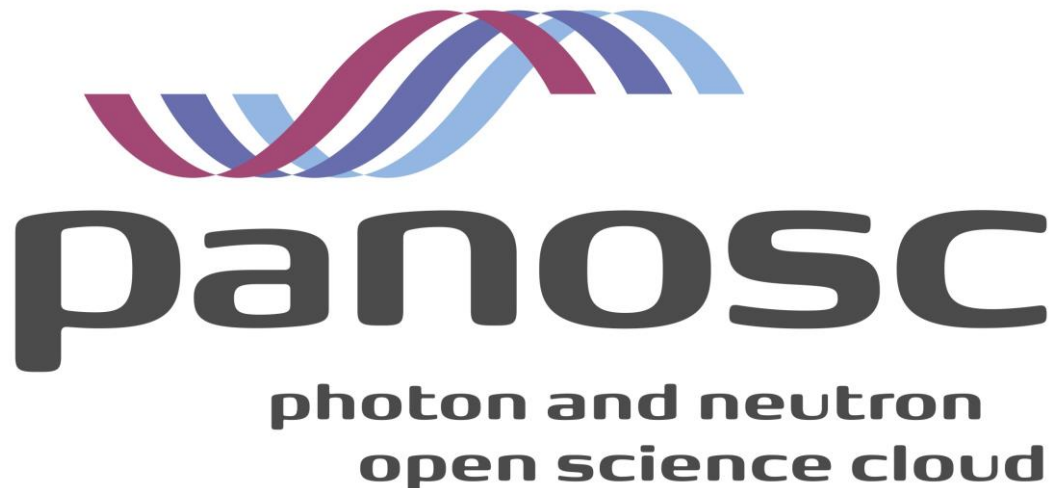


PaNOSC

Photon and Neutron Open Science Cloud

H2020-INFRAEOSC-04-2018

Grant Agreement Number: 823852



Deliverable: D6.6 Integration of RIs in EOSC

Project Deliverable Information Sheet

Project Reference No.	823852
Project acronym:	PaNOSC
Project full name:	Photon and Neutron Open Science Cloud
H2020 Call:	INFRAEOSC-04-2018
Project Coordinator	Andy Götz (andy.gotz@esrf.fr)
Coordinating Organization:	ESRF
Project Website:	www.panosc.eu
Deliverable No:	D6.6
Deliverable Type:	Report
Dissemination Level	Public
Contractual Delivery Date:	30/11/2022
Actual Delivery Date:	15/12/2022
EC project Officer:	Flavius Alexandru Pana

Document Control Sheet

Document	Title: Integration of RIs in EOSC
	Version: 1
	Available at: https://github.com/panosc-eu/panosc
	Files:1
Authorship	Written by: Giuseppe La Rocca (EGI.eu), Andrea Manzi(EGI.eu)
	Contributors: Jean-François Perrin(ESRF), Fabio Dell'Antonia (XFEL.eu)
	Reviewed by: Jordi Bodega Sempere (ESRF)
	Approved: Andrew Gotz (ESRF)

List of participants

Participant No.	Participant organisation name	Country
1	European Synchrotron Radiation Facility (ESRF)	France
2	Institut Laue-Langevin (ILL)	France
3	European XFEL (XFEL.EU)	Germany
4	The European Spallation Source (ESS)	Sweden
5	ELI European Research Infrastructure Consortium (ELI-ERIC)	Belgium
6	Central European Research Infrastructure Consortium (CERIC-ERIC)	Italy
7	EGI Foundation (EGI.eu)	The Netherlands

Table of Content

Project Deliverable Information Sheet	2
Document Control Sheet	2
List of participants	2
Table of Content	3
1. Introduction	5
2. On-boarding PaN services in EOSC	6
2.1. The EOSC onboarding procedure in a nutshell	6
2.2. The EOSC Portal Onboarding Team (EPOT)	7
2.3. Benefits of becoming a service providers for EOSC	7
2.4. How to onboard resources in the EOSC Portal Registry	7
2.4.1. Comply with the EOSC Provider inclusion criteria	8
2.4.2. Validation of the criteria for becoming an EOSC service provider	8
2.4.3. Onboard as a service provider	8
2.4.4. Onboard your first resource	8
2.5. Documentation	8
2.6. Training	9
3. PaN providers and resources integrated in EOSC	10
3.1. PaN providers registered the EOSC Portal	10
3.1.1. Institute Laue-Langevin (ILL)	10
3.1.2. European Synchrotron Radiation Facility (ESRF)	11
3.1.3. European XFEL (XFEL.EU)	11
3.1.4.	11
3.1.5. European Spallation Source (ESS)	12
3.1.6. Central European Research Infrastructure Consortium (CERIC-ERIC)	12
3.2. PaN resources registered in the EOSC Portal	12
3.2.1. Institute Laue-Langevin (ILL)	13
3.2.2. European Synchrotron Radiation Facility (ESRF)	13
3.2.3. European Spallation Source (ESS)	13
3.3. Task-Force	13
3.3.1. Service providers onboarded in EOSC	13
3.3.2. PaN services onboarded in EOSC	14
3.4. PaN providers and services planned to be onboarded in the EOSC Portal	14
3.4.1. European XFEL (XFEL.EU)	14
3.4.2. Central European Research Infrastructure Consortium (CERIC-ERIC)	14
3.4.3. European Spallation Source (ESS)	14

