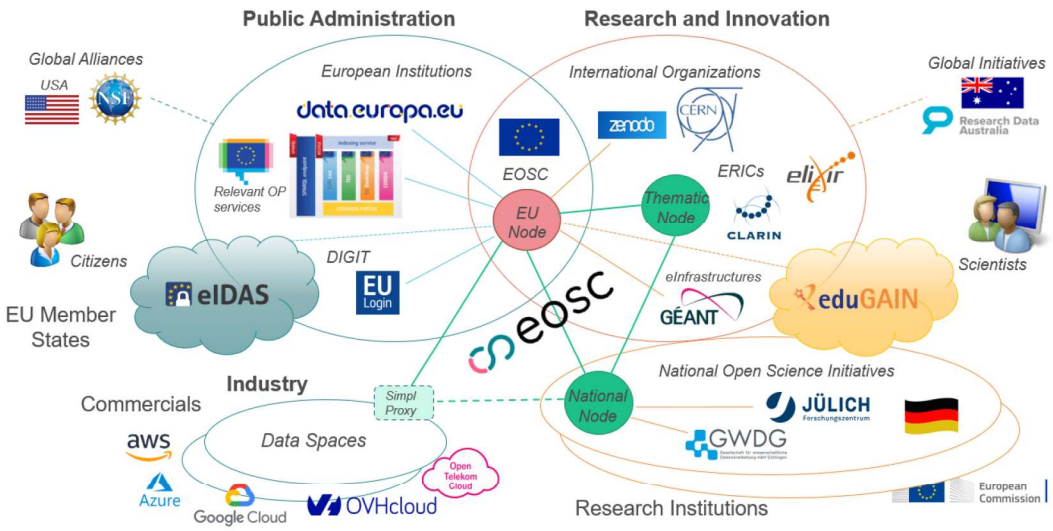


Fig. 1. Interrelations and interdependencies

Services typically used by the targeted user community today in the public administration domain include:

- EU login (eIDAS) for authentication and access control
- Cellar/CORDIS for EU research results and other resources
- data.europa.eu for open data sets
- Open Research Europe for open publications workflow
- EU Vocabularies for knowledge management

Services regularly used by the targeted user community today in the scientific research domain include:

- Research Infrastructures (ERICs) in 5 disciplinary clusters under ESFRI
- eInfrastructures for generic compute, storage, network (GEANT, EGI, EUDAT) as well as the EuroHPC JU.
- Data Infrastructures for open resources and knowledge management (OpenAIRE, Zenodo, Software Heritage, etc.)
- eduGIAN (GEANT) for inter-federating national access federations (AAI) for research and education.

[redacted]

[redacted]

[redacted]

2.3 Expected Outcomes

The expected outcomes of EOSC submitted in the PIR remains valid:

- A fully operational, secure cloud-based EOSC infrastructure, including a federated core platform and the EOSC Exchange, offering high quality professional services and providing for a superior user experience, usability and ease of use for a very large number of users, with the functionalities available (i.e. operated and maintained) 24/7.
- Population of EOSC with a rich set of innovative, modular, customisable and composable services for a wide variety of users from the research communities and beyond.
- A large number of data and service communities aligned in terms of standards and consolidated at subdomain, domain and interdisciplinary levels.
- Established links with common European data spaces in crucial sectors, such as green deal or health, and synergies with the work on the European cloud federation as described in the Member States' joint declaration on building the next generation of cloud in Europe.
- Increased discovery and reuse of European research output as a result of FAIR data and services provided through EOSC, and cross-fertilisation and a wider sharing of knowledge and technologies.

In addition to these, the following specific outcomes of the procured EOSC EU Node platform services have been clarified after the procurement:

The managed services of the EOSC EU Node constitute a European level multi-disciplinary and multi-national scientific data/service portfolio for all research users and citizen scientists. Until 2027, the minimum viable services are owned by the EC and governed by the EOSC Tripartite Governance (European Commission, EOSC Association, and Member States/Associated Countries representatives). The future ownership is under discussion with the stakeholders.

- In particular, the procured EOSC EU Node has the following value propositions: Facilitates the creation of the "Web of FAIR data and interoperable services" (referred as the EOSC Federation) under the Open Science Policy.
- Puts a "seed in the ground" by operationalizing the first recognised/reference EOSC Node at the European level for the initial 3 years of service duration.
- Offers core services for scientific research infrastructures to federate (i.e., EOSC Core services such as: single-sign-on, catalogues, knowledge graph, application workflow, monitoring, accounting, helpdesk) and common horizontal services for end-users to benefit from (i.e., EOSC Exchange services such as: compute, containers, data transfer, notebooks, file sharing, open research data).
- Defines the pathway and blueprint (i.e., EOSC Architecture and Interoperability Framework) for other potential EOSC Node operators to join the federation.

The EOSC EU Node has an open concept and is only the first node of the federation. National, regional and/or thematic service providers as well as autonomous EOSC Nodes can connect to the federation via the established interoperability frameworks and policies. The SIMPL (Smart Middleware) Agent proxy implementation ensures connectivity to other industrial Data Spaces. EuroHPC resources may also be offered to the EOSC Federation.

3 PROJECT DESCRIPTION

3.1 Scope

3.1.1 Includes ("IN" Scope)

The European Open Science Cloud (EOSC) is to build and deploy a fully operational enabling reference infrastructure for EOSC – referred to as the EOSC EU Node – providing access to a rich portfolio of FAIR (Findable, Accessible, Interoperable, Reusable) data and professional quality interoperable services in all relevant domains from data handling to computing, processing, analysis, sharing/sending and storing.

The following business capabilities and technical functionalities of the EOSC EU Node are in scope of the project:

Core Federating Services for the EOSC EU Node

EOSC Core is the set of internal services which allow the EOSC federation to operate. It includes a core technical platform (i.e., EOSC EU Node Core Platform) which facilitates EOSC operations in which the researcher-facing resources in the EOSC Exchange can be integrated as appropriate. It also includes non-technical coordination functions which operate and facilitate the technical platform, such as the onboarding and security coordination.

- The winner of the lot is the Open Science Agora Consortium.
- The consortium is coordinated by Athena Research Center (ARC), and participated by EGI Foundation, OpenAIRE A.M.K.E and Netcompany Intrasoft SA. The subcontractors are: GÉANT Association, Greek Research & Technology Network (GRNET S.A.), and Scientific Compute and Competence Centre of University of Göttingen and the Max Planck Society (GWDG mbH). The underlying infrastructure providers are CESNET z. s. p. o, Interdisciplinary Centre for Mathematical and Computational Modelling (ICM) at University Warsaw, and GRNET S.A. offering their community and public cloud tenancies.
- The consortium is going to provide professionally managed services for the core components of the EOSC EU Node including functions such as:
    - Web Portal Front Office,
    - Multifaceted Resource Catalogues and Registry Services, - connected to open.data.eu, Cellar/CORDIS, Open Research Europe, EU Vocabularies and others.
    - Application Workflow Management engine,
    - Federated Identity Management and Single-Sign-On solution, - reusing EU Login
    - Monitoring and Accounting function, and
    - overall Service Management System and service integration.

EOSC Exchange is the set of services and other resources registered into EOSC by Service/Data Providers (such as eInfrastructures, Research Infrastructures, Science Clusters, Commercials) to serve the needs of research communities and will widen its offering to the public and private sector. Generic services and resources which target heterogeneous scientific domains and research communities are identified as 'horizontal services'. Resources which target users from a specific science, community, and/or regional domain are identified as 'thematic and/or regional resources'.

There are two types of EOSC Exchange services: infrastructure services and application services.

All the operational procedures related to running managed IaaS and PaaS will be preformed by the hosting sites of awarded contractors' groups. IaaS and PaaS services that are backing the managed service offerings (VMs, containers, data transfers, file sync and share, data science environment, file transfers) are operated by two partners: PSNC and Safespring, at their premises, namely data centres using servers, storage systems and network devices owned by PSNC and Safespring. Each of these two organisations will operate their own 'regions', where each region includes and independent set of physical resources: data centre, server hardware etc. as well as software resources.

The operational procedures implemented constantly and periodically by these sites are relevant to the usual operational processes implemented in data centres and cloud operators. They will include the general operations, 24/7 monitoring, periodic security assessments and audits, changes management related system and software components updates, maintenance process, failure handling activities, internal issues solving, problem escalations, resource usage monitoring and reporting etc.

No dependency of any of these processes is assumed related to any internal EC resources, systems and processes.

Exchange Infrastructure Services for the EOSC EU Node:

- The winner of the lot is the Poznan Supercomputing and Networking Center (PSNC).
- The subcontractors are: NORDUnet A/S representing the Norwegian Agency for Shared Services in Education and Research (Sikt), and Blue Safespring AB. The underlying network provider is the GÉANT Association.

- The contractor and its subcontractors are going to provide fully managed services for the infrastructure services component of the EOSC EU Node including:
    - Container Platform service,
    - Compute service and
    - Bulk Data Transfer service.

Exchange Application Services for the EOSC EU Node:

- The winner of the lot is the Poznan Supercomputing and Networking Center (PSNC).
- The subcontractors for this lot are: NORDUnet A/S representing the Norwegian Agency for Shared Services in Education and Research (Sikt), Blue Safespring AB, EGI Foundation, CESNET z. s. p. o, and ownCloud GmbH.
- The contractor and its subcontractors are going to provide fully managed services for the application services component of the EOSC EU Node including:
    - EFSS – Enterprise File Synchronisation and Sharing service,
    - Interactive Notebooks service and
    - Large File Transfer service.

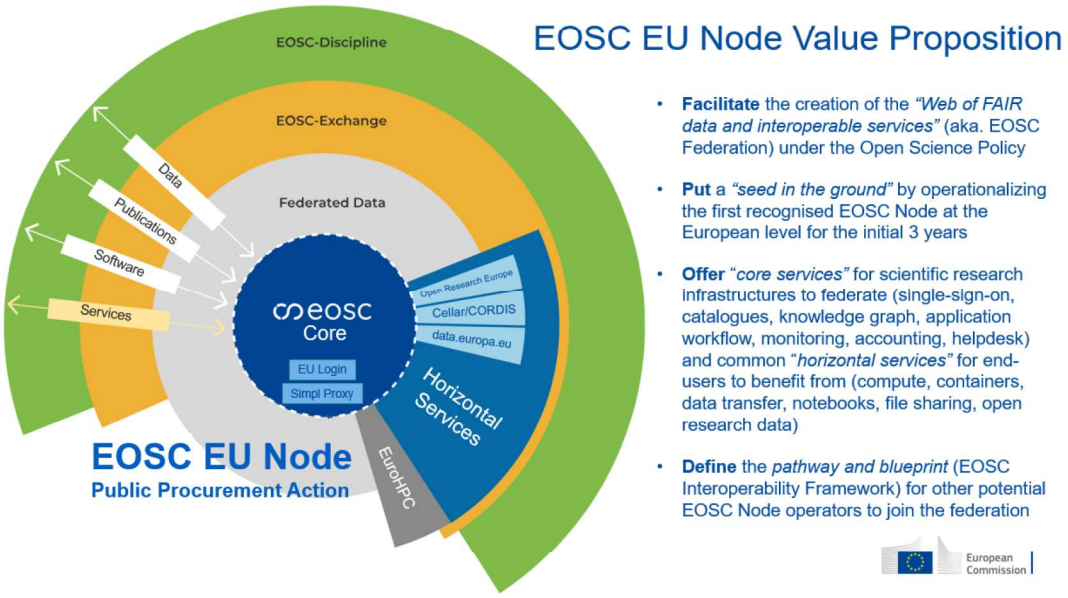


Fig. 2. In scope (blue) and out of scope (rest) of the EOSC EU Node platform services.

