Organisation ESRF

Project Charter

PaNOSC

(Photon and Neutron Open Science Cloud)

Version 31 March 2025

Document History

- Version 10/3/2025 - first version submitted to EC + EOSC-A
- Version 31/3/2025 - reviewed following the Build-up Kickoff meeting; reduced number of use cases to 5; added milestones for Phase 2; included Dectris Cloud as service to onboard during Phase 2

1. Project Summary

The Photon and Neutron (PaN) community has been closely involved with the EOSC through the PaNOSC and ExPaNDS projects which ran from 2018 to 2023. These projects have already provided significant outcomes which boosted Open Science and data services for users. With the build-up of the EOSC Federation the ESRF will host the EOSC Node for the Photon and Neutron community. The PaNOSC Node will start off with 11 facilities, ESRF, HZDR, ESS, ILL, DESY, ALBA, ELETTRA, MAX IV Laboratory, European XFEL, PSI, ISIS, all of which are members of LEAPS[1] and LENS[2], to federate and provide data and services to the EOSC Federation. The LEAPS and LENS facilities have been collaborating together since over 20 years. The PaNOSC Node will onboard services from PaN partners providing cloud like services e.g. Dectris Cloud, to enable EOSC Users to access these cloud services more easily.

The outcomes of the PaNOSC EOSC Node will be higher scientific and societal impact of research data from the facilities. Linking the datasets from numerous experiments in multiple scientific domains performed at PaN facilities to other EOSC Federation data repositories will increase the quality and visibility of PaN data in general leading to more data being found and reused. Being part of the EOSC Federation will increase the adoption of Open Science practices in the research conducted at PaN facilities which will lead to better data management as well as more data being re-used by scientific communities.

The availability of high-quality data will enable AI algorithms to be trained for different purposes including analysing data and verifying results. The PaNOSC EOSC Node will ensure data become a permanent part of the LEAPS/LENS infrastructures and that unique processed data e.g. the Human Organ Atlas and Palaeontology repositories, will enable new scientific collaborations. The PaNOSC Node will eventually create a new class of Open Science users (and algorithms) producing new results as Open Data. The PaNOSC Node will promote the adoption of common tools and standards for metadata, software, and data repositories. The PaNOSC EOSC Node will reuse and link to services from the EU Node and other Nodes to make them available to the PaN community.

2. Value Proposition

- **Main Goal:** Provide FAIR data for a wide variety of scientific domains. Increase the scientific impact of PaN data and outcomes (results, training, publications etc.). Improve the quality of resources provided by PaNOSC by joining the EOSC Federation. Increase the adoption of FAIR in the PaN facilities. Learn from other Nodes by linking data between Nodes and aligning best practices with more mature Nodes.

[1] https://leaps-initiative.eu/
[2] https://lens-initiative.org/

- **Needs addressed:** make data FAIR with richer metadata to; provide more tools for managing, processing, and sharing data; link PaN data to *de-facto* standard community data sources; streamline access to PaN open data on the computing instances**;** provide alternative solutions for basic IT services required by scientists like AAI, file sharing, data transfer etc.
- **Key Benefits:** Provide quality FAIR data to the EOSC Federation; To increase the scientific output in all scientific domains covered by the PaN RIs; improve the quality of FAIR data from PaN RIs.
- **Unique capabilities**: Data availability from multidisciplinary research at Europe's large RIs
- **Who Benefits:** PaN Users (academic & proprietary), FAIR data Users, developers of AI/ML algorithms, society