3.4.4. European Synchrotron Radiation Facility (ESRF)	14
4. Transverse/support services	16
4.1. Helpdesk integration (use case for EOSC-Future)	16
4.2. Authentication and Authorization Infrastructure (AAI)	17
4.3. Data Transfer	18
4.3.1. Data Archiving	18
4.3.2. Transparent Data Access for Data Analysis	18
4.3.3. Data Transfer to Home PCs	20
5. Difficulties	21
5.1. Issues with the onboarding of PaN resources in EOSC	21
5.2. Lack of data transfer service for service composability in EOSC	21
5.3. Helpdesk and its sustainability	22
6. Summary and Outlook	23
7. References	24

1. Introduction

The Photon and Neutron Open Science Cloud (PaNOSC) project aims at providing a remote access infrastructure to enable and contain FAIR data services for users of the community and ideally scientists across domain borders, through the European Open Science Cloud (EOSC). In the centre of this project, taken care of by Work Package 6, lies the integration of the Photon and Neutron (PaN) data catalogues and services in the EOSC. As part of this Work Package, D6.6 provides the plan and the strategy to provision domain-specific services developed by WP3 - Data Catalog Services, WP4 - Data Analysis Services, WP5 - Virtual Neutron and X-ray Laboratory (VINYL) and WP8 - Staff and User Training into the EOSC Portal and make them available to the scientific community.

The present deliverable is organized as follows:

- Section 2 introduces the service onboarding process in EOSC. A prerequisite to onboarding new services in the EOSC portfolio is to have production level services with high Technology Readiness Levels (TRL)¹ (TRL \geq 7) and a legal entity responsible to operate and maintain the service in production level for end-users and provide them support.
- Section 3 provides an overview of the current PaN service providers and resources currently available in EOSC and their future plans for onboarding new resources.
- Section 4 introduces the main EOSC core services of interest for the PaN community and transversal services that are needed by their end users. For each of them we describe the activities that have been carried out or future plans of integration when possible.
- Section 5 highlights the difficulties and challenges raised by the PaN community during the project's lifetime. Some of them are still open (e.g.: the Data Transfer solution in part of the EOSC Procurement), for others an initial solution is proposed, or is under investigation (e.g.: agree a co-ownership in the PaN community for on-boarding new PaN resources in EOSC).
- Summary and outlook are reported in Section 6.

¹ https://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp1415-annex-g-trl_en.pdf

2. On-boarding PaN services in EOSC

2.1. The EOSC onboarding procedure in a nutshell

The EOSC Portal Onboarding Process is the process (see also Figure 1) that a service provider must follow in order to register new “Resources” in the EOSC Portal Registry². Overall, the EOSC Portal Registry provides EOSC users with a list of live/ready-to-use descriptions of EOSC Resources offered by EOSC for supporting advanced research and helping scientific researchers and communities to run their daily work. Each entry of the EOSC Portal Registry must comply with and be described and updated following the EOSC Portal Interoperability Framework³.

The EOSC Portal Interoperability Framework includes, among others:

- the EOSC Profiles⁴,
- the EOSC Rules of Participation (RoP)⁵
- the EOSC Portal Application Programming Interface⁶ (APIs) methods for the automatic provisioning and synchronisation of information between Providers’ systems and the EOSC Portal Registry.

From a technical perspective, domain-specific Photon and Neutron services can be registered in the EOSC Portal Registry from the dedicated EOSC Providers Portal⁷, a platform where service providers can onboard their organisations, register and manage their resources, and track the usage of the resources. More specifically, through the EOSC Providers Portal, service providers can:

- Manage their complete resource portfolio in the EOSC Portal,
- Update the information about their organisation and resources,
- Configure the ordering settings,
- Create draft resources,
- Publish or unpublish resources in the Marketplace,
- View resources shared and co-provided with other organisations in the EOSC Portal,
- View messages coming from users,
- Get access to statistics about visits and use of their offerings, as well as a complete timeline of changes and actions applied to their resources.

² <https://marketplace.eosc-portal.eu/>

³ <https://eosc-portal.eu/news/eosc-interoperability-framework-consultation>

⁴ <https://eosc-portal.eu/eosc-providers-hub/what-are-eosc-profiles>

⁵ <https://op.europa.eu/en/publication-detail/-/publication/a96d6233-554e-11eb-b59f-01aa75ed71a1>

⁶ <https://providers.eosc-portal.eu/developers>

⁷ <https://providers.eosc-portal.eu/>

2.2. The EOSC Portal Onboarding Team (EPOT)

A dedicated EOSC Portal Onboarding Team validates the provided information by the service providers against the Inclusion Criteria derived from the Rules of Participation. In the case of invalid or outdated information (during and after onboarding) regarding their offerings, the providers will be notified by EPOT and proceed with improvement actions. The EPOT is fundamental in achieving trust in the EOSC by checking compliance to the EOSC Rule of Participation and the EOSC Profiles.

2.3. Benefits of becoming a service providers for EOSC

The benefits of becoming a service providers for EOSC and contributing to the EOSC Exchange, are the following:

- Being able to add new resources to the EOSC Portal Registry.
- Promote your services in EOSC and attract new users through the EOSC Portal Registry.
- Integrate new resources in the EOSC core resources. For example, by integrating with:
 - EOSC Helpdesk
 - Access Monitoring
 - Usage Accounting
 - Order Management
 - Federated AAI
 - Open Science Monitor

2.4. How to onboard resources in the EOSC Portal Registry

Fig. 1 shows the four step-staged process to follow in order to onboard new services and resources on the EOSC Portal Registry through the EOSC Provider Portal.