3.1.2 Excludes ("OUT" Scope)

In principle, the EOSC is not a single monolithic infrastructure or resource provider but is rather a federation (federated architecture) built out of many independent organisations and resource providers as in a System of Systems approach (i.e., web of FAIR data and services). As such, it ensures independence and autonomy of resource providers. Scientific resource/service providers are widely distributed across Europe, have the mandate to serve one or more research disciplines in various geographies, and have to comply with different national and European legislations. The vision of EOSC is to serve a wide variety of users and stakeholders (e.g., researchers, research infrastructures, service providers, service developers, funders, organizations, project managers, SMEs, citizens etc.). It is to create an ecosystem that facilitates easy access to already existing resources and to allow the EOSC users to build complex solutions out of a variety of resources.

As the EOSC is recognized as a System of Systems, its design principles are to be flexible and inclusive rather than selective, i.e., various metadata standards from communities should be acceptable, multiple service framework standards (service pipelining and application workflows, infrastructure/platform

services, etc.) adopted by the communities should be applicable. However, the system of systems approach to work properly and to ensure interoperability must define some boundaries and to select deployment options. Therefore, choices have to be made for this Project to select appropriate standards, best practices, tools and APIs that are suitable for the EOSC platform and services implementation at the European level.

The following elements are out of scope of the EOSC EU Node platform project:

Maintaining and governing the EOSC Interoperability Framework

- EOSC Interoperability Framework (EOSC IF) provides a flexible framework of standards and guidelines to support the interoperability and composability of resources in the EOSC Core and EOSC Exchange according to the recommendations of the recommendation of the European Interoperability Framework. The EOSC IF will allow EOSC providers to specify with which interoperability frameworks (e.g., standards, APIs, data models, exchange formats, protocols) the resources they are sharing comply. It will therefore act as the glue to connect services and research products (e.g., publications, datasets, and software) across providers in terms of their interoperability capability. 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;
- The Contractors of the EOSC Node at the European level are going to contribute to the EOSC Interoperability Framework with the implementation choices they apply. The coordination and maintenance of the EOSC IF is outside of the scope of the Project, the European Commission provides funding for that activity by other means.

Creating and governing the EOSC Federation

- The EOSC EU Node is the first reference implementation of the federated EOSC Node concept as a system of systems, it may serve as a blueprint for other potential node operators to define and establish their EOSC Nodes, either national or thematic focused. The Commission is only responsible and ultimately accountable for the operations of the EOSC EU Node and not the entire federation of autonomous EOSC Nodes in the research infrastructures' ecosystem.
- The facilitation and nurturing of the EOSC Federation and the governance of the federation level policies are funded by the Commission by other means (e.g., grants for CSAs under Horizon Europe Research Infrastructures Work Progamme).

Implementing SIMPL

- DG CNECT has successfully completed the procurement action for the Smart Middleware (Simpl) framework contact in parallel with the EOSC procurement contract. Simpl meant to be a preferred best practice protocol stack to interconnect European Common Data Spaces and potentially penetrate into the data spaces by promoting industry best practices for interoperability. EOSC is committed to run a feasibility study with Simpl, referred as Simpl-Live-EOSC Study, where the compatibility and re-usability of the Simpl stack will be investigated and assessed. EOSC EU Node is not going to implement Simpl as such but it is going to ensure a proxy approach to Simpl interfacing with other potential data spaces such as Green, Health, etc.

Competing with EuroHPC

- The infrastructure services of the EOSC EU Node exchange platform are primarily targeting the long tail of research users dealing with multi-disciplinary and multi-national scientific data sets. Having said that, particular demand for HPC compute-intensive resources will not be met by the EOSC EU Node. However, the application workflow mechanisms of EOSC is planned to be extended to HPC resource providers, including EuroHPC, that is not a technical issue rather a policy question. These policy questions are out of scope of the Project.

3.1.3 Scope Statement

The deployment of the European Open Science Cloud (EOSC) EU Node platform services by the Commission facilitates the creation of the federated "Web of FAIR data and interoperable services" promoting Open Science principles at large. The EOSC EU Node is operated and maintained in production 24/7 by third-party contractors and made available to all multi-disciplinary and multi-national scientific research communities in the EU Member States and Associated Countries and beyond, aiming to federate

existing research data infrastructures and scientific clusters in the European Common Data Space for Research and Innovation.

3.2 Success Criteria

The success criteria of the EOSC EU Node deployment and operations Project are as follows:

- Deployment of a fully operational, secure cloud-based EOSC infrastructure, including a federating EOSC Core platform and the EOSC Exchange, offering high quality professional services and providing for a superior user experience, usability and ease of use for a very large number of users, with the functionalities available (i.e. operated and maintained) 24/7.
- Population/Onboarding of EOSC with a rich set of innovative, modular, customisable and composable services for a wide variety of users from the research communities and beyond.
- A large number of data and service communities (ESFRI and clusters) aligned in terms of standards and consolidated at subdomain, domain and interdisciplinary levels.
- Established links with Common European Data Spaces in crucial sectors, such as green deal or health, where possible using the SIMPL Proxy and synergies with the work on the European Cloud Federation as described in the Member States' joint declaration on building the next generation of cloud in Europe.
- Increased discovery and reuse of European research output as a result of FAIR data and services provided through EOSC federation, and cross-fertilisation and a wider sharing of knowledge and technologies.

The success criteria of the EOSC EU Node platform services as a product are as follows:

- be robust, secure, scalable, flexible, data-driven and user-centric;
- constantly be improved and upgraded following user feedback and the state-of-the-art of the underlying core technologies;
- offer high quality of service management compliant with industrial standards and
- provide a superior user experience, usability and ease of use for a very large number of users, with the functionalities available 24/7;
- self-contained and managed to be easily handed over to a new legal entity (when necessary);
- provide access to the EOSC federation via the EOSC EU Node's web front-office hosted under the ec.europa.eu web domain.

3.3 Stakeholder and User Needs

ID	Need Description	Priority
1	***Researchers*** *Principal investigators of publicly funded research projects affiliated with academic institutions and research labs. They need to access FAIR data and interoperable services in order to design, execute, disseminate and assess research in a collaborative environment fostering Open Science principles. Researchers are typically funded in national context or disciplinary communities lacking multi-national and multi-disciplinary collaborations. EOSC is there to break the silos and onboard the key stakeholders into the EOSC federation.*	*High*
2	***Citizen scientists*** *The long tail of research users includes citizen scientist willing to experiment with publicly available FAIR data sets and benefiting from interoperable scientific services, tools and applications hosted on sovereign cloud environment for*	*Low*

	curiosity-based research subjects. Currently citizen scientists have difficulties to find trusted data sources and execute research workflows on dedicated research infrastructure. EOSC EU Node is these to provide a neutral place to any citizen scientist to engage with muti-disciplinary experiments.	
3	***IT administration staff*** *Not all researchers are experts of digital tools and IT systems needed to execute research workflows often in a complex hybrid cloud environment. IT administration staff is often tasked to support researchers making the right infrastructure and application choices available to them. EOSC is there to provide a catalogue of these infrastructure and application services that are interoperable and generating FAIR research outputs by design.*	*Mid*
4	***Research Infrastructures*** *Research Infrastructures are primarily funded to serve a particular need of their core researchers working on a given filed of science. These are often specialized in a given instrumentation and generating huge volume of data. Making these data FAIR resulted in an increasing number of data visitations by alien researchers of other fields. RIs are not funded to accommodate the growing demands. EOSC EU Node is there for RIs to connect to and to burst out scientific workflows generated by data visitation of multi-disciplinary scientific use cases.*	*High*
5	***eInfrastructure Service Providers*** *eInfrastructure service providers are offering horizontal services commonly available to all researchers such as compute, storage, network, data analytics, publications, etc. These horizontal services are currently often siloed with various access policies, engagement models, funding options. EOSC is there to harmonize and federate these eInfrastructure services under common service delivery workflows that are data-driven and user-friendly.*	*High*
6	***Research Data Providers*** *Research data is stored in the plethora of repositories with various levels of FAIR-ness implanted in them. Some research data often remains with scientific instruments or the researcher themselves not being able to be shared or reused in any ways. EOSC is there to harmonize schemas and harvest metadata from repositories making their data sets findable and accessible. EOSC also helps researchers to deposit their data sets in repositories ensuring FAIR treatment by design.*	*High*
7	***Software/Tool Providers*** *Opensource software tools are developed and made available under various opensource license schemes that are often not backward compatible and preexisting rights are undocumented. Tools developed by researchers are typically not mature and sustainable enough to be mainstreamed in any major distributions of scientific software or code. EOSC is there to share and reuse opensource software in a trusted and reproducible environment.*	*High*
8	***Publishers*** *Today, researchers directly or indirectly by their institutions pay significant amount of money for publishing in scientific*	*Low*

	journals dominated by strong publishers. The Open Research Europe publishing platform of the Commission is addressing this problem by offering a free alternative. EOSC is there to integrate the open publishing service with the rest to the research lifecycle making the design, execution, dissemination and assessment phases of research seamlessly coupled in a simple workflow.	
9	***Members States and Associated Countries*** *Member States are defining their open science policies and facilitating their domestic research communities to open up the scientific workflows and outcomes. The EOSC EU Node provides a blueprint architecture and reference implementation for interested MS/AC to define and deploy their national EOSC Nodes and join the EOSC Federation mainstreaming open science principles across Europe and beyond.*	*High*
1o	***Science Clusters*** *Science clusters are serving thematic research communities that are typically already multi-national across Europe. Having issues with the growing demand of cross-disciplinary use of secondary data, science clusters – primary those incorporated into ERICs already – have the opportunity to define their own thematic EOC nodes and join the EOCS federation facilitating multi-disciplinary scientific user scenarios.*	*High*

3.4 Deliverables

ID	Deliverable Name	Deliverable Description
1	***IT Governance documentation*** *Project/Programme Charter* *Architecture Design Plan and the Architecture Canvas: including hosting, DNS, network* *IT Security Plan: security model, security architecture and IT security impact assessment* *Technical Data Protection Plan and Data Protection Impact Assessment (questionnaire)* *Operational Disaster Recovery Plan (including implementation plan)* *Evaluation of the IT Security Plan: provide input to the IT Security Risk Report* *Service Interoperability Plan*	*All the documents, templates and information provided for online tools such as the Architecture Canvas, Global Risk Catalogue and others that are needed for IT Governance processes, approvals and production launch of the EOSC EU Node services.*
2	***Data protection policies and procedures*** *Risk and Compliance Assessment Plan: Controls for risk, compliance, continuity & recovery + cost-benefit analysis* *Risk Registry* *Personal Data Protection Impact Assessment (GDPR)* *Data Processing Agreement*	*All documents, templates and information provided for online tools such as the Data Protection Impact Assessment and others that are needed for approvals and production launch of the EOSC EU Node services.*

3	***Deployment strategies, integration plans, charters, progress, risks and timelines*** *Deployment Plan* *Configuration Plan*	*All the deliverables needed for documenting the deployment and configuration plans and actions.*
4	***Operational quality plans, service reviews, verification and test session results and defects status and resolution*** *Capacity Plan* *Verification, Validation and Testing Plan* *Operations, Maintenance and Support Plan* *Incident Reporting Plan (including setup of ticketing system and workflows)*	*All the operational related deliverables including testing, Quality Assurance Management, operations, maintenance and support including incident handling.*
5	***Stakeholder and community engagement strategy*** *Communications Plan* *Documentation Plan: user/admin manuals and release notes* *Training Development Plan*	*All the documents and plans related to stakeholder management addressing all the profiles described in the previous section.*
6	***Production roll-out of service components*** *EOSC Core* • *Web Portal Front Office,* • *Multifaceted Resource Catalogues and Registry Services, - connected to open.data.eu, Cellar/CORDIS, Open Research Europe, EU Vocabularies and others.* • *Application Workflow Management engine,* • *Federated Identity Management and Single-Sign-On solution, - reusing EU Login* • *Monitoring and Accounting function, and* • *Service Management System* *EOSC Exchange* • *Container Platform Service,* • *Compute Service,* • *Bulk Data Transfer Service.* • *EFSS – Enterprise File Synchronisation and Sharing Service,* • *Interactive Notebooks and* • *Large File Transfer Service.*	*The actual managed service components of the EOSC EU Node including the EOSC Core and the EOSC Exchange Infrastructure and Application services and their integrations.*

3.5 Features

Architecture, functionalities, features and integration plans of all the 6 EOSC Core capabilities as well as the 3 EOSC Exchange infrastructure services and the 3 EOSC Exchange application services (listed under Deliverable 6 above) are detailed in the Appendix. The table below is only giving an overview of the service features related to the key stakeholder and user needs.

