

PaNOSC Project brief on the EOSC ecosystem

Document	Title: PaNOSC's project brief on the EOSC ecosystem
	Available at: https://github.com/panosc-eu/panosc
Authorship	Written by: Andy Götz (ESRF), Jordi Boderà (ESRF), Rudolf Dimper (ESRF), , Ornela De Giacomo (CERIC-ERIC), Nicoletta Carboni (CERIC-ERIC), Fabio Dall'Antonia (EuXFEL), Jean-Francois Perrin (ESRF), Thomas Rod (ESS)
	Contributors: Jayesh Wagh (ESRF), Tobias Richter (ESS), Teodor Ivănoaica (ELI-ERIC). Erwan Le Gall (ILL)
Version	10 January 2022

Contents

Introduction.....	2
1. Integration with the EOSC infrastructure	2
Recommendations on follow-up priorities for the EOSC	4
2. FAIR principles implementation and repositories	4
Recommendations on follow-up priorities for the EOSC	5
3. Technical, semantic, legal and organisational interoperability.....	6
Recommendations on follow-up priorities for the EOSC	7
4. Stewardship of data	7
Recommendations on follow-up priorities for the EOSC	8
5. Cross-cluster collaboration activities and achievements	8
Recommendations on follow-up priorities for the EOSC	9
Conclusion.....	10

Introduction

The Photon and Neutron Open Science Cloud ([PaNOSC](#)) is one of the cluster projects financed by the H2020 INFRAEOSC-4 call to align the efforts of research infrastructures (RIs) in Europe to enable FAIR data and services to be contributed to build the EOSC.

PaNOSC is composed of 5 photon and neutron (PaN) RIs on the ESFRI roadmap and one ERIC. The partners are: European Synchrotron Radiation Facility (ESRF), Institut Laue-Langevin (ILL), European X-ray Free Electron Laser (EU.XFEL), Extreme Light Infrastructure ERIC (ELI-ERIC), The European Spallation Source ERIC (ESS) and the Central European Research Infrastructure Consortium (CERIC-ERIC). The partners are complemented by an e-infrastructure - EGI Foundation (EGI.eu) and GÉANT as service providers.

In addition to the PaNOSC partners, the national PaN RIs participate in the [ExPaNDS](#) project, the goal of which is to adopt and in some cases extend the PaNOSC outcomes at the PaN national RIs. The close collaboration between the two projects enables the PaN community to take advantage of synergies to better serve the PaN scientific community in Europe as a whole. The goal of the two projects is to have a common approach to data stewardship, FAIR data, and the EOSC, so that users of open data from all photon and neutron sources in Europe have a common experience and share common solutions.

1. Integration with the EOSC infrastructure

PaNOSC has a full work package (WP6) dedicated solely to EOSC integration, to engage with EOSC stakeholders to define the EOSC, contribute data and services to the EOSC catalogue and make use of relevant EOSC-wide services.

PaNOSC has been actively engaging with the EOSC and has submitted position papers¹, attended EOSC-related events, participated to consultations and WP leaders of PaNOSC have contributed to EOSC working groups and other EOSC-related projects and initiatives. Moreover, most of the PaNOSC RIs are members of the EOSC Association and are therefore committed to Open Science and to contribute to the implementation of the Partnership with the EC.

A major outcome of PaNOSC is to have ensured that 5 RIs on the ESFRI roadmap and an ERIC to make their data open and available to the EOSC. Two of the RIs (ESS and ELI) are

¹ PaNOSC position paper on the EOSC. Zenodo. <https://doi.org/10.5281/zenodo.3689420>
ESFRI Science Clusters Position Statement on Expectations and Long-Term Commitment in Open Science (1.02). Zenodo. <https://doi.org/10.5281/zenodo.4892245>

still in the process of construction and ramping up to user mode respectively. Therefore, they do not have significant amount of data available yet; however they are ready to publish Open Data from the beginning of User Operation thanks to their Open Data policies based on the PaNOSC policy framework and the implementation of data catalogues developed as part of PaNOSC.

PaNOSC has also allowed the partners to offer data analysis services by developing an Open Source data analysis portal VISA (<https://visa.readthedocs.io/en/latest/index.html>), which is being deployed at all PaNOSC and most ExPaNDS sites. VISA is ready to be deployed as a generic service at an e-infrastructure and/or other RIs as an EOSC service. VISA offers remote control of experiments, remote data analysis and simulation services, experimental set-ups, and analysis of data available via the data portals. VISA has been in operation at the ILL since 2020 and has been highly appreciated by hundreds of scientists during and in the future after the pandemic, as explained in this video interview with instrument scientists <https://youtu.be/w5DIN9nYsyA>).