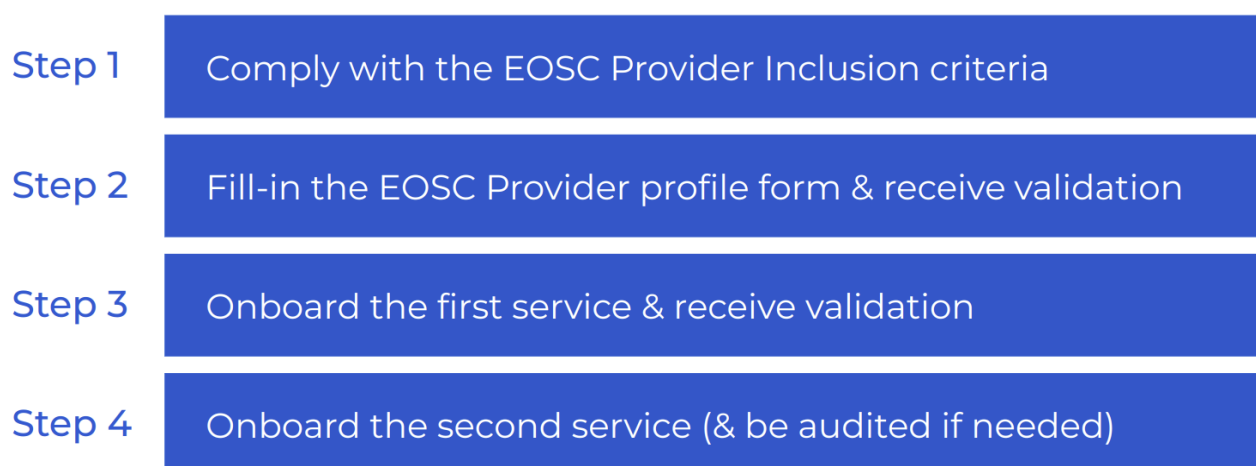


Figure 1 Onboarding providers and resources in the EOSC Portal Registry

2.4.1. Comply with the EOSC Provider inclusion criteria

In this section we report the criteria service providers have to comply with in order to become an EOSC service provider.

More specifically:

- **“Resources” must be onboarded by a legal entity** (although the legal entity may do so on behalf of a project or consortium in which they participate, with the agreement of those groups).
- **Providers** onboarding a resource **must assert that** they are able to ensure **the resource is delivered by them or their collaborators and agree to remove resources which are no longer operational or available**.
- Comply with the EOSC Rules of Participation.
- Provide the service provider profile in the EOSC Portal Registry.
- Onboard resources to target EOSC and EOSC community and contribute to the EOSC Exchange, or build or leverage EOSC capabilities to serve some other communities.
- Commit to an effective service delivery.
- Agree on periodically updating data about the service provider and the resources served in the EOSC Portal Registry.

2.4.2. Validation of the criteria for becoming an EOSC service provider

The EOSC Portal On-boarding Team (EPOT) manages and assesses compliance with the inclusion criteria. For further information, use the EPOT mailing-list (onboarding@eosc-portal.eu) for queries about the EOSC onboarding procedure.

2.4.3. Onboard as a service provider

Go to the EOSC Providers Portal to apply to become a provider and fill-in your submission form.

2.4.4. Onboard your first resource

Once you have been onboarded as a provider, you will be able to onboard services.

Go to your provider account in the [EOSC Providers Portal](#) and start providing information about the first resource. As soon as the resource onboarding submission is reviewed, the EOSC Portal Onboarding Team will validate the first resource, or ask for clarifications.

2.5. Documentation

- EOSC Portal Registry: <https://eosc-portal.eu/for-providers/>
- Onboarding of a service provider: <https://providers.eosc-portal.eu/>
- EOSC “Becoming a provider”: <https://providers.eosc-portal.eu/becomeAProvider>
- EOSC Provider Documentation: <https://eosc-portal.eu/providers-documentation>

2.6. Training

- Onboarding of a Provider: <https://youtu.be/wWGN7WzNuCQ>
- Validating/auditing of a Provider: <https://youtu.be/FW1ly-IE3Gs>
- Onboarding of a Resource: https://youtu.be/nsWPW7Qjc_s
- Validating/auditing of a Resource: <https://youtu.be/nHULTPs5SI8>

All these training materials are currently available in the EOSC Future Youtube channel⁸ and we plan to onboard them into the EOSC Portal as soon as the procedure to onboard training resources is in production.

⁸ <https://www.youtube.com/c/EOSCPortal>

3. PaN providers and resources integrated in EOSC

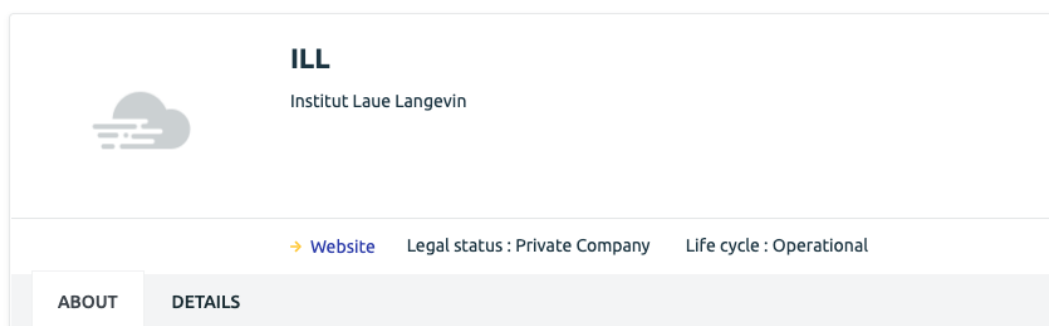
In this section we report the list of PaN providers and resources already on-boarded in EOSC.


3.1. PaN providers registered the EOSC Portal

As previously stated, prior to the onboarding of services to the EOSC portal, the providers of the services need to be registered and approved by the onboarding team. The EOSC portal shows the information about the registered providers at: <https://marketplace.eosc-portal.eu/providers> (238 providers registered by the end of September 2022). In this section we report the PaN providers that have been successfully onboarded in the EOSC Portal at the time of the deliverable writing.