Related Need	Features	Deliverable(s)

All	*Common web portal front-office with dedicated user space for all the ESOC EU Node services.*	*5*
All	*Multifaceted Resource Catalogues and Registry services for all public users open access (no login required).*	*All*
1, 2, 3	*Application Workflow Management*	*All*
All	*Federated Identity Management and Single-Sign On*	*All*
All	*Monitoring and Accounting*	*All*
All	*Service Management System*	*All*
All	*Container Platform Service*	*All*
All	*Compute Service*	*All*
All	*Bulk Data Transfer Service*	*All*
1, 2, 3	*Interactive Notebooks Service*	*All*
1, 2, 3	*EFSS - Enterprise File Sync & Share Service*	*All*
1, 2, 3	*Large File Transfer Service*	*All*

3.6 Alignment with EC Digital Strategy (ECDS)

EC digital principle	How does the solution follow the EC digital principle?
Digital by default and Once-only	*The EOSC EU Node is a digital platform supporting the creation of the Web of FAIR data and interoperability services for the scientific research community and citizen scientist. In the federated EOSC Nodes ecosystem users can get access to EOSC from any of the Nodes relying on a common information model and connected infrastructure.*
Security and privacy	*The EOSC EU Node digital platform will be owned by the Commission, hosted and operated by the third-party contractors. As such, the solution fully complies with the EC security and data protection requirements.*
Openness and transparency	*EOSC is facilitating the Open Science principles. It aggregates scientific data and metadata from various repositories, including the Open Data Portal and provides a Metadata Knowledge Graph for federated search, findability and discoverability features. This will increase transparency of information to all stakeholders.*
Interoperability & Cross border	*EOSC is facilitating both multi-national and multi-disciplinary research scenarios by federating national, regional and thematic EOSC Nodes. The EOSC Interoperability Framework is defined based on the European Interoperability Framework and provides industry best practices and consensus-based protocols, interfaces, schemas and vocabulary to foster cross-border and cross-disciplinary*

	research engagement. This includes the sharing of data, publications, software and services across the EOSC federation.
User-centric, data-driven, agile	*EOSC is a de-facto data-driven environment and the particular EOSC EU Node is designed to be user-centric as much as possible providing superb user experience for all scientific research workflows. EOSC is in the heart of the EU Data Strategy and identifies as the European Common Data Space for Research and Innovation.*

3.7 Constraints

[redacted]

The major constraints of the EOSC EU Node deployment lays in the fact that it must be an open platform primarily targeting the scientific research community at large, governed by the EOSC Tripartite Governance including the Members States and Associated Countries, the EOSC Association and the EC. The platform must fulfil the expectations of the joint Strategic Research and Innovation Agenda (SRIA) of EOSC and serve the main purpose of multi-disciplinary and multi-national research workflows of ERA.

At the same time, the EOSC EU Node is also a Commission service. It must comply with the IT Governance processes of the Commission and fulfil all operational, maintenance and support criteria of a managed service platform in production as a public service. Therefore, we have a given constrain on budget, scope, legal/admin compliances, and human resources involved defined by the manged service contracts.

The EOSC EU Node services have been procured and being owned by the Commission which implies that the EOSC Web Front-Office (landing website) will be hosted under ec.europa.eu domain. The domain request (open-science-cloud.ec.europa.eu) and its approval process is pending on the ITCB approval of this Project Charter.

The requirement to use free open-source software (FOSS), wherever it is possible, in order to maintain a choice of implementation options that can serve as a blueprint for other potential EOSC Node candidates in the community to follow is also a technological constraint. If licensed products are used, open API definitions are mandatory. This leads to the constrains of the EOSC Interoperability Framework that must be established and maintained during the lifetime of the service provisioning of the EOSC EU Node.

Concerning data confidentiality, security, and data protection, the EOSC EU Node predominantly handles open data in its main business processes serving the scientific research community and mainstreaming open science principles. The main purpose of the platform is to act as a node for open, easy to access, scientific FAIR data sets and interoperable services (i.e., no SNC data). The service management system and operational layer of the EOSC EU Node may handle sensitive information: provider contacts, user behaviour, user details etc. We are currently in the process of evaluating data sensitivity (Data Protection Impact Assessment) together with the DPO of the Commission and the contractors, creating data records and establishing data classification in DPMS.

3.8 Assumptions

The underlying key assumption of the EOSC Federation is that there will be multiple interoperable EOSC Nodes established and operated at the national, regional, and thematic communities level, of which the EOSC EU Node is only one of them. The EOSC EU Node therefore also serves as a reference node or blueprint architecture/deployment concepts that paves the way for further deployments of other EOSC Nodes in the near future.

[redacted]

- [redacted]

[redacted]

- [redacted]
- [redacted]

[redacted]

- [redacted]
- [redacted]
- [redacted]

[redacted]

- [redacted]
- [redacted]

3.9 Risks

[redacted]

ID	Risk Description & Details	Status	Likelihood[1]	Impact[2]	Risk Level[3]	Risk Owner	Risk Response Strategy[4]	Action Details
1	The Contractors of the three separate service lots (EOSC Core, Infrastructure and	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]	[redacted]

[1] A numeric value denoting the relative probability that the risk should occur. The possible values are: 5=Very high, 4=High, 3=Medium, 2=Low, 1=Very low.

[2] A numeric value denoting the relative severity of the impact of the risk if it should occur. The possible values are: 5=Very high, 4=High, 3=Medium, 2=Low, 1=Very low.

[3] The risk level is the product of the likelihood and impact (RL=L*I).

[4] The possible risk response strategies are: Avoid / Accept / Reduce / Transfer for negative risks (threats) and Exploit / Accept / Enhance / Share for positive risks (opportunities).

ID	Risk Description & Details	Status	Likelihood[1]	Impact[2]	Risk Level[3]	Risk Owner	Risk Response Strategy[4]	Action Details
	Applications) fail to coordinate							
2	Initial services deployment and integration timeline is unrealistic							
3	User policies and Rules of Participation for services cannot be supported by implementation choices							
4	Production launch of the EOSC EU Node services is delayed							
5	Necessary compliances of security, data protection, IT operations, quality assurance, web guidelines, etc. fail to pass							
6	Resource management of the platform services does not meet the usage demand							

ID	Risk Description & Details	Status	Likelihood[1]	Impact[2]	Risk Level[3]	Risk Owner	Risk Response Strategy[4]	Action Details
7	Verification and validations of the services delivered fails or shows major issues							
8	Production operation, maintenance and support of managed services does not meet the SLAs defined							
9	Handover/Takeover requirements of the managed services cannot be fulfilled							
10	User feedback on additional features requirements or changes							

4 COST, TIMING AND RESOURCES

4.1 Cost

4.2 Timing and Milestones

The milestones and timing below have been defined in the final tender specifications and accepted by all Contractors. More granular service development and delivery timeline will be made available by the Contractors by the end of Month 1.

ID	Milestone Description	Target Delivery Date
Phase-in Period	Contract signature and kick-off meeting maximum 15 workdays after the contract signature. Review the planning and support documents (marked with *) being ready for the preliminary approvals of the Contracting Authority once the service period starts. Start preparing the Project Management Plan with detailed timeline, activities, tasks, milestones and deliverables (according to PM2) Build up capabilities to the level that is needed for the defined SLRs in the contract from Day 1 of the service delivery.	***Signature: 17 November 2023*** ***Kick-off meeting: 5 December 2023*** ***Start of delivery: 14 February 2024***
M1	1st Monthly Meeting Timeline for the delivery of the required planning and support documents should be finalized taking into account the approval processes of the Contracting Authority. First versions of the individual service components and functionalities should be made available to the Contracting Authority for testing/reviewing purposes Detailed information on final system architecture, integration/harmonization and configuration actions as well as service coordination and support activities in production should be delivered. Integration plan of Lot 2 EOSC Exchange Infrastructure Services and Lot 3 Exchange Application Services into the Lot 1 EOSC Core Services must be orchestrated. JIRA ticketing system and other reporting procedures must be up and running. Initial web presence must be established and provide updates on the deployment status.	**14 March 2024**
M2	2nd Monthly Meeting Project Management Plan with timeline, activities, tasks, milestones and deliverables (according to PM2) should be finalized and approved. Provide final architecture diagram, schematic, risk registry and change management plan to all the system/service components. Seek approvals on all IT Governance and Data Protection related deliverables from the Contracting Authority, as deemed necessary. Demonstrate all ITSM functions and processes ready for production roll-out.	**14 April 2024**

ID	Milestone Description	Target Delivery Date
M3	1st Steering Committee Meeting/combined 3rd Monthly Meeting All required planning and design documents submitted for sign off and becoming living documents to be followed up on and maintained throughout. Provide detailed documentation of the system/service components and functionalities, including details of contingency, back-up, performance, monitoring, training materials and support. Support the Contracting Authority in announcing and introducing the EOSC Node services to the early community users at events and conferences. Checking the delivery and reporting of the services by the Contracting Authority, verify and authorize payment request based on invoicing. Continue with the roll out of all system/service functionalities and the corresponding coordination/support capabilities until M6.	**14 May 2024**
…	Monthly activity report with updated documentation & work plan for next month	
M6	2nd Steering Committee meeting/combined 6th Monthly Meeting The Lot 1 EOSC Core Services are up and running in production managed, operated, maintained, coordinated and supported 24/7. Lot 2 EOSC Exchange Infrastructure and Lot 3 EOSC Exchange Application Services are on-boarded and integrated with the Lot 1 EOSC Core Services functionalities. On-boarding, verification, validation and testing, policy compliances and other process are established, documented (see EOSC Interoperability Framework update) and tested. Third-party EOSC services providers are being identified and gradually on-boarded to the Lot 1 EOSC Core Services business as usual. Planning for additional (above the baseline) Project-based activities are presented and approved by the Contracting Authority. Checking the delivery and reporting of the services by the Contracting Authority, verify and authorize payment request based on invoicing.	**14 August 2024**
…	Monthly activity report with updated documentation & work plan for next month Quarterly Service Committee Meetings conducted regularity and actions followed up	
M30	10th Steering Committee Meeting/combined 30th Monthly Meeting Prepare draft documentation outlining development and deployment work, methodology and provision for handover and work plan for the last 6 months. Prepare the necessary handover procedures with the Contracting Authority. Checking the delivery and reporting of the services by the Contracting Authority, verify and authorize payment request based on invoicing.	**14 August 2026**

ID	Milestone Description	Target Delivery Date
…	Monthly activity report with updated documentation & work plan for next month	
M34	34th Monthly Meeting Changes to draft final report based on input from the Contracting Authority, support and maintenance, customisations as required for ensuring seamless operations during the handover phase.	**14 December 2026**
M35	35th Monthly Meeting Finalise report and provisions for handover, support and maintenance, customisations as required. Initiate the handover process gradually as planned.	**14 January 2027**
M36	12th Steering Committee Meeting/combined 36th Monthly Meeting Final report meeting with provisions for handover and final documentation. Final services review and acceptance Checking the delivery and reporting of the services by the Contracting Authority, verify and authorize payment request based on invoicing. Closing financial management, last invoicing and payment. Agree on the post-transfer services and support to be maintained according to the contract.	**14 February 2027**
Post-transfer Period	At the request of the Contracting Authority the contractor must provide all necessary assistance, including information, documents and files, to allow the Contracting Authority to complete, continue or transfer the services to a new contractor or internally, without interruption or adverse effect on the quality or continuity of the services	**Defined by legal terms**

4.3 Planned Resources

Under the public procurement action, the Commission procured fully managed services provided by the third-party Contractors for a fixed price (all inclusive). The Contractors are required to manage their own resources in order to deliver the managed services according to technical specifications and SLR/SLAs defined in the final tender specifications and annexed to the services contracts.