3. USE CASE(S)

Use Case ID	***Use Case Description***	***Federation Contributions & Value to Users***
1 – ESRF + PSI + all	Align PaNOSC AAI with EOSC AAI so that PaNOSC AAI is compatible with the EOSC AAI and can accept EOSC Users	GEANT eduTEAMS + EOSC AAI Resources: umbrellaID, eduTEAM
2 - ALL	Integrate the PaNOSC data catalogues into the EOSC Federation so they can be searched for FAIR data and data can be downloaded and/or uploaded to other EOSC Nodes	Connect all PaN catalogues to the PaNOSC federated catalogue and EOSC Federation metadata search engines e.g. EUDAT Resources: https://leaps-wg3.desy.de/open-data-resources.html
2.1 – ESS + ESRF + all	Demonstrate finding curated FAIR data for a specific X-Ray or neutron technique, CryoEM or Sample. Use the OSCARS projects as test cases for searching for data i.e. surface diffraction data for AI-Scope, spectroscopy data for SHARE, materials science powder diffraction data for MatScatNet, human organs, palaeontology samples, etc.	EOSC Federation: PaNOSC federated search, PaNET data ontology for techniques Users: curated high quality data for reuse or AI/ML training
2.2 – ESRF + all	Link raw curated data to *de-facto* standard database in life sciences for X-Ray & neutron (PDB) and CryoEM (EMPIAR) for the tens of thousands protein structures from the PaNOSC sources, CXIDB for FELs	EOSC Federation: close collaboration with ELIXIR and LS Node database team EOSC Users: higher data impact through linking to de facto standard community databases; access to data-repositories, including raw and curated-experimental data and unified API
3 – ESRF + ILL	Provide access to EOSC Users to process open data on the PaNOSC VRE	EOSC Federation: provide the AAI for EOSC Users to login Resources: https://visa.esrf.fr https://visa.ill.fr EOSC Users: access to data and compute

4 - HZDR	Provide a common training platform with ELIXIR LS Node to collect high quality training material, training events and to describe relations between these training resources, involved analysis services and data products, as well as background information in overall descriptive training path workflows.	EOSC Federation: provide AAI for EOSC Users to login + common training material shared with ELIXIR LS Node + further communities EOSC Users: Common cross-disciplinary space for knowledge exchange and community building.

Use Case ID	***Use Case Description***	***Federation Contributions & Value to Users***
5 – ESRF + EuXFEL	A visualisation and verification service to accept FAIR, open and curated data that adheres to some community metadata standard e.g. like a service to validate NeXus like H5Web	Support harvesting metadata, to integrate datasets within various knowledge graphs. Support cross-walks to established metadata standards from other communities. Integrate H5Web with PaNOSC federated search. EOSC Users: visualise data in common data formats

In- Scope

Core Functions: 11 data catalogues, a federated search over the catalogues, a common training platform, AAI integration into the EOSC AAI, workflows for data processing, VISA VRE, NeXuS metadata standards, PaNET ontology, helpdesk, software packaging and distribution via CVMFS.

Stakeholder Support: 11 Research Institutes support the PaNOSC EOSC Node through Open Data Policies and providing FAIR data , support of the community initiatives LEAPS + LENS, RIs will provide Competence centres,

Integration: Federating data between partners and other EOSC Nodes, AAI federation, File sync&share federation, Bulk data transfer federation

Out of Scope

Excluded Activities: Data production, data curation

Limitations: CPU/GPU resources are limited to users of PaN Open Data

Dependencies: EOSC AAI, data repositories from scientifically related communities, resource catalog, options for data transfer, Sync&Share

3.1 Use Case 1 – Align and integrate PaNOSC AAI with EOSC AAI

Description: The fundamental build block of the EOSC Federation and all EOSC Nodes is a common AAI solution which is AARC BPA compliant. The PaNOSC EOSC Node will support the EOSC AAI.

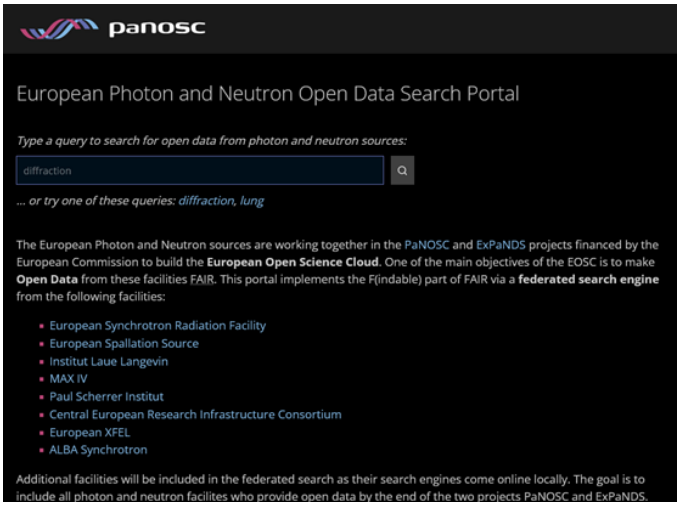
PaNOSC has developed a community AAI called UmbrellaID[3]. In 2020 to be compliant with AARC BPA, we have collaborated with GÉANT to transform the architecture of the UmbrellaID by setting up an eduTEAM proxy to be compliant with the AARC BPA. We are now working with GÉANT to adopt MyAccessID on top of our current architecture to be fully compliant with the ESOC AAI. Technically this evolution represents to major steps: 1) Migrate users legacy id to the ones of MyAccessID internal Ids and open the access from eduGAIN idps 2) Implement authorisation management at the level of the community using eduTEAMS services. The creation of the PaNOSC Node will be the occasion to finally move umbrellaID to MyAccessID and become compliant with the EOSC AAI.