3.1.1. Institute Laue-Langevin (ILL)⁹


[Providers](#) > Institut Laue Langevin




ILL

 Institut Laue Langevin

[Website](#) Legal status : Private Company Life cycle : Operational

ABOUT DETAILS

The Institut Laue-Langevin is an international research centre at the leading edge of neutron science and technology.

As the world's flagship centre for neutron science, the ILL provides scientists with a very high flux of neutrons feeding some 40 state-of-the-art instruments, which are constantly being developed and upgraded.


As a service institute the ILL makes its facilities and expertise available to visiting scientists. Every year, about 1400 researchers from over 40 countries visit the ILL and 640 experiments selected by a scientific review committee are performed. Research focuses primarily on fundamental science in a variety of fields: condensed matter physics, chemistry, biology, nuclear physics and materials science, etc.

Figure 2 The ILL service provider in the EOSC Portal Registry

⁹ <https://marketplace.eosc-portal.eu/providers/eosc.ill>

3.1.2. European Synchrotron Radiation Facility (ESRF)¹⁰

🏠 > Providers > European Synchrotron Radiation Facility



ESRF
European Synchrotron Radiation Facility

[→ Website](#) Legal status : Non For Profit Company

ABOUT


DETAILS

The ESRF is one of the world's largest synchrotron science centres. Every year, 9000 scientists from 22 partner countries and from around the world travel to Grenoble to use its extremely brilliant X-rays for leading-edge research. This fundamental and applied research often addresses how to better use resources, develop more sustainable materials or improve our daily lives. The ESRF Upgrade Programme 2009-2018 maintains our world-leading role through a continuous quest for higher performance figures, meeting the needs of returning users, and attracting scientists from new disciplines.

Figure 3 The ESRF service provider in the EOSC Portal Registry

3.1.3. European XFEL (XFEL.EU)¹¹

🏠 > Providers > European X-ray Free Electron Laser Facility GmbH



European XFEL
European X-ray Free Electron Laser Facility GmbH

[→ Website](#) Legal status : Non For Profit Company Life cycle : Operational

ABOUT

DETAILS

European XFEL in the Hamburg area is an international research facility of superlatives: 27,000 X-ray flashes per second and a brilliance that is a billion times higher than that of the best conventional X-ray sources open up completely new opportunities for science. Research groups from around the world are able to map the atomic details of viruses, decipher the molecular composition of cells, take three-dimensional "photos" of the nanoworld, "film" chemical reactions, and study processes such as those occurring deep inside planets. The operation of the facility is entrusted to European XFEL, a non-profit company that cooperates closely with its main shareholder, the research centre DESY, and other organisations worldwide. European XFEL has a workforce of more than 450 employees and started user operation September 2017. At present, 12 countries have signed the European XFEL convention: Denmark, France, Germany, Hungary, Italy, Poland, Russia, Slovakia, Spain, Sweden, Switzerland, and the United Kingdom. For more information on European XFEL go to www.xfel.eu.

Figure 4 The XFEL service provider in the EOSC Portal Registry


3.1.4.

¹⁰ <https://marketplace.eosc-portal.eu/providers/eosc.esrf>

¹¹ https://marketplace.eosc-portal.eu/providers/eosc.european_xfel

3.1.5. European Spallation Source (ESS)¹²

🏠 > Providers > European Spallation Source ERIC



ESS

European Spallation Source ERIC

[→ Website](#)
Legal status : European Research Infrastructure Consortium (ERIC)
Life cycle : Under Construction

ABOUT


DETAILS

The European Spallation Source is a research infrastructure based in Sweden and Denmark. The organisation is constructing a spallation neutron source, when operational ESS will provide the brightest source of cold neutrons and next generation scientific instrumentation that will allow visiting researchers from a wide variety of scientific backgrounds to investigate the structure and dynamics of materials. Along side experimental infrastructure ESS delivers data and data services to the scientific user community, with an emphasis on FAIR data and Open Science.

Figure 5 The ESS service provider in the EOSC Portal Registry

3.1.6. Central European Research Infrastructure Consortium (CERIC-ERIC)¹³

🏠 > Providers > Central European Research Infrastructure Consortium



CERIC-ERIC

Central European Research Infrastructure Consortium

[→ Website](#)
Legal status : European Research Infrastructure Consortium (ERIC)
Life cycle : Operational

ABOUT

DETAILS

CERIC is a European Research Infrastructure Consortium (ERIC) integrating and providing open access to some of the best facilities in Europe, to help science and industry advance in all fields of materials, biomaterials and nanotechnology. With a single entry point to some of the leading national research infrastructures in 8 European countries, it enables the delivery of innovative solutions to societal challenges in the fields of energy, health, food, cultural heritage and more. More info <https://www.ceric-eric.eu/about-us/who-we-are>

Figure 6 The CERIC-ERIC service provider in the EOSC Portal Registry

3.2. PaN resources registered in the EOSC Portal

This section reports the PaN services that have been successfully registered in the EOSC Portal at the time of the deliverable writing, grouped by provider.