[redacted]

[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

5 APPROACH

5.1 Methodology

The EOSC EU Noce Contractors' consortium is internally organised and governed with the following instruments, roles, and responsibilities, applying the PM²-Agile methodology and associated terms/roles, which is seamlessly integrated and extends PM². Any underlined roles refer to the administration of the governance processes of the Contract as prescribed by the CA (Req. 1.9.1/2) as per the PM² methodology.

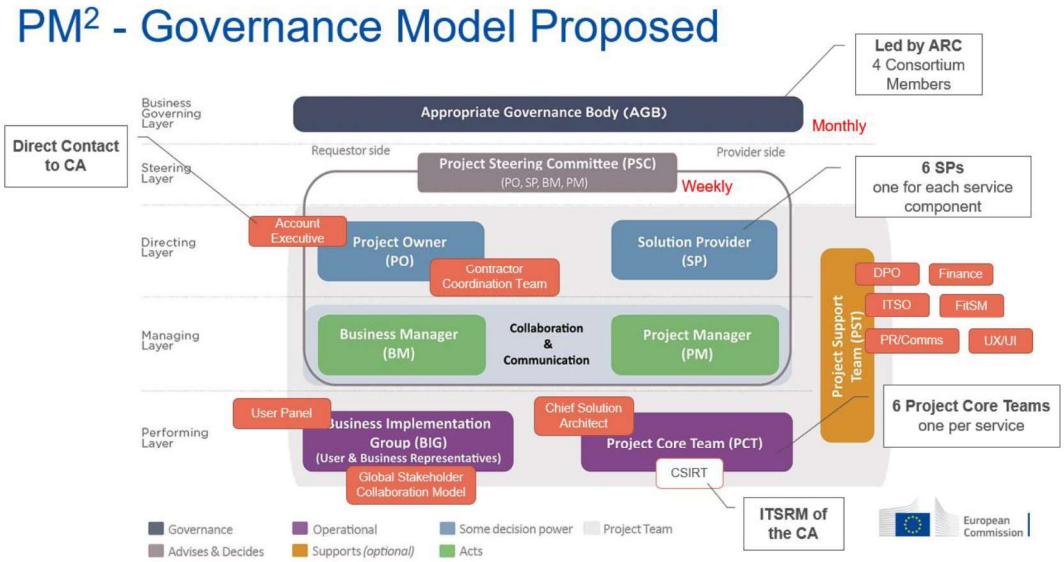


Fig. 3. Proposed methodology for the project management and governance model

The participating personnel are members of the Consortium's Project Office, Technology Office, Legal/Financial Unit, Communications/Marketing Unit:

- Appropriate Governance Body (AGB). This is the governance body of the Consortium, operating under the scope of our Consortium Agreement, with one representative per partner (4) and led by the Group Leader. It is responsible for high-level decisions of the project (resource/financial management/reallocation, risk mitigation, strategic planning, etc.) with direct communication lines with the Project Owner and Project Steering Committee (PSC). Its members will be the Senior Executives participating in the Service Committee. AGP convenes monthly (and ad hoc for urgent matters) inviting other team members as appropriate; its decisions are conveyed and implemented by the Project Owner and PSC members across their teams.
- Project Steering Committee (PSC). It will comprise one Project Owner (PO; key project decision maker across all managed services, accountable for project success) which will assume the role of the Account Executive, 6 Solutions Providers (SPs; one for every managed service) which will assume the roles of the six Service Delivery Managers, one Lead Project Manager and four Deputy Project/Business Managers (one per Consortium partner), which will assume the roles of the Project Managers. The PSC convenes on a weekly basis, supported by the Project Support Team (PST) (see next) inviting A-PCT members as appropriate.
- Six Project Core Teams (A-PCT; one per managed service), comprising a TeCo, Product Owner, Architecture Owner, and ATeMs (OME, WD, DC, ITSO, DPO, CRM, PCO). Members of the A-PCT may serve in more than one team and under different roles. The Architecture Owners form a separate subgroup (Arch-Team) under the Chief Solution/System Architect to ensure uniform and effective solution architecting at the project level.
- BIG (Business Implementation Group), comprising the Service Delivery Managers, the Product Owners and the Architecture Owner. A-PCTs perform all formal PM²-Agile Ceremonies and maintain all Artefacts.
- Project Support Team. These comprise Consortium personnel providing project-supporting activities to the directing, managing, and performing organisation layers via the following sub-teams: (i) Legal/financial, which includes the DPO and admin/financial personnel of the Consortium, (ii) Comms/Media, assembling all PCOs and led by the PR/Communications

Manager, responsible for organising, supporting and performing all comms and marketing actions, including copywriting, graphic design, video production, etc., (iii) UI/UX, assembling WDs under our UX/UI Lead, responsible for designing, reviewing, enforcing, and accepting a commercial-grade and streamlined user experience across all EOSC node touchpoints, (iv) CSIRT, assembling all ITSOs and led by the Chief ITSO, (v), Training and Support, assembling all CRM under the Support Lead, responsible for the end-to-end effective and high-quality support and training activities across all EOSC node services, (v) FitSM, responsible for monitoring, enforcing, preparing required reports/SLRs, and auditing, across all teams the FitSM framework and practices applied in the SMS internally and externally (Lot 2/3, Exchange services). Members of all sub-teams are embedded (permanently or on an ad hoc basis) across the APCTs to increase collaboration and agility.

5.2 Change Management

The main task of the Change Management (CHM) in the context of the EOSC EU Node managed services is to ensure that changes to the Services Configuration Items (CIs) are planned, approved, implemented and reviewed in a controlled manner avoiding adverse effects to services or customers. Additionally, it provides rules on how to handle Releases, so that these new changes can be tested and deployed to the live environment together. It also oversees (approves, reviews, etc) their implementation in the same manner as single changes.

5.2.1 Project Change

Project Team and Collaboration will be implemented applying the formal methodology (PM^2- Agile), collaboration instruments and processes. In this context, the Project will be approached as an EPIC, involving all A-PCTs and hence services.

Project Inception. The following activities will be conducted at the project's inception:

- Requirements. Formal requirements for the project will detail all required technical and procedural development, risk assessment, client relationship management (e.g. the target research infrastructure in onboarding project A), communication/training/outreach, testing, integration, and deployment effort. These will be transferred as issues in the relevant service boards (under one EPIC). This activity will be led by the Chief Solution/System Architect and the Service Delivery Managers, and its output will be approved by the PSC.
- Issue Formalization. The list of requirements expressed as issues in the EPIC will be reviewed against the existing backlog so that dependencies and resource gaps can be identified. This activity will be performed by the BIG, approved by the PSC and submitted for approval to the CA (in the context of the Monthly Meetings or as otherwise requested by the EC). If functionality for a special project will be delayed because of other priorities, the timing and phasing of the special onboarding project may need to be adjusted to accommodate this.
- UI/UX mockups. Impacts on user experience will be identified in advance in order to assure clarity for users and understanding by stakeholders such as the CA.
- Infrastructure allocation. The need for additional resources will be assessed in advance, although this is not likely given the scope of these projects.
- Lot 2/3 Contractor collaboration. Specific rule changes that impact the presentation and operation of Lot 2/3 Contractor services will be assessed and communicated in advance through relevant coordination bodies.
- Time-planning and control points. We will prepare a detailed plan of activities, clearly marking the EPIC sprints, milestones, and acceptance criteria (per issue, sprint, and EPIC).
- Sizing and Time Planning. Finally, and based on the previous output, we will review and revise the sizing of effort required to implement the project, detailing the involvement of Consortium personnel, while also considering their availability.

Project Implementation. The actual project implementation commences, with activity balanced among relationship management and communications, data analysis and migration, procedural reviews and updates, as well as the iterative development, testing, and acceptance of required technical changes. As per our methodology, development is in self-contained sprints, leveraging all collaboration tools, CI/CD and QA pipelines.

Project Acceptance. The progress of each special onboarding project will be communicated to the CA on a weekly basis, highlighting any deviations from estimated effort and time planning. For Onboarding Project B, the CA will be specifically asked to approve advanced communications to Providers about requested changes to ensure full transparency and clarity, as well as final "go live" steps before new rules are implemented. The Consortium will inform the CA as key milestones in the plan are reached and when the project overall has been successfully completed, along with all related material and integration testing results.

5.2.2 Configuration Management

Service quality verification within the EOSC EU Node will be ensured by rigorous procedures defined within the SMS according to the FitSM/ITILv4 standards. The Service Portfolio Management (SPM) process defines what is required for delivering production-grade services and enforces that all requirements are satisfied before a state transition (e.g., from pre-production to production). Changes to every service are required to be controlled via the Change Management (CHM) process, which ensures adequate testing prior to rollout of changes. After changes are validated, new releases can be deployed in production according to the procedures of the Release and Deployment Management (RDM) process. To ensure stability of the system and its functionalities and to automate the validation of the services against the multiple quality criteria defined in the EOSC SMS and its processes, a rigorous testing infrastructure will be built. In the proposed approach, the testing is performed at three different levels, implementing a hierarchical testing architecture (see next figure).

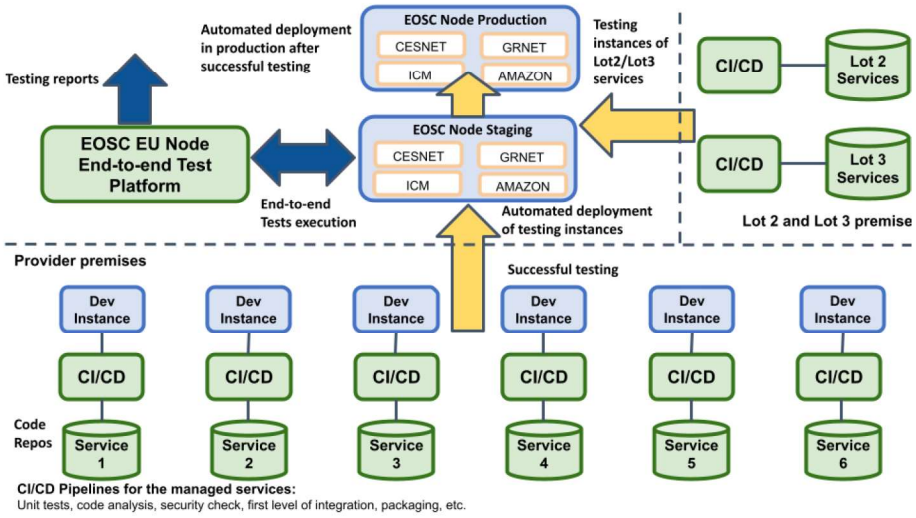


Fig. 4. Configuration management and CD/CI pipelines

The bottom part of the diagram represents the internal release cycles of the six managed services, together with all their components. For each of the services, the service provider employs an automated CI/CD pipeline to locally deploy new releases of the service and/or its components and validate that it passes the quality criteria of the service provider, including unit, integration (between service components), and system tests, any additional validations the provider might use for the service (performance, stability, various security and penetration tests, etc.), as well as packaging for release. The use of CI/CD pipelines automate the series of steps that must be performed to deliver new versions of service components, in a way that is aligned with the quality process defined in the EOSC SMS, minimising human error, and maintaining a consistent software release process. Each product team owns its own CI/CD pipelines and configures them to properly adhere to the rules defined in the EOSC SMS.