Goal: use MyAccessID to be fully compliant with the EOSC AAI.

Needs: work closely with GÉANT to migrate our infrastructure to MyAccessID and work with the PaNOSC community to move to federated authorisation mechanisms to allow cross Organisations/Nodes services that request authorisations.

Multi-Node: PaNOSC (ESRF, PSI), EU Node (GÉANT) and all the others communities/Nodes that are adopting EOSC AAI.

3.2 Use Case 2 – Provide, Search & Find curated FAIR data for a specific X-Ray or neutron technique



https://data.panosc.eu

Description: The aim of the EOSC Federation is to provide a web of data. The PaN RIs together provide over 600 000 Open Datasets[4] which will be made available through the PaNOSC Node. The PaNOSC catalogues provide OAI-PMH endpoints and the majority implement the PaN Federated search endpoint. This foundational use case will address how scientific data are made available via the EOSC Federation. It will identify the Federating Capabilities required for research data and how these will be made Findable through examples for specific applications (see Use Case 3.2.1).

[3] https://umbrellaid.org/

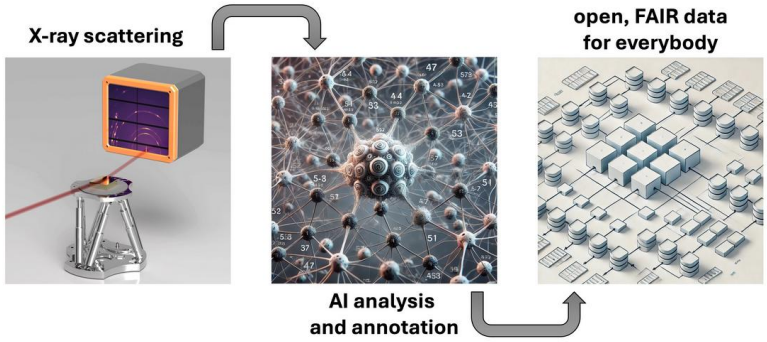
[4] https://leaps-wg3.desy.de/open-data-resources.html

Goal: the goal of the use case is to provide access to EOSC Users of Open Data.

Needs: EU Node Resource Catalogue, PaNET Ontology for techniques, datasets to be correctly labelled

Multi-Node: PaNOSC (all partners)

3.2.1 Use Case 2.1 – Search & Find curated FAIR data for a specific X-Ray or neutron technique



Description: Find data for specific applications e.g. e.g. OSCARS AI-Scope project[5] needs GIsaxs data to train their AI algorithm.

Goal: the goal of the use case is to provide access to EOSC Users of Open Data.

Needs: PaNET Ontology for techniques, datasets to be correctly labelled

Multi-Node: PaNOSC (all partners), University Graz

3.2.2 Use Case 2.2 – Link raw curated data to de-facto standard database in life sciences for photon & neutron sources

Description: Scientific communities conducting research at PaN facilities have been diligently curating scientific information databases and databanks for extended periods of time. The Protein Databank in Europe (PDBe) is collecting, curating and disseminating data on structure of biological molecules. The Electron Microscopy Data Bank (EMDB) is a public repository for data from cryogenic electron microscopy (cryoEM), an experimental technique that was awarded the Nobel prize in 2017 and is rapidly developing today. It was not possible to store original experimental data in the indexed files in the early times of x-ray crystallography or in personal computers later. The digital revolution, adoption of open science practices and related changes in scientific data management policies at PaN facilities allow original experimental datasets to be archived nowadays. The

[5] https://oscars-project.eu/projects/ai-scope-ai-driven-enhancement-surface-scattering-data-open-science-platforms-across

goal is to link this raw curated data to de-facto standard databases managed by LSRI Node for photon & neutron data.