¹² <https://marketplace.eosc-portal.eu/providers/eosc.ess>

¹³ <https://marketplace.eosc-portal.eu/providers/eosc.ceric-eric>

3.2.1. Institute Laue-Langevin (ILL)

ILL as provider, has onboarded the following services to the EOSC portal:

- **PaNOSC Software Catalogue**¹⁴
- **VISA - Virtual infrastructure for scientific analysis**¹⁵
- **ILL Data Portal**¹⁶

3.2.2. European Synchrotron Radiation Facility (ESRF)

ESRF as provider, has onboarded the following services to the EOSC portal:

- **Jupyter Notebooks**¹⁷
- **The ESRF Data Portal**¹⁸

3.2.3. European Spallation Source (ESS)

ESS as provider, has onboarded the following services to the EOSC portal:

- **Pan-Learning.org**¹⁹

3.3. Task-Force

In order to facilitate the onboarding of new domain-specific services in the EOSC Portal Registry, a dedicated task-force (TF) was launched in collaboration with the ExPaNDS²⁰ project. In the framework of this TF, members of the two projects were supported in the onboarding procedure, providing technical support in collaboration with the EOSC Portal Onboarding Team. The TF was officially launched in January 2022. After the kick-off of the TF dedicated one-to-one meetings were organized with the service providers to help their onboarding process. Since the TF was set-up, the following service providers and PaN services were onboarded in the EOSC Portal Registry.

3.3.1. Service providers onboarded in EOSC

- **Paul Scherrer Institute (PSI)**²¹

¹⁴ <https://marketplace.eosc-portal.eu/services/panosc-software-catalogue>

¹⁵ <https://marketplace.eosc-portal.eu/services/visa-virtual-infrastructure-for-scientific-analysis>

¹⁶ <https://marketplace.eosc-portal.eu/datasources/eosc.ill.d422cba59746f39d10bdfea5c9cf8511>

¹⁷ <https://marketplace.eosc-portal.eu/services/jupyter-notebook-can-be-used-to-create-and-share-documents-that-contain-live-code-equations-visualizations-and-text>

¹⁸ <https://marketplace.eosc-portal.eu/datasources/eosc.esrf.ecc74ab09791c52aa238ee77ae988874>

¹⁹ <https://marketplace.eosc-portal.eu/services/pan-learning-org-f68a8ee0-ed50-4eb1-b51a-b28516df7966>

²⁰ <https://expands.eu/>

²¹ <https://marketplace.eosc-portal.eu/providers/eosc.psi>

- Helmholtz-Zentrum Dresden-Rossendorf (HZDR)²²
- Synchrotron SOLEIL²³
- European Synchrotron Radiation Facility (ESRF)²⁴

3.3.2. PaN services onboarded in EOSC

- PSI Public Data Repository²⁵
- PSI Remote Desktop Service²⁶
- The European Synchrotron Radiation Facility Data Portal²⁷

3.4. PaN providers and services planned to be onboarded in the EOSC Portal

3.4.1. European XFEL (XFEL.EU)

XFEL.eu, although registered in the portal, has not yet onboarded any service.

As a plan for the next months, XFEL.eu is going to contribute registering the following resources in the EOSC Portal:

- **Data catalog myMDC**²⁸
- **VISA - Virtual infrastructure for scientific analysis**

3.4.2. Central European Research Infrastructure Consortium (CERIC-ERIC)

CERIC-ERIC has been already registered in the EOSC portal. The registration of the **CERIC Data portal**²⁹ is in progress.

3.4.3. European Spallation Source (ESS)

ESS has been already registered in the EOSC portal. The registration of the **e-learning PaN platform**³⁰ is in progress.

3.4.4. European Synchrotron Radiation Facility (ESRF)

²² <https://marketplace.eosc-portal.eu/providers/eosc.hzdr>

²³ <https://marketplace.eosc-portal.eu/providers/eosc.soleil>

²⁴ <https://marketplace.eosc-portal.eu/providers/eosc.esrf>

²⁵ <https://marketplace.eosc-portal.eu/datasources/eosc.psi.f1a79f572f95bc2fba5cdc40ef4eb22>

²⁶ <https://marketplace.eosc-portal.eu/services/remote-desktop-service>

²⁷ <https://marketplace.eosc-portal.eu/datasources/eosc.esrf.ecc74ab09791c52aa238ee77ae988874>

²⁸ <https://in.xfel.eu/metadata>

²⁹ <https://data.ceric-eric.eu/>

³⁰ <https://e-learning.pan-training.eu/>

As a plan for the next months, ESRF is going to contribute registering the following resources in the EOSC Portal:

- The **ESRF VISA instance**³¹ providing direct access to the ESRF data is in progress.
- The **Human Organ Atlas**³² which provides a specific view to the ESRF data catalog dedicated to the health user community.

³¹ <https://visa.esrf.f>

³² <https://human-organ-atlas.esrf.eu/>

4. Transverse/support services

4.1. Helpdesk integration (use case for EOSC-Future)

One of the main EOSC core services is the EOSC Helpdesk³³ which provides an entry point and ticketing system/request tracker for issues concerning the available EOSC services. The service also implements incident and service request management, and provides efficient communication channels between customers, users and providers of the IT resources and services. The high-level architecture of the EOSC Helpdesk is shown in Figure 7.