5.2.3 Organisational Change

The tender specified that the basic principles of the governance structure, roles and responsibilities must be in line with the PM² (Project Management Methodology) that is a generic project management methodology developed by the European Commission and open to all. Its purpose is to enable project managers deliver solutions and benefits to their organisations by effectively managing the entire lifecycle of their projects. PM² incorporates elements from a wide range of globally accepted project management best-practices.

Resource Management Services include the activities associated with the provision and adjustment of appropriate human resources, according to workloads, expertise, continuity and cost optimization

requirements, to perform the required Services at the required Service Level Requirements for the agreed price. They include but are not limited to the following tasks:

- Continuously monitor the performance of all the human resources involved in delivering EOSC Core Services to ensure that the Services comply with the SLRs
- Analyse the impact of optional requests made by the Contracting Authority and to be implemented by the Contractor and propose solution
- Monitor the workload of the various Contractor (and sub-contractor) human resources and adjust as needed to meet SLRs
- Ensure that staffing and skill levels are adequate to achieve contract objectives
- Ensure reliability of personnel and inform them of their non-disclosure agreement (NDA) obligations with the CA
- Manage the Contractor Staff time off and replacement
- Designate certain members of Contractor Staff as key team members ("Key Personnel"), as listed in Requirement 1.9.2
- Inform the Contracting Authority of any potential key personnel staffing changes and of any new personnel assignments planned for new projects and Services
- Review and authorize key changes to personnel for existing services and personnel for new projects and services
- Assign a new Contractor Account Executive and/or Contractor Service Delivery Manager(s) upon the Contracting Authority request

6 GOVERNANCE AND STAKEHOLDERS

6.1 Structure

The current tripartite governance structure of EOSC is depicted in the diagram below. It's been in place since 2021. The service/business owner of the EOSC EU Node is the European Commission, represented by DG CNECT Unit C.1. The Service Owner reports to the EOSC Tripartite Governance, that includes the representatives of the Members States and Associated Countries as well as the representatives of the EOSC Association members, approx. 4 times a year.

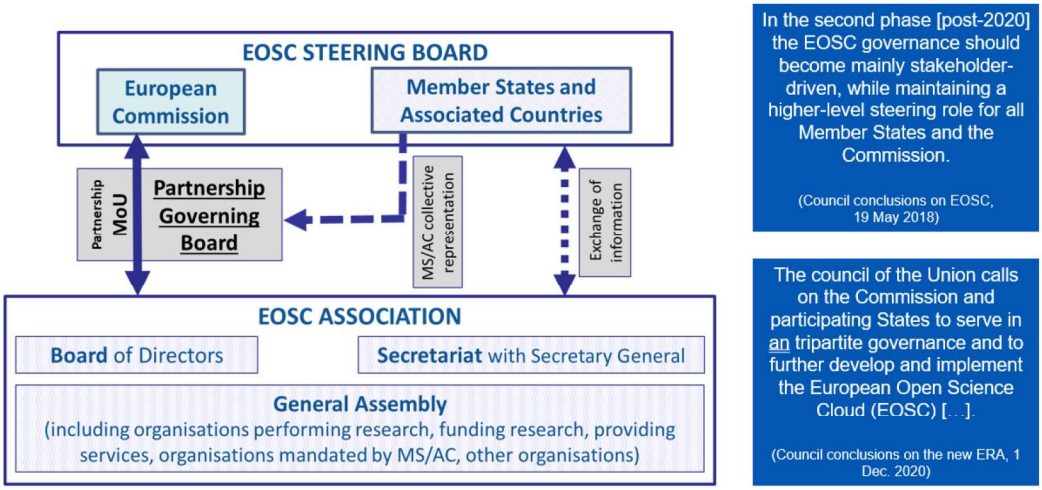


Fig. 5. EOSC Tripartite Governance structure and reporting lines

6.2 Roles and Responsibilities

[redacted]

[redacted]

6.3 Other Stakeholders

EOSC is a concept of an important core scientific infrastructure for European science that will need to lead the future development of infrastructures far into the future. From a legal point of view, the EOSC vision could benefit from the involved stakeholders defining a legal entity to take forward a number of practical matters from handling finances to hiring dedicated staff etc. This could be facilitated for instance by the establishment of an EOSC PPP or a similar vehicle.

On an operational basis, the governance of EOSC should enable the MVE process. Furthermore, the proposed structure needs to be capable of running a fully-fledged EOSC. As mentioned above, the users have to play a central role in the design and implementation of the EOSC. Requirements, standards, operating procedures, etc., should be defined through close collaboration between:

- the end user - scientific community;
- the service providers – developers, intermediaries and operators;
- funding agencies and scientific policymakers.

Rapid realisation of an EOSC governance is needed to move from vision to implementation. The two-stage approach for the implementation of EOSC has received general support, the first stage being the process of developing the EOSC and the second stage its management, operation and development.

In full conformance with what presides and the SWD, a three-layer governance model, based on the EOSC Declaration is proposed, as depicted in the following Figure. The three layers are:

- Strategic Layer in the form of an EOSC board to combine state-of-the-art expertise on scientific cloud infrastructures with the Funders and Policy Makers. The Board will therefore include EU Member States and Associated Countries representatives. The EOSC board will mainly make strategic decisions on the development and evolution of the EOSC.
- Executive Layer in the form of an executive board to manage day-to-day operation of the EOSC and procurers designing and planning work-related future developments. This, the only full-time staffed layer, will be supported by Working Groups, and will have the responsibility of ensuring that user needs are met and strategic requirements addressed.
- Stakeholder layer organised in the form of a stakeholders forum to provide a medium for stakeholders: Users (Consumers), Providers and Intermediaries of EOSC Resources. This would have the main role to discuss, supervise and channel communication between the EOSC and the communities across all three layers.

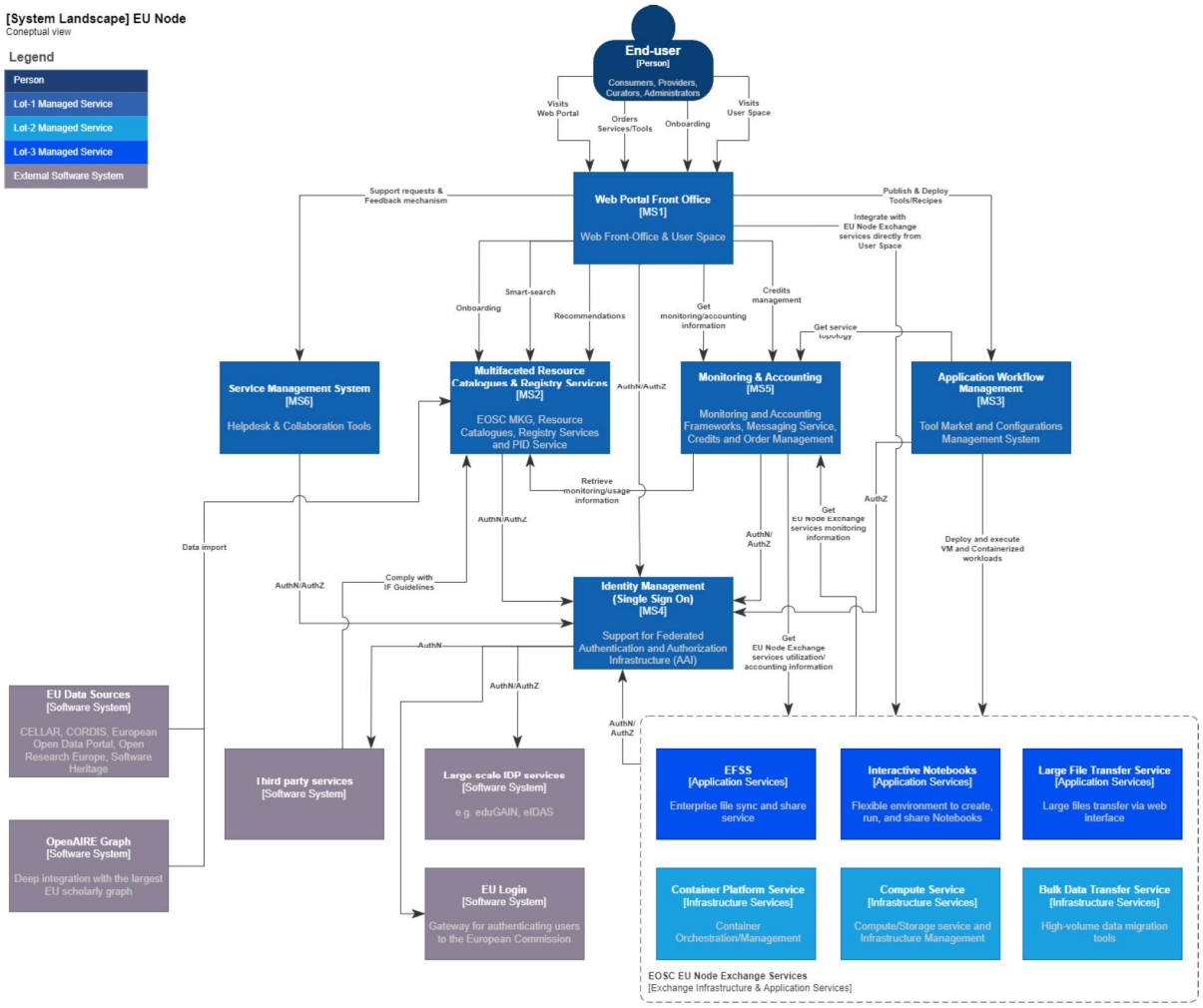
The EOSC Secretariat is a Coordination Support Action Structure, providing support to the EOSC Governance while working openly and inclusively with communities to co-create an all-encompassing European Open Science Cloud.

APPENDIX 1: REFERENCES AND RELATED DOCUMENTS

[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]

APPENDIX 2: SYSTEM ELEMENTS – ARCHITECTURE DETAILS

1 HIGH LEVEL ARCHITECTURE DIAGRAM



2 COMPONENTS DESCRIPTION

2.1 LOT1 - Core Federating Services

2.1.1 Web Portal Front Office

Description

The Web Portal Front Office is the main European Open Science Cloud (EOSC) EU Node platform's entry point, framing, organizing, communicating, and recommending access to a plethora of content, resources, and services. It seamlessly integrates all the architectural components of the EOSC EU Node platform around a cohesive front-end delivering comprehensive, intuitive, and low-effort approach towards framing, exposing, and delivering EOSC EU Node's resources, services, and training material in a manner that serves the scientific workflows and open science practices of all scientists, researchers, and innovators, regardless of their domain, seniority, knowhow, and location. On the one hand it enables the effective discovery and utilization of cross-disciplinary research output, and on the other, provides streamlined access to commoditized cloud storage and compute services to facilitate research within a collaborative environment. It provides a unique place for service and data providers, as well, to advertise their offerings to the community and brings together National/Thematic Science clusters and EOSC Node candidates. Finally, it serves as the first Node of the EOSC federation, also acting as a 'blueprint' for other national/regional/thematic nodes, gradually implementing and supporting EU's Policy of Open Science and European Data Strategy.