Goal: Link raw data to entries in Protein Data Bank (PDBe), Electron Microscopy Data Bank (EMDB), (references: DOI: 10.1016/j.jbc.2021.100560)

Needs: Work with LSRI Node to identify APIs to link and verify

Multi-Node: PaNOSC (all partners) + LSRI node

3.3 Use Case 3 - Provide access to EOSC Users to process open data on the PaNOSC VREs

https://visa.esrf.fr

Description: The PaNOSC Virtual Research Environment (VRE) has been developed by the ILL during the PaNOSC project and deployed by the community RIs. It runs on OpenStack and is used by hundreds of users at the ILL (main site), ESRF (secondary site), EuXFEL, ESS, ALBA, SOLEIL. It provides interactive notebooks and Virtual Machines close to the (huge) data with various CPU+GPU configurations and scientific software suites pre-packaged and tailored for the PaN data analysis. The current VRE service is reserved for Users of the PaNOSC institutes.

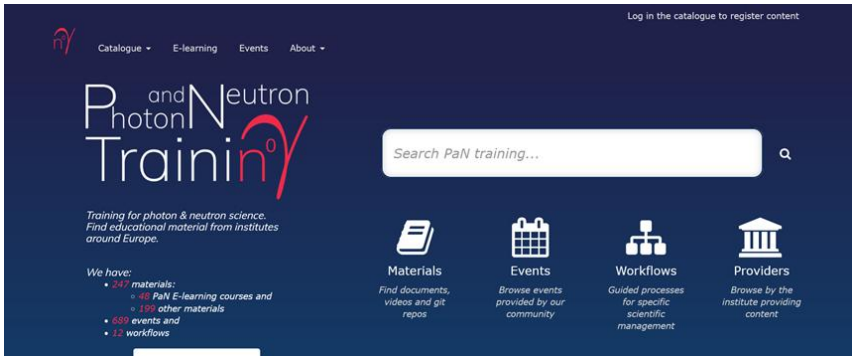
Additional data analysis services include Webknossos and Neuroglancer. These are used to annotate and segment large 3D data volumes from different scientific domains e.g. digital histology and connectomics, etc.

Goal: the goal of the use case is to provide access data analysis services to EOSC Users of Open Data.

Needs: integrate the PaNOSC facilities AAI to authenticate the EOSC AAI; extend the access rights to use the PaNOSC VRE to EOSC Users of Open Data.

Multi-Nodes: EOSC EU Node AAI + PaNOSC(ESRF)

3.4 Use Case 4 - Provide a common training platform with LSRI Node



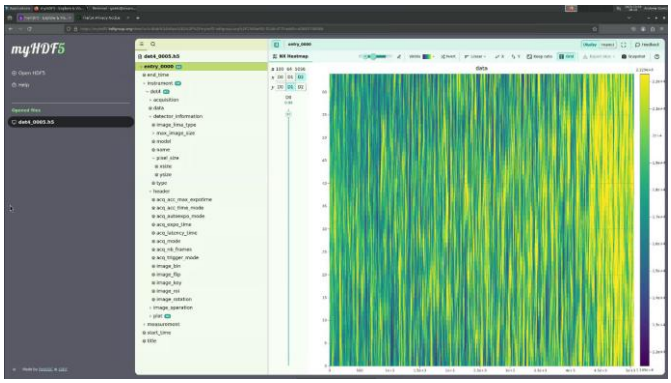
Description: The PaNOSC cluster has a common training catalogue developed during the PaNOSC project – https://pan-training.eu/. The training catalogue is based on Tess developed by Elixir. An OSCARS funded project, mTess-X, is working on federating the PaNOSC catalogue with the ELIXIR catalogue.

Goal: the goal of the use case is to provide access to the PaNOSC training platform to EOSC Users and to federate training catalogues.

Needs: Integrate EOSC AAI into PaNOSC training platform, federate PaNOSC training catalogues, PaNET Ontology for techniques.