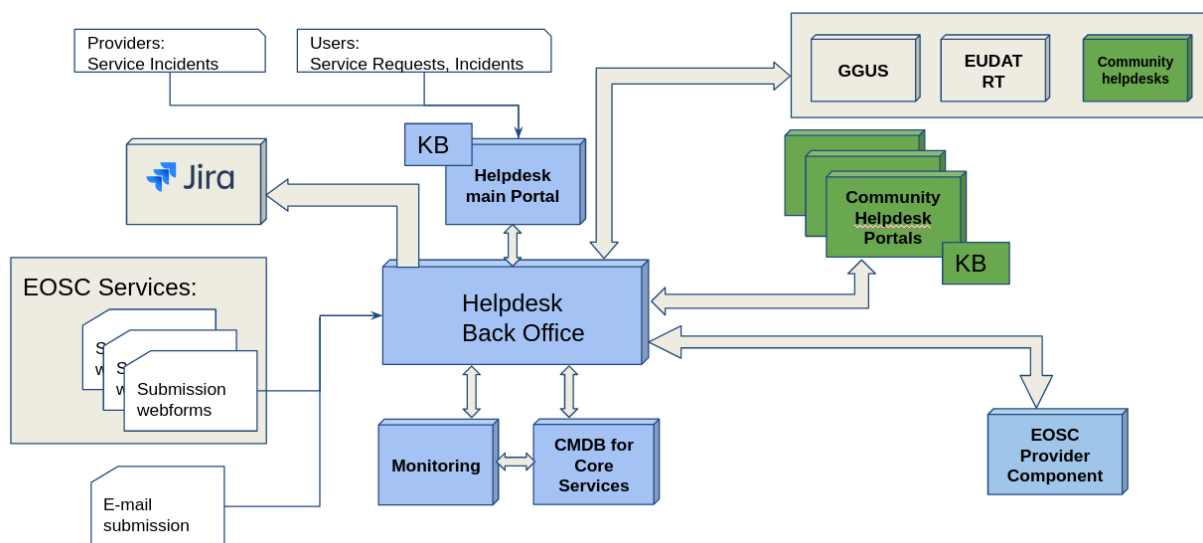


Figure 7 The EOSC Help-desk high-level architecture

As shown in the figure above, the EOSC Help-desk service is already integrated with multiple components including the:

- **EOSC submission webforms** provides ready to use webforms to be embedded on the dedicated portals of the EOSC services for easy submission of the request without logging to the helpdesk.
- **E-Mail submission** enables the integration of the helpdesk with a specific mailbox which can be offered for users to submit their requests via e-mail.
- **Community Helpdesks** allows bi-directional synchronization of the user requests created in EOSC Helpdesk for further processing in the integrated helpdesk.
- **Community Helpdesk Portals** enables the integration of the multiple user-faced community branded portals.
- **Integration with Configuration Management Database (CMDB) for EOSC-Core Services** enables assignment and tracking of the service request for the dedicated EOSC Core service.
- **Integration with the EOSC Monitoring system** provides easy access to live information about supported services, their production statuses, availability information.
- **Integration with EOSC Resource Catalogue** allows direct submission of issues and requests from the

³³ <https://eosc-helpdesk.eosc-portal.eu>

Provider Dashboard interface to the Helpdesk.

For the PaN service providers, EOSC offers the following three levels of interoperability with the service:

- **Full integration.** This integration path also includes full synchronization between the EOSC Helpdesk and the community helpdesk. This integration can be achieved by application of a set of helpdesk REST APIs. The exact integration guidelines should be defined based on the specifications agreed with the community.
- **Ticket redirection.** In this integration path the EOSC helpdesk is used only as a contact point to redirect the initial request to the provider's or community mailing list without further integration.
- **Direct usage.** In this last integration path the EOSC helpdesk can be used as the ticketing system for the community and their onboarded services.

4.2. Authentication and Authorization Infrastructure (AAI)

Since its inception in 2012, UmbrellaID³⁴ has been serving the AAI needs of the PaN community. One of the main results in PaNOSC WP6 - EOSC Integration has been the integration of the UmbrellaID AAI within the EOSC AAI general architecture (Figure 8) as Community AAI. This work has been partially supported by the EOSC Future project³⁵.

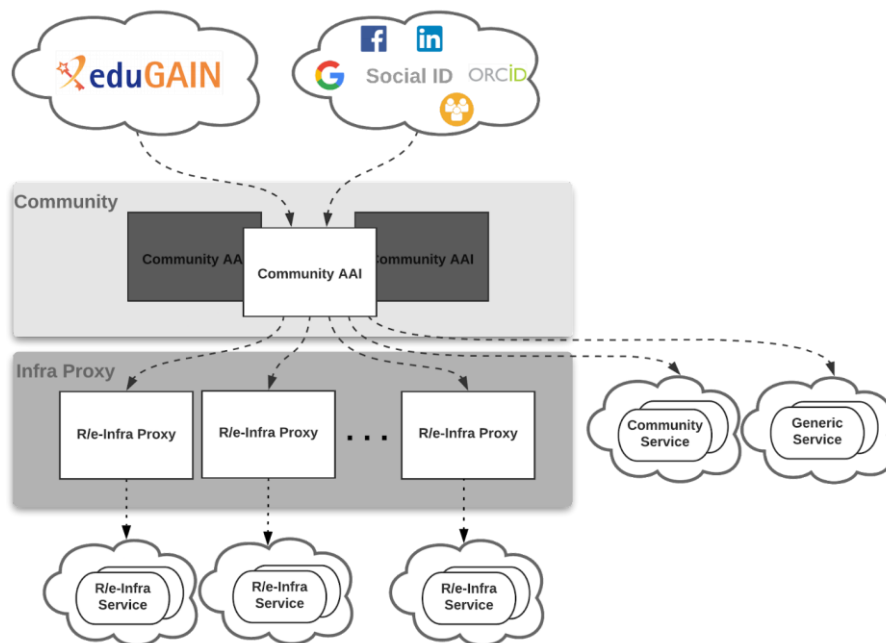


Figure 8 The EOSC AAI Architecture

From a technical point of view, the integration process is detailed in the PaNOSC deliverable D6.3 [1]. One of

³⁴ <https://umbrellaid.org/>

³⁵ <https://eoscfuture.eu/>

the first achievements has been the adoption of the eduTEAMS³⁶ service, operated by GÉANT³⁷, as the core infrastructure for the UmbrellaID AAI. The eduTEAMS infrastructure for UmbrellaID has been established in Autumn 2020 and a series of technical workshops have been organized to allow the migration of all PaN Service Providers by March 2021. The migration to eduTEAMS simplify the integration of UmbrellaID with EOSC AAI, where eduTEAMS is one of the Community AAI proxy, by:

- Enabling UmbrellaID users to access EOSC services, which are integrated with the EOSC AAI.
- Enabling access to PaNOSC services via EOSC AAI.