The EOSC EU Node Web Front Office (website) will be hosted under the *ec.europa.eu* web domain following the Open Europa web/style guides and other recommendations by DG CNECT D.4 and DG DIGIT.

Component Details

The Web Portal Front Office full-stack managed service deployment instantiates a micro-services architecture integrating several software applications, components, external services, and back-office APIs, deployed over a Kubernetes container orchestration framework. As such, the managed service has excellent scaling, performance, cost-effectiveness, and flexibility characteristics, employing best practices in large-scale web applications. The use of virtualization, containerization, and cloud-based hosting infrastructure also enables efficient resource utilisation and management while implementing robust network security measures to protect the service and its data. The inherent scalability features allow to efficiently handle increases in demand which is critical for the component's performance given that it is the main integration point of all the architectural components of the EOSC EU Node platform, directly handling end-user traffic and forwarding/consuming traffic from/to all other EOSC EU Node's components. The entire Web Portal Front Office comprises two logical software components, the Web Portal and the User Space. The Web Portal serves the static content of the platform and provides access to all the other Front Office services. It also provides access to training resources/courses/tutorials and functions as the entry point for all users (registered and unregistered) to search, discover, and view detailed information (resource-specific views) for all EOSC EU Node's assets (resources, data, services, Exchange services, tools, etc.) as exposed by the "Multifaceted Resource Catalogues and Registry Services" component detailed in the Architecture Canvas. The User Space comprises a collection of sub-components exposing the underlying back-office EOSC EU Node platform's applications and services via a coherent and intuitive interface, integrated through their corresponding APIs. The content and structure of the User Space (both its landing Dashboard and individual sub-components) are dynamically adapted based on the specific role/capacity the user has. As such, it supports the needs of all EOSC EU Node platform's users within the same stable and intuitive environment, while securely exposing the required back-office functionalities per user role/state.

2.1.2 Multifaceted Resource Catalogues and Registry Services

Description

The Multifaceted Resource Catalogues and Registry Services component streamlines search and access to a wide variety of high-quality, cross-disciplinary scientific content by aggregating in the platform's Metadata Knowledge Graph highly esteemed EU catalogues and repositories, the EU's open scholarly communication graph (OpenAIRE), along with content directly contributed in the EOSC platform in a unique value proposition. The component collects and interlinks metadata records of research products (including publications, data, and software/code), services, deployment tools and recipes (provided through the Tools Market of the Application Workflow Management component) with authors, communities, organisations, services, funders, and projects. The quality of the provided information is guaranteed by applying dedicated and extensive curation and quality assurance processes throughout the lifecycle of the Metadata Knowledge Graph's production. The component is deeply integrated with the Web Portal Front Office component through which it allows the registration of organisations (EOSC Providers) that aim to publish their offerings (services, data, software, publications) via the EOSC platform. It also provides dedicated APIs to enable the creation of added value services.

Component Details

The Multifaceted Resource Catalogue and Registry Services is provided as a full-stack managed service deployment providing to the end-users a uniquely positioned Metadata Knowledge Graph (MKG) which is built on an architecture that features the following sub-systems: metadata aggregation (harvests research products metadata records from OpenAIRE Graph, CELLAR, CORDIS, European Data Portal, Open Research Europe, Software Heritage; harvests EOSC EU Node Service Catalogue for service profiles, IF registry for IF guidelines, Tool Market for recipes, projects, and bundles), metadata harmonisation (transforms metadata records into the MKG data model), metadata deduplication by PID (merges records with the same PIDs, e.g. DOI), Provider User Space (specialised section of the Web Portal Front Office that supports Providers for onboarding research products from their Data Sources, checking status of harvesting, validating (and feedback) compliance and FAIRness), and metadata curation environment (enables data curators to perform adaptive sampling of MKG to investigate and identify issues; reported to original sources, authors, or internally fixed) with very high quality assurance requirements. The MKG supports deduplication at the PID level, to improve performance and enable fine tuning of the process (graph deduplication and optimization techniques). The component also delivers as-a-Service, customized views of the MKG, to better serve communities and organisations with dedicated discovery and research assessment capabilities. The component is deployed over a Kubernetes container orchestration framework and as a result it has excellent scaling, performance, cost-effectiveness, and flexibility characteristics, employing best practices in large-scale web applications and information systems to effectively address the baseline sizing and scaling requirements of the EOSC EU Node.

2.1.3 Application Workflow Management

Description

The Application Workflow Management component enables the EOSC EU Node end-users to compose and execute complex tools (applications and application workflows) across the EU Node Exchange infrastructure services. The added value of this component is two-fold. On the one hand, the involved tools (in the form of user-contributed 'recipes', such as TOSCA templates and Ansible scripts), provide ready-to-use solutions that streamline the execution of complex infrastructure provision and configuration tasks. This functionality greatly facilitates research activities by simplifying time-consuming and cumbersome tasks. On the other hand, it simplifies access for end users to the underlying Exchange infrastructure services of the EOSC EU Node (Virtual Machines, container orchestration framework) , thus ensuring their effective utilization. These capabilities greatly benefit researchers as they enable them to focus on their activities and not spend time and effort on potentially complex IT tasks. The recipes can be created, edited, and shared by users (researchers) of the EOSC EU Node, and are findable through the Multifaceted Resource Catalogues and Registry Services component. This allows end-users to search for recipes based on their supplier, components used, referred publication or compatibility with datasets. All recipes in the catalogue are validated before being published to ensure they are deployable, executable and security-compliant with all the components of the EOSC EU Node. The Application Workflow Management is integrated with the Web Portal Front Office so that the users can manage their tools lifecycle by leveraging all the front-end capabilities.

Component Details

The Application Workflow Management component is provided as a full-stack managed service deployment comprising different sub-components. These are the Tool Market - Access Layer, the

Infrastructure Manager (IM) – Application Manager Layer, and the Configuration Management System. The Tool Market delivers all the front-end capabilities to allow the end-users to navigate over the Multifaceted Resource Catalogues and Registry Services component, identify the tools of interest and deploy them by leveraging the Exchange infrastructure services of the EOSC EU Node. The Tool Market publishes information about the usage of tools in the Monitoring and Accounting component of the canvas. The IM will be integrated with the Resource Management and Provisioning Layer provided by the EOSC EU Node's Infrastructure services providers. IM enables the deployment of containerised applications on the EOSC EU Node Exchange infrastructure services (or potentially on any other infrastructure provider offering their services via the EOSC EU Node). Containerised applications are described through TOSCA templates and can be deployed by IM in multiple cloud infrastructures. This is possible thanks to the large and easy extensible set of interfaces and plugins that IM supports (public cloud providers such as Amazon EC2, Google Cloud Platform, Microsoft Azure, T-Systems OTC, Orange flexible engine, on-premises cloud management platforms such as OpenNebula, OpenStack and federated environments such as the EGI Cloud Compute). The Configuration Management System GOCDB is a central information repository providing a web portal interface for CRUD operations and a REST API for data queries. It is a key tool for the configuration management of the EOSC EU Node components and is intentionally designed to have no dependencies on other operational tools other than the Identity Management (Single Sign On) component. The IM depends on the Configurations Management System to retrieve topology information for the underlying EOSC EU Node infrastructure services. The Application Workflow Management component is deployed over a Kubernetes container orchestration framework and as a result it has excellent scaling, performance, cost-effectiveness, and flexibility characteristics, employing best practices in large-scale web applications and information systems to effectively address the baseline sizing and scaling requirements of the EOSC EU Node.

2.1.4 Identity Management (Single Sign On)

Description

The Identity Management component of EOSC EU Node platform provides a convenient and secure way for users to access multiple systems and services within the platform's domain using a single set of credentials (Single-Sign-On, SSO), while ensuring the privacy and security of user data. It also simplifies the management of user accounts, reduces the risk of password fatigue, and enables organizations to collaborate and share resources more easily. The component also provides federated Authentication and Authorization Infrastructure (AAI) that enables users to authenticate with their institutional identity provider and authorize access to resources across multiple providers' domains that participate in the EOSC Access Federation.

The Federated Identity Management Solution provided by the EOSC EU Node will be integrated with EU Login step-up authentication solution of DG DIGIT.

Component Details

The component provides a full-stack managed service for Identity Management (Single Sign On) and federated AAI based on the AARC Blueprint Architecture. The solution is based on the GEANT Core AAI Platform and RCIAM that are already used to deliver Identity Management services to ESFRI Clusters (Lifesciences, the Photon and Neutron RIs and SSHOC), the FENIX RI, the Puhuri RI, to EuroHPC Systems (currently LUMI and Leonardo), etc. The above demonstrate the component's mature operational capabilities (performant, well-tested, serving already thousands of users), highlight that it leverages existing community best practices in AAI and supports federation with existing large-scale IDP services. The component supports authentication via Identity Providers from national federations in eduGAIN, national eIDs via the eIDAS Bridge, x.509v3 certificates via the IGTF Bridge and common social media authentication methods. The service is technically ready to integrate with EU Login using the OpenID Connect protocol and will be able to integrate IDPs from outside the research and education community (Guest IDPs) that support any of the federated authentication protocols (OAuth2, OpenID Connect and SAML2).

2.1.5 Monitoring and Accounting

Description

The Monitoring and Accounting component provides respective frameworks that integrate with the other components presented in the architecture canvas and used to track and analyse IT operations. The service

availability and reliability monitoring (EOSC EU Node Monitoring) is a framework that provides insights into an infrastructure, the applications, services, and even into processes/behaviours. It provides a comprehensive view of the infrastructure and applications running on the EOSC EU Node, allowing admin teams and users to quickly identify and troubleshoot issues before they impact the service levels and performances. The Monitoring framework also integrates with collaboration platforms and configuration management system that the admin teams use to facilitate their activities and has inherent scalability and flexibility capabilities. The component is also able to monitor the conformance of multiple SLAs. The Accounting framework tracks and records the usage of services and research products both within the core components of the EOSC EU Node and from the externally contributed resources listed in the Multifaceted Resource Catalogues and Registry Services component. The Accounting framework ensures that the various resource providers are delivering the committed resources and enables the optimal allocation of resources to the platform's end-users through a virtual credits mechanism which can be used to regulate access and usage of the EOSC EU Node Exchange infrastructure services and Exchange application services.

Component Details

The Monitoring and Accounting component is provided as a full-stack managed service. Monitoring is the key service needed to gain insights into a distributed infrastructure. It needs to be continuous and on-demand to quickly detect, correlate, and analyse data for a fast reaction to anomalous behaviour. The key functionalities offered by the Monitoring framework are: (i) Monitoring EOSC EU Nodes components, Exchange infrastructure services, Exchange application services, and all other Exchange services (ii) Reporting availability and reliability, (iii) Visualisation of services status, (iv) dashboard interface, (v) real-time alerts to services providers and EOSC EU Node service operators to varying levels of complexity. Further, APIs are provided to allow third parties to gather monitoring data from the system. The Accounting Framework is designed to efficiently collect, aggregate, and exchange usage metrics across various infrastructures, providers, and projects. The system provides a REST API, which accepts input from diverse resources, stores it in a database, and aggregates the incoming data. It also offers an intuitive user interface that allows clients to interact with the platform and access accounting data for specific time periods. All API resources are only accessible to authenticated clients, ensuring secure access to sensitive data. The key functionalities offered by the Accounting Service are: (i) Efficient collection, aggregation, and exchange of metrics, virtual credits and service limits, (ii) REST API that accepts input from diverse resources, (iii) Database storage and aggregation of incoming data, (iv) Intuitive user interface for accessing accounting data, (v) Secure access to sensitive data through authenticated clients. From an architectural aspect, the Monitoring and Accounting component has the main characteristics of scalability, availability, and latency. It is designed to be highly reliable and scalable. It provides very high durability and redundancy as all of its underlying components will be deployed in a high availability mode to minimise single points of failure. The deployment architecture supports horizontal scaling by provisioning more nodes, if required to increase service capacity. Nodes for each component can be easily added or removed on demand without interrupting the overall availability of the service.