Multi-Nodes: PaNOSC(HZDR) + EOSC EU Node + LSRI-Node

3.5 Use Case 5 – Data visualisation and verification service



https://h5web.panosc.eu

Description: Visualise and verify scientific data as a service. The PaNOSC community has developed the h5web visualisation tool which is widely used by tens of thousands of scientists from a wide variety of fields including ones beyond the PaNOSC community e.g. genomics. This service will be provided and extended to verify metadata for the NeXuS metadata standard. A future development will add AI tools to do quality control of large volumes of data.

Goal: provide a service for visualising data and verifying metadata

Needs: H5Web, H5Grove backend for Nodes wanting to integrate it.

Multi-Node: PaNOSC (all partners)

4. EXTERNAL DEPENDENCIES & KEY RISKS

- **External Dependencies:** EOSC AAI; LSRI ELIXIR Node Databases + Training platform; NFDI data repositories
- **Key Risks & Mitigation Measures:** EOSC AAI not readily implementable;

External Dependencies & Risks	Actions	Deadline
EOSC AAI	Align the PaN UmbrellaID with the EOSC AAI	September 2025
EU Node File Sync & Share	Link PaNOSC partners Nextcloud platform to the EU Node File Sync&Share	September 2025
LSRI Node	Link to PaN data repositories to ELIXIR life sciences repositories	November 2025
NFDI Node	Link PaN data repositories + training platforms to NFDI	November 2025

5. Contributions

ID	Deliverable Name	Deliverable Description	Deliverable Owner (the partner within the organisation responsible for producing the deliverable, if applicable)
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	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	ESRF, MAXIV, DESY,

	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	of the EOSC 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 services.	ALBA,
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.	ESRF, DESY, …
4	Operational quality plans, service reviews, verification and test session results and defects status and resolution Capacity Plan	All the operational related deliverables including testing, Quality Assurance Management, operations, maintenance and support including incident handling.	ALL PaNOSC partners

	Verification, Validation and Testing Plan Operations, Maintenance and Support Plan Incident Reporting Plan (including setup of ticketing system and workflows)		
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.	ESRF, DESY, EuXFEL, HZDR
6	Production roll-out of service components including web-service API's and associated documentation	The actual managed service components of the PaNOSC EOSC Node onboarded services and API's.	ESRF, HZDR, ILL, ESS
7	Datasets + Data repositories	ESRF, HZDR, ESS, ILL, ALBA, ELETTRA, MAXIV, EuXFEL, PSI, ISIS, DESY*	ALL PaNOSC partners
8	Software + Tools	VISA VRE, WEBKNOSSOS, EWOKS workflow engine, XRFitVis, Darkiver	ILL, ESRF, ELETTRA
9	Services	PaN H5Web, PaN Training platform, PaN federated search, PaN Software catalogue, PaN VISA VRE	ESRF, HZDR, ESS, ILL

6. Timing and Milestones

- **Start Date:** 17/03/2025
- **Expected Duration:** 2 years
- **Planning**: will be done in 2 Phases - 1 & 2

ID	Milestone Description for build-up Phase 1	Target Delivery Date
0	Specify requirements for PaNOSC Node partners	March 2025
1	Setup PaNOSC Node governance	April 2025
2	Setup PaNOSC Node website	May 2025
3	Onboard PaNOSC partners to PaNOSC node	July 2025
4	Connect PaNOSC to EOSC AAI; Federate PaNOSC data and services	September 2025
5	Setup Helpdesk and monitoring service	October 2025
6	Demonstrate Use Cases for PaNOSC Node	November 2025

Planning after November 2025 will be done based on the progress and results achieved during Phase 1 which runs from March 2025 to November 2025. There are still a lot of unknowns in the federating services offered by the EU Node due to lack of documentation.

ID	Milestone Description for build-up Phase 2	Target Delivery Date
0	Specify requirements for Onboarding commercial service(s)	February 2026
1	Onboard Dectris Cloud	March 2026
2	Onboard additional PaN clouds / services	May 2025
3	Review UX and improve as required	July 2025

7. **Contact & Submission**

Role	Name	Email
Coordinator	Andy Götz	andy.gotz@esrf.fr
Operation Manager	Oliver Knodel	o.knodel@hzdr.de
Security Officer	Jean-Francois Perrin	jean-françois.perrin@esrf.fr
Scientific Officer	Zdenek Matej	zdenek.matej@maxiv.lu.se
Privacy Officer	Jean-Francois Perrin	jean-françois.perrin@esrf.fr
Legal Officer	Renata Gibson	TBC