4.3. Data Transfer

4.3.1. Data Archiving

The use case is about implementing data archiving for raw data from experiments to an external facility and the restore of the data. STFC³⁸ has worked in this use case, both as a provider of storage to be used for archiving (ECHO Storage at STFC) and for investigating solutions to perform efficient backup. Initially the Rucio³⁹ and FTS⁴⁰ services have been evaluated in order to move data from ILL as a facility selected as pilot. Both services are available in the EOSC portal^{41, 42} and offered by STFC as a provider.

In order to use Rucio and FTS, there has been the need to install a gridftp server on the ILL facility as the data transfer is implemented only between storage endpoints. After the initial tests, the archive solution via Rucio has been discarded due to the need to ingest datasets inside the Rucio database which is a requirement to perform the data archiving operations.

The usage of rclone⁴³ has then been piloted, given that the storage at STFC offers also an S3 interface which is easily accessible via that tool. The pilot via rclone has given successful results in terms of usability and performance, and the amount of data archived at STFC storage has reached 200 TB.

4.3.2. Transparent Data Access for Data Analysis

The second use case deals with transferring datasets to external facilities (e.g. EOSC compute platform) to run data analysis. The requirements for the use case are:

- The data should be transferred transparently from the point of view of the users,
- Users should be able to transfer the result back to the originated RI,
- Data access and transfer should be authenticated via UmbrellaID AAI,
- The solution should guarantee access to both open and embargo data.

³⁶ <https://eduteams.org/>

³⁷ <https://geant.org/>

³⁸ <https://www.ukri.org/councils/stfc/>

³⁹ <https://rucio.cern.ch/>

⁴⁰ <https://fts.web.cern.ch/fts/>

⁴¹ <https://marketplace.eosc-portal.eu/services/scd-stfc-rucio-data-management-service>

⁴² <https://marketplace.eosc-portal.eu/services/egi-data-transfer>

⁴³ <https://rclone.org/>

The pilot setup for this use case and the initial results has been documented in the deliverable D6.1[2], where EGI has deployed a dedicated instance of EGI DataHub⁴⁴ for the project together with a data analysis service (JupyterHub) hosted at CESNET⁴⁵. EGI DataHub offers a way to expose the RIs data using a locally deployed component (Oneprovider) and implement custom mapping between local user accounts or credential on storage resources (e.g. POSIX user ID/group ID, LDAP DN, Ceph username, GlusterFS UID/GID, etc.) to AAI credentials.

The pilot demonstrated the access of data located at CERIC-ERIC from the Jupyterhub installation at CESNET-MCC, data processing and transferring of data back to the original RIs (see Figure 9), and everything was integrated with the UmbrellaID AAI.

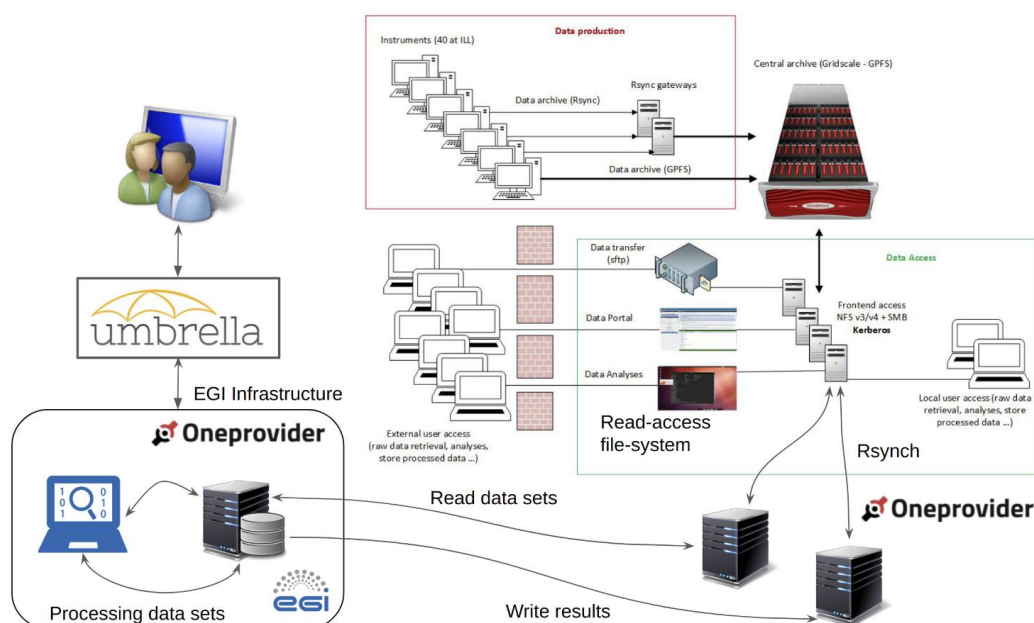


Figure 9 The DataHub PaNOSC Data transfer pilot

After this initial proof of concept, other RIs have started to plan the installation of the DataHub Oneprovider component which is needed in order to import the data into DataHub.