2.1.6 Service Management System

Description

The Service Management System (SMS) component services and operational roles in delivering the quality-assured provision of EOSC EU Node. It includes a set of roles, responsibilities, procedures, policies, and other documentation and tooling to support management of services according to the FitSM family of standards, that is also suited for federated scenarios making it ideal for use in the EOSC EU Node environment. The SMS will plan, monitor, coordinate, and report the provision of the offered EOSC EU Node components and services under the direction of policies and organised by processes and supporting procedures. The SMS will help ensure that the contractual obligations of service delivery are fulfilled by controlling and supporting management of all EOSC EU Node components. This includes all aspects of service performance management, service provisioning, system integration, performance monitoring, support and coordination activities, as well as service quality across all components of the EOSC EU Node. As part of the SMS component the EOSC EU Node, a Helpdesk platform and technical support teams will be provided to manage service incidents and service requests. The Helpdesk includes all basic features of a helpdesk system needed to provide effective user support, such as a form to submit a ticket, ticket search, email notifications on change of the ticket status and user feedback mechanisms. The workflows and tickets triage will be organised according to the established procedures of the Incident Service Request Management (ISRM) process within the SMS. The Helpdesk support unit structure will be

organised in three levels and comprises multiple Support Units (SU) for effective resolution of the incidents and user requests within the EOSC EU Node.

Component Details

The SMS is based on FitSM family of standards to govern all aspects of service delivery. FitSM was developed to ensure a seamless user experience when integrated services are provided by multiple providers and is particularly suited to the federated nature of service delivery, not only within the EOSC EU Node, but also across the Exchange services and the multiple envisaged EOSC Nodes that will form the EOSC federation. FitSM (and similar SMS frameworks) are the ideal tools to ensure that end-to-end Service Level Requirements (SLRs), defined as part of Service Level Management, can be consistently achieved and customer experience is continually improved. FitSM is organized around a simple set of 14 ITSM processes that govern service management and their requirements, and application frame how the SMS responds to inputs from the (services, research products) provider and user communities, who can use the Helpdesk, as well as feedback mechanisms integrated directly into the Web Portal Front Office, to raise questions, complaints and offer suggestions for improvement. The Helpdesk, the Collaboration Tools (wiki, ticket tracker, workflow engine, mailing lists) underpinning the entire SMS, and the Monitoring and Accounting Systems will be linked to enable both consistent reporting, as automated as possible, as well as tools for visualisation, alerts, and exception reporting. Regarding the Helpdesk, dedicated personnel ensure continuous managed operations and administration. The Helpdesk, from an architectural perspective, features a high availability and performant installation that guarantees high quality of service for the EOSC EU Node end-users. It enables custom workflows, automation of procedures, an extended knowledge base, notifications, as well as feedback forms to continuously improve the end-user engagement. Withing the SMS, and to maintain a security environment within the EOSC EU Node, it is necessary that security is implemented as an ongoing process that covers all components and data. The process will be driven by assessment of security risks, allowing us to effectively target control measures. The standard IT Security and Risk Management (ITSRM) methodology will be employed, specifically designed for analyzing shared responsibility models, which is particularly pertinent given the distributed nature of EOSC EU Node.

2.2 LOT2 - Exchange Infrastructure Services

2.2.1 Container Platform Service

Description

Managed Container Platform Services provide containerized platform-level cloud services (PaaS) and resources for computing, data storage and access and network communication based on infrastructure-level cloud resources provided by Managed Compute (Virtual Machine) Infrastructure service.

The service is built using open source de facto standard components. Computing service is based on containerisation mechanisms Docker and CRI-O the as well as platform-level orchestration solution OKD that is an open version of RedHat Openshift. Data storage and access is provided by Ceph beneath the OpenStack platform (RBD volumes of VMs) and directly (Cinder for RBD, CephFS for file shares). OKD can also use NFS shares configured on the external NFS appliances or data storage clusters.

All services components are programmable and API driven which enables automation of deployment, configuration, provisioning (Infrastructure as a Code) and orchestration using Terraform, Ansible etc.

Set of tools for monitoring (Nagios, Zabbix) and control purposes as well as configuration management (GitHub) is used within the platform. Also auxiliary systems and tools are used in order to facilitate integration of the application workflows including Prometheus, Velero, Restic, Let's Encrypt, Harbor).

OKD provides the GUI for managing user projects and resources, configuration, monitoring and virtual machines access.

The computing, data and network services are run on top of the on-premise cloud infrastructure in the two hosting sites (PSNC, Safespring) of the contractors, interconnected by dedicated network links configured through NORDUnet.

The platform-level cloud services will be deployed in both data centers, and integrated in a loosely-coupled setup, i.e. active-backup high availability configuration. Basic instances of the services will be run at sites (user load will be distributed), and critical data and meta-data will be subject of local backup procedures and will be replicated to the other site, to enable recovery in case of failure.

Component Description

Business aspects:

Basic compute, data and network services of the containerized cloud platform are provided by the on-premise infrastructure at PSNC and Safespring – the contractors' hosting sites (data centers). This approach, accompanied by relevant security processes and procedures ensures that data sovereignty, data privacy and GDPR aspects are handled properly, according to best practices and EU regulations and recommendations. In particular no user data nor meta-data are stored outside EU. Also the actual network communication within the infrastructure is performed based on dedicated dark fiber links.

Economic scalability of the service is ensured by the fact that the basic service tiers (T-shirt sizes), are clearly defined along with transparent cost model. Data ingress/egress- or I/O-related fees are not applied (otherwise they could constitute big component of the cost in the I/O intensive application).

Costs are related to the compute resources, data storage and I/O handling capability and network resources offered as well as actual resources usage that can be monitored and controlled using functionality provided by the contractor. Service tiers can be flexibly adopted based on the EC request.

Usage of open source components for the cloud platform and infrastructure management, resources provisioning and orchestration (OKD, Docker, CRI-I, Linux) enables keeping unit cost relatively low, compared to e.g. systems where costly, yearly subscription-based models are currently promoted.

Functionality aspects:

The platform-level services support for API-based integration and self-service capability. They implement the concept of projects (OKD project) that group entities (containers, storage volumes etc.) and resources (computing, memory, storage) into pools dedicated to users or user groups.

Users may allocate the chunks of the resources to particular entities e.g. containers within their per-project quotas (related to processing, memory, data network addresses etc.). In such model of containerized infrastructure management the daily platform owner intervention is not required. The operator only sets the projects' limits and quotas. The actual detailed resources' provisioning is handled on the user level (by user, or orchestration solutions, on behalf of the user).

2.2.2 Compute Service

Description

Managed Compute (Virtual Machine) Infrastructure service provides virtualized infrastructure-level cloud services (IaaS) for computing, data storage, access and management and network communication.

The service is built using open source de facto standard components. Computing service is based on OpenStack, data storage and access service uses Ceph, and network stack is based on Neutron.

All services components are programmable and API driven which enables automation of deployment, configuration, provisioning (Infrastructure as a Code) and orchestration using Terraform, Ansible etc. In addition OpenStack provides the GUI (Horizon) for managing user projects and resources, configuration, monitoring and virtual machines access. Also Ceph provides monitoring and management GUIs,

The computing, data and network services are run on top of the on-premise infrastructure including computing servers, disk servers, storage and network infrastructure in the two hosting sites (PSNC, Safespring) of the contractors, interconnected by dedicated network links configured through NORDUnet. The infrastructure services will be deployed in both data centers, and integrated in a loosely coupled setup, i.e. active-backup high availability configuration. Basic instances of the services will be run at sites (user load will be distributed across sites), and critical data and meta-data will be subject of local backup procedures and will be replicated to the other site, to enable recovery in case of failures.

Component Details

Business aspects:

Basic compute, data and network services are provided by the on-premise infrastructure at PSNC and Safespring – the contractors' hosting sites (data centers). This approach, accompanied by relevant security processes and procedures such services implementation ensures that data sovereignty, data privacy and GDPR aspects are handled properly, according to best practices and EU regulations and

recommendations. In particular no user data nor meta-data are stored outside EU. Also the actual network communication within the infrastructure is performed based on dedicated dark fiber links.

Economic scalability of the service is ensured by the fact that the basic service tiers (T-shirt sizes), are clearly defined along with transparent cost model. Data ingress/egress- or I/O-related fees are not applied (otherwise they could constitute big component of the cost in the I/O intensive application).

Costs are related to the compute resources, data storage and I/O handling capability and network resources offered as well as actual usage that can be monitored and controlled using functionality provided by the contractor. Service tiers can be flexibly adopted based on the EC request.

Usage of open source components for infrastructure management, resources provisioning and orchestration enables keeping unit cost relatively low, compared to e.g. VMware-based systems and other commercial solution, where costly, yearly subscription-based models are currently promoted.

Functionality aspects:

The infrastructure-level services support for programmatic integration and self-service capability. They implement the model of projects (tenants in OpenStack terminology) that is the concept of grouping entities (VM, storage pools) and resources (computing, memory, storage) into projects dedicated to users or user groups. Users may allocate the chunks of the resources to particular entities e.g. VMs within their per-project quotas (related to processing, memory, data network addresses etc.). In such model of of infrastructure management the daily infrastructure owner intervention is not required. The only intervention / activity relates to setting up the projects' limits and quotas, while the actual detailed resources' provisioning is handled on the user level (by user, or orchestration solutions, on behalf of the user), This enables elastic and flexible as well as scalable management of physical vs logical resources.

Architecture aspects:

The service is provided based on open/modular architecture that facilitates integration and reusability:

- Modular architecture:
    - The architecture components – OpenStack-based orchestration layer, KVM-based server virtualization, Software Defined Storage based on Ceph, and Network based on Neutron – are standard elements of the IaaS architecture. This enables integration with various types of physical infrastructure (including regular servers running Linux OS with KVM) and deploying higher-level services on top of IaaS including container platform (see Managed Container Platform Service).
    - The components of the software stack can be exchanged, for instance the network stack of OpenStack can be implemented using various products including Neutron or Calico.
    - Virtualized resources can be organized into separated data centers that are managed independently and integrated on the level of container platform or application or the can be configured into regions thus providing the integrated view and management.
- Integration;
    - OpenStack provides a range of API for manipulating projects, quotas and infrastructure components including compute, storage, network, authentication and authorization etc. It enables Infrastructure as a Code model and orchestration using Terraform, Ansible etc. It also provides interfaces needed to integrate with provisioning mechanics of LOT1.
    - Ceph exposes management API as well as various data storage and access protocols including RBD (volumes), S3 (object storage) and CephFS (filesystem);

2.2.3 Bulk Data Transfer Service

Description

Managed Bulk Data Transfer service enables performing massive data transfer into and from the visualized and containerized infrastructure of the EOSC EU Node. It supports data transfer parallelism and applies performance optimisation as well as proactive monitoring of the transfer status and progress and restarts. The default use case of the solution is to stage large datasets owned by users, communities or projects into the back-end storage of the cloud platform and make this data accessible for computing and data analysis processes and other applications run in the virtualised infrastructure and containerized platform. The solution also provides functionality to export large datasets from the EOSC EU Node to the external user, community or project-owned data repositories or infrastructure.

The Managed Bulk Data Transfer can be also used as a solution for performing massive data migration within the user's or institution opt-in and opt-out processes.