First ILL has tried the installation and configuration, which has been stopped for lack of effort. Then ESRF started the installation and found some technical difficulties. In particular the access to the storage at ESRF via NFS requires root access from the Oneprovider component in order to import data, which was not possible for security reasons. The EGI DataHub team has therefore developed new functionalities in order to implement a non root user data import, which has been successfully tested at ESRF. Unfortunately the installation has not been further exploited for testing and piloting due to effort spent in installing new versions of Oneprovider (not initially foreseen) and other parallel activities ESRF has onboarded, like the integration of Globus to support another data transfer use case.

⁴⁴ <https://marketplace.eosc-portal.eu/services/egi-datahub>

⁴⁵ <https://www.cesnet.cz/>

4.3.3. Data Transfer to Home PCs

In order to transfer data to home PCs there is the need to have a client software easily installable on the users side which can react to network instabilities and resume transfers of data after disconnections or simply PC restarts.

At the moment there is no service in the EOSC portal which offers this capability. On the other hand Globus⁴⁶ and Aspera IBM⁴⁷ offers solutions for users to transfer his/her data between endpoints which can be easily configured on home computers. The main outcome of the evaluation that have been carried out during the project is the following:

The Aspera subscription is very expensive and it has been considered very complex to be set-up.

Globus offers two different options: A free at the point of use solution with some drawbacks, and a commercial solution with additional services, offering also the possibility to purchase services just for some techniques/data.

After discussions with the Globus team in order to agree on some GDPR requirements, the Globus solution was piloted then at ESRF (which also acquired a subscription) and other facilities are already using the solution and found it very useful for their needs.

⁴⁶ <https://www.globus.org/>

⁴⁷ <https://www.ibm.com/products/aspera>

5. Difficulties

5.1. Issues with the onboarding of PaN resources in EOSC

The onboarding of new PaN services and resources in EOSC had some bugs and several iterations with the EOSC Portal Onboarding Team (EPOT) were necessary in order to solve the issues and publish new PaN service providers with their resources in the EOSC Portal. In particular the main issue was related to the upload of the service providers' privacy policy and terms of use which were considered optional in the previous release of the portal, but that were requested in order to complete with the registration workflow. These issues have been solved with the recent new release of the onboarding procedure and the update of the workflow. Historically many services provided by the communities and in particular the PaN Community, are collaborative efforts between different partners formalized in a Memorandum of Understanding (MoU) but without a specific legal entity formally responsible for the service. Registering these services in the EOSC portal under the umbrella of a single entity was causing two main difficulties:

In terms of communication, the registration of a single entity gives the feeling that this service is no longer part of the PaN community but that one entity is taking over under its name the development activity. This is detrimental to the collaboration spirit and does not reflect the efforts of the other partners. A partial answer was provided in the form of co-ownership by the EOSC Onboarding team.

Identifying a partner inside the PaN collaboration that could assume all the responsibilities of the service. As services are often relying on different sub parts managed by different partners, having a single entity that takes over all legal aspects instead of having each partner assuming their part of the responsibility is a real issue and a show-stopper.

5.2. Lack of data transfer service for service composability in EOSC

At the moment there is no service in the EOSC Portal which offers a data transfer solution which is suitable for the PaNOSC use cases. The various RIs part of PaNOSC have heterogeneous storage solutions and their integration with a solution for data transfer in EOSC which support access via the Umbrella AAI is not an easy task.

As mentioned in the previous section, the EGI DataHub solution has been tested to solve a particular use case, but it requires agreements on each facility in order to install the component needed to expose the data and still has to be tested in PaNOSC to scale with the number of users.

As well there is no suitable solution in EOSC to transfer data reliably from/to user home PCs (as for instance Globus provides). As far as the archiving use case, the services currently available in the EOSC portal (Rucio and FTS) have been tested but finally discarded in favor of a client only application (rclone). The future EOSC procurement will most probably tackle these issues and provide a general solution for data transfer in EOSC which could be beneficial from any Research Infrastructures (RIs)

5.3. Helpdesk and its sustainability

The EOSC Helpdesk comes quite late for the PaNOSC project, so there was not enough time to test the integration with the PaN services developed by the project. In addition, there is not a clear understanding of how the EOSC Helpdesk will be sustained as a part of the EOSC Procurement

6. Summary and Outlook

This deliverable described the plans to enable PaN RIs to integrate and provision services in EOSC and contribute to the EOSC Exchange. The various PaN providers registered in the EOSC Portal and the service onboarded at the end of the PaNOSC project have been documented together with the plan for new services to be onboarded.

The work around integration of PaNOSC services with the EOSC transversal/support services has been documented with a section describing as well the difficulties encountered by the project in the various activities.

From a technical point of view, thanks to the collaboration and support of the EOSC Future AAI Task-Force, the UmbrellaID is now part of the EOSC Federation AAI.

The Data Transfer solution is still an open issue in EOSC. Several solutions were investigated during the project's lifetime but an official solution was not identified yet. This issue will be followed in the framework of the EOSC Future project and in the next EOSC Procurement (Lot no. 3 - Managed Collaborative Data Platform Services for the EOSC Exchange)⁴⁸.

To continue the support of the PaN community integration with EOSC beyond the PaNOSC project, common work is being investigated together with other science cluster projects where large commonalities have been identified. The work with the e-infrastructure has also been largely beneficial and should continue, even be extended, beyond the scope of the PaNOSC project.

⁴⁸ <https://ted.europa.eu/udl?uri=TED:NOTICE:234660-2022:TEXT:EN:HTML&src=0>

7. References

- [1] Perrin, Jean-François (2022). D6.3 - Integration of the PaN AAI into the EOSC. DOI: 10.5281/zenodo.5913472
- [2] Perrin, Jean-François (2022). D6.1 - Data Hub https://www.panosc.eu/wp-content/uploads/2020/12/D6.1_DataHub.pdf