It is an infrastructure-level solution, as opposed to file transfer solution delivered within LOT3 based on FileSender (that is an end user-oriented tool for handling application-level data sharing workflows).

The service is accessible through a set of CLI tools, APIs. Web-based transfer management tool(s) can be used to initiate and monitor the massive data transfers.

Component Details

Business aspects:

The data transfer service is delivered based on the on-premise infrastructure and services and composed of reusable functional modules with APIs that enable integration into various work flows.

The infrastructure and low level services are provided by PSNC and Safespring, the contractors' hosting sites (data centers). Accompanied by relevant management processes and security procedures such implementation ensures that data sovereignty, data privacy and GDPR aspects are handled properly, i.e. according to with best practices and EU regulations and recommendations. No user data nor meta-data are stored outside EU. User authentication and authorisation is based on federated AAI, provided by European organizations. The actual data transfer is kept within the EU borders (which is not the case of tools such as Globus Online used in past for scientific projects or WeTransfer used by individuals).

Economic scalability of the service is ensured by the fact that the basic service tiers (T-shirt sizes), chunks of resource used etc. are clearly defined and can be influenced by EC and contractors. Data ingress/egress fees are not applied (otherwise they could constitute big component of the cost in the data transfer service). Costs are related to the data storage capacity offered and actual storage space usage as well as to performance factors (number of parallel sessions to be handled by the service).

2.3 LOT3 - Exchange Application Services

2.3.1 EFSS – Enterprise File Synchronisation and Sharing Service

Description

Managed File Synchronization and Sharing Service provides functionality for data and files storage, management, sharing, manipulation, editing, collaborating and transferring.

The functionality and features are provided based on the on-premise infrastructure and low-level virtualization and private cloud services (compute, storage, network) provided by the hosting sites (PSNC, Safespring) which ensures that data sovereignty, data privacy and GDPR aspects are handled according to with best practices and EU regulations and recommendations.

The service is built based on the *de facto* standard file sync & share application OCIS (ownCloud Infinity Scale). It supports relevant industry protocols related to data and meta-data exchange, data sharing and data-based collaboration workflows including OCM and ScienceMesh.

ownCloud also provides APIs necessary for integration of user-level functionality (file storage, access, transferring, sharing, editing, moving, copying, meta-data assignment and meta-data based search) and performing administration-level work flows (creating accounts, setting quotas, checking resources usage, freezing accounts etc.).

Component Details

Business aspects:

Managed File Synchronization and Sharing Service provides file sync & share functionality including data and files storage, sharing, manipulation, editing, collaborating and transferring.

It is delivered based on the on-premise infrastructure and services and composed of reusable functional modules with APIs that enable integration into various users', groups' and institutional work flows.

Managed File Synchronization and Sharing Service is deployed based on the on-premise infrastructure at PSNC and Safespring, the contractors' hosting sites (data centers) that provide basic infrastructure capabilities (hosting, power supply, air conditioning, rooms, physical protection, security) and virtualization and cloud services including cloud computing (based on OpenStack and KVM), data storage (based on Ceph) and network based on orchestrated network configuration and dark fiber links.

Accompanied by relevant management processes and security procedures such implementation ensures that data sovereignty, data privacy and GDPR aspects are handled properly, i.e. according to best practices and EU regulations and recommendations. No user data nor meta-data are stored outside EU. User authentication and authorization is based on federated AAI, provided by European organizations.

Economic scalability of the service is ensured by the transparent cost model, with no hidden costs. The basic service tiers (T-shirt sizes), are well-defined upfront and they and can be changed over the contract execution period based on the consumption analysis and control by EC and contractors. Data ingress/egress fees are not applied (otherwise they could constitute big component of the cost in case of I/O intensive application). Costs are related to the data storage capacity provisioned and consumed as well as to performance measures (number of parallel session to be handled by the service).

Architecture aspects:

The service is provided based on open and modular architecture enabling integration and reusability:

- Modular architecture:
  - The architecture components – server: ownCloud OCIS server and clients: Web, workstation, mobile - represent well-understood building blocks of sync & share system;
  - the current set of clients can be extended using the existing APIs; custom integrations can be worked out based on the extensive API;
- Integration;
  - The file sync and share application server provides the integration APIs:
    - Proprietary APIs for file sync & share – used to implement the ownCloud clients
    - Standard APIs for data storage /access WebDAV – can be used for integration with external systems (e.g. data push, data download using standard tools)
  - Compliant to OCM (Open Cloud Mesh) standard:
    - Can exchange data with other file sync & share products: NextCloud, Seafile etc.
    - Can be integrated with EOSC ecosystem (various local EFSS solutions may exist)
    - *There is an exit path in case the product is no longer supported / developed*
  - Supports ScienceMesh invitation workflow:
    - Currently based on REVA, can be provided as more lightweight integration
    - Can be part of file sync & share systems federation ScienceMesh
- Reusability:
  - Server:
    - Set of clients exist, more clients can be developed using documented API
    - Server is compliant to OCM (Open Cloud Mesh) standard
      - Might be applied in other data management/compute services stacks
      - Can be replaced with products of similar functionality (if needed)
  - Clients:
    - Functionality: clients support various work flows and user-service interactions:
    - Web-based portal:
      - Access ubiquity - available through typical web browser
      - Web app has no direct access to user files – no automated sync
    - Workstation clients/mobile clients
      - Workstation client for mainstream platforms (Windows, MacOS, Linux)
      - Mobile clients: Android, iOS

Direct access to user data – enables changes tracking, synchronization, massive data transfers / metadata updates.

2.3.2 Interactive Notebooks

Description

Managed Interactive Notebook service provides Data Science environment, data analytics solution, based on Jupyter Notebooks platform and on-premise compute infrastructure.

The service is built based on the de-facto standard and widely-adopted platform: Jupyter Notebooks. This ensures the overall usability and approachability of the service from the perspective of the user base assumed (data scientists from various disciplines, users with various computing resources needs).

Jupyter Notebooks architecture ensure functional flexibility. It has extensive integration possibilities including support for executing the computing kernels various compute environments (local, cloud, HPC), using various data sources and data storage systems etc. Its interface for developing system extensions (plug-ins) can be used for offering specialized functionality or perform integrations .

The computing and data analytics environment is run on top of the on-premise infrastructure including compute, storage and network systems provided by hosting sites (PSNC, Safespring), which addresses data sovereignty, data privacy concerns and proper handling of GDPR.

Cost efficiency, predictability and control is achieved by applying transparent cost model: clearly defined service tiers (T-shirt sizes), no hidden costs (I/O or data transfer-related) and providing usage monitoring capabilities. Scalability is ensured by possible scale-up of consumption using LOT2-provided resources.

Component Details

Business aspects:

Managed interactive Notebook service based on Jupyter Notebooks and on-premise infrastructure enables developing and executing data analysis workflows and kernels based on various technologies including Python, Octave, R-Studio and Matlab. It provides Web-based interface for direct user interaction. The interface can also be embedded in and integrated with other systems (portals etc.).

Jupyter Notebooks enable using various computing resources including scale-out to cloud and HPC as well as various data storage systems (e.g. based on sync & share system, cloud storage, data lakes etc.) for executing and storing the kernels as well as handling input and output data for analytic processes.

Basic compute and data resources for Managed interactive Notebook Service are provided by the on-premise infrastructure at PSNC and Safespring – the contractors' hosting sites (data centers) - that operate cloud computing (OpenStack, OKD), storage (Ceph) platforms and network links (LAN, WAN).

Accompanied by relevant management processes and security procedures such implementation ensures that data sovereignty, data privacy and GDPR aspects are handled properly, i.e. according to with best practices and EU regulations and recommendations. No user data nor meta-data are stored outside EU. User authentication and authorization is based on federated AAI, provide by European organizations.

Economic scalability of the service is ensured by the fact that the basic service tiers (T-shirt sizes), are clearly defined along with transparent cost model. Data ingress/egress- or I/O-related fees are not applied (otherwise they could constitute big component of the cost in the I/O intensive application). Costs are related to the compute resources, data storage and I/O handling capability offered and actual usage that can be monitored and controlled using functionality provided by the contractor.

Architecture aspects:

The service is provided based on open/modular architecture that facilitates integration and reusability:

- Modular architecture:
  - Server provide a thin resource access and integration layer as well as computing and resources orchestration functionality.
  - Web based access ensures the access ubiquity, prevent access barriers (approachability) and integration options (considered within LOT1 user portal)
- Integration;
  - Jupyter Notebooks provide range of API for manipulating user files, kernels, sessions, terminals, and developing / hooking system extensions (plug-ins).
  - Developing or integrating specialized functionality is possible, e.g. for collaborative notebooks editing within a user group, storing the data in EFSS etc.
- Reusability:
  - Notebook Service can be deployed on various types and 'tiers' of compute, storage and network resources (infrastructure components, low-level cloud services components)

Notebook service can use resources pools assigned to LOT3 by its contractors. There is also technical and contractual option to scale-out to larger resources based on LOT2 managed Compute Infrastructure Services and Managed Container Platform Service.

2.3.3 Large File Transfer Service

Description

Managed Large File Transfer service provides data transfer capability accessible from Web front-end, that allows authenticated user to securely, reliably and easily transfer large files to the end user. Users can upload files, create download links and inform by email the recipients on the data made available.

The service is provided based on FileSender, a data transfer solutions developed and widely adopted by academic network operators, computing centers, universities and NRENs (national research network operators). It has AAI federation capability as well as provides integration APIs. FileSender-provided data transfer mechanics can be integrated with application workflows, including sync & share, data science.

FileSender APIs enable programmatic interaction with the service including possibility to initiate data transfer, triggering file upload/download, defining and updating transfer recipients. manipulating files' meta-data, defining recipients and vouchers etc.

The services is deployed on top of the on-premise infrastructure and cloud services including compute, storage and network provided by hosting sites (PSNC, Safespring), which addresses data sovereignty, data privacy concerns and proper handling of GDPR as well as service delivery and operation cost efficiency and usage transparency of cost model applied and costs predictability.

Component Details

Business aspects:

The data transfer services is delivered based on the on-premise infrastructure and services and composed of reusable functional modules with APIs that enable integration into various work flows.

The infrastructure and low level services are provided by PSNC and Safespring, the contractors' hosting sites (data centers). Accompanied by relevant management processes and security procedures such implementation ensures that data sovereignty, data privacy and GDPR aspects are handled properly, i.e. according to with best practices and EU regulations and recommendations. No user data nor meta-data are stored outside EU. User authentication and authorization is based on federated AAI, provided by European organizations.

Economic scalability of the service is ensured by the fact that the basic service tiers (T-shirt sizes), chunks of resource used etc. are clearly defined and can be influenced by EC and contractors. Data ingress/egress fees are not applied (otherwise they could constitute big component of the cost in the data intensive application). Costs are related to the data storage capacity offered and actual storage space usage as well as to performance factors (number of parallel session to be handled by the service).

Architecture aspects:

The service is provided based on open and modular architecture that enables integration and reusability:

- Modular architecture:
    - The architecture components – represents commonly-understood building blocks.
    - Server – the FileSender application that provides core service functionality and exposes the integration APIs.
    - Clients – Web based access ensures the access ubiquity, prevent access barriers (approachability), deep integration options (considered within LOT1 user portal)
- Integration;
    - Data transfer integrations can be developed for other services e.g. Jupyter Notebooks – that would enable transparent transfer of data used by notebooks for data analysis and/or making the computing results available/downloadable) to/by other users.
- Reusability:

Part of the Filesender-based data transfer service can be used as functional components for other services – e.g. data transfers within LOT2 (as an alternative to relatively heavy and complex FTS-based bulk data transfer solution).