PaNOSC has onboarded the following services in the EOSC Portal:

- [PaNOSC Software Catalogue](#): giving access to over a hundred standard software tools used for analysing data from PaN RIs;
- [Pan-learning.org](#): giving access to e-learning materials for PaN students and researchers, and plans to submit many more before the project ends;
- [Human Organ Atlas](#): an open data portal of 3D scans of human organs with micron resolution for different pathologies including COVID-19;

A number of other services including the 6 data portals of the PaNOSC partners and the search API service are in the process of being registered.

PaNOSC, in collaboration with ExPaNDS, has been developing and provisioning an EOSC training platform for the photon and neutron scattering community, available at <https://pan-training.eu>. It includes a catalogue developed in ExPaNDS which contains metadata and links to existing training material in the community and will be an entry point to [pan-learning.org](#) developed in PaNOSC. Pan-learning.org is an e-learning platform integrating moodle and Jupyter and provides support for using services developed in PaNOSC for training activities. It has been adapted and further developed from e-neutrons.org. It has been successfully used by a number of training sessions and workshops. It currently has more than 500 registered users, is registered as an EOSC service, and is accessible with UmbrellaID (see below).

PaNOSC, together with GÉANT, has put in operation a standard Authorisation and Authentication Infrastructure (AAI) common to all PaNOSC RIs and EOSC contributors to facilitate access to the different services. UmbrellaID² has now integrated the GÉANT's

² <https://umbrellaid.org>

EDUTeams as a proxy platform and is in line with the EOSC AAI architecture and operation requirements. It will soon allow the authentication of users through both the community ID provided by the RI, ORCID and the users' affiliation Authentication infrastructures (EduGAIN). All the PaNOSC partners have now integrated UmbrellaID to authenticate users for their data portal and proposal submission; community services, like the software catalogue³, benefit from these developments at the same time.

While work is ongoing and PaNOSC partners hope that the EOSC will end up being the entry to federated services and Open Data for scientists and citizens, several challenges remain to make the integration of PaNOSC services in the EOSC a success. The only EOSC infrastructure visible to users today is the [EOSC portal](#), however, the portal has certain limitations today:

- The EOSC Portal is for the time being limited to being an aggregator of data and services hosted and delivered elsewhere, adding very few benefits to the scientific communities that know very well the RIs. As such, it is strongly dependent on the sustainability of those services and data providers.
- The EOSC Portal registration of services remains a relatively heavy administrative process for providers, while the advantage to such providers is not clear.
- Currently the EOSC portal does not offer infrastructure to help address the sustainability problems that the provision of FAIR data and related services are generating by increasing IT infrastructure and human resources costs for the RIs. The incentives (financial or other) for RIs to provide these services beyond their own communities need to be clarified.

Recommendations on follow-up priorities for the EOSC

1. Provide a common solution for AAI for all EOSC services;
2. Improve the portal for enhanced user experience and make it a platform which provides added value for providers and users;
3. Make service registration (onboarding) more streamlined and provide more support for communities like PaNOSC.

2. FAIR principles implementation and repositories

One of the main objectives of PaNOSC (and the EOSC) is to **"make FAIR data a reality"**. Therefore, the associated European funding that has been provided to all its partners has been instrumental in increasing the priority of implementing the FAIR principles at the PaNOSC partners. Many activities related to FAIR have been started within PaNOSC:

- Development of a FAIR Data Policy Framework⁴;

³ <https://software.pan-data.eu/>

⁴ PaNOSC FAIR Research Data Policy framework (1.1). Zenodo.

- Update of the existing data policies at 4 ESFRI RIs to enable FAIR data;
- Introduction of FAIR data policies at 1 ESFRI RI and 1 ERIC that did not have one yet;
- Scoring and certification of dataset FAIRness (following work done by ExPaNDS);
- Development of a federated search API to enable easier access to open data;
- Introduction of standard AAI for access to data portals;
- 6 OAI-PMH endpoints registered with OpenAIRE;
- Increased discoverability of PaN metadata in the EOSC ecosystem through B2FIND;
- General adoption of the community metadata standard Nexus e.g. all partners have now joined the advisory committee of NeXus;
- Collaborated with FAIRsFAIR to CTS certify one of the PaNOSC data repositories.
- Close cooperation with scientific journals and learned societies e.g. the International Union of Crystallography is mandating for FAIR data to be mandatory in all its publications (cf. <https://doi.org/10.1107/S2052252521012999>).

PaNOSC is contributing a semi-automatic processing tool in Python for a new journal, the **IUCr raw data letters** which will be launched in 2022 for publishing raw data references with crystallographic metadata. The tool will facilitate the collection and conversion of metadata information, for easy validation and embedding into manuscripts.

The 6 data repositories from the PaNOSC partners are being registered in the EOSC as part of the PaNOSC services. Together they provide over 60 PBs of data which will become open after the embargo period. The data are made up of over 1 million open datasets, 7762 data DOIs and 857 DOIs cited in publications. The medium term goal is to setup a PaN Data Commons where all Open Data from all PaN RIs in Europe can be searched and downloaded, and to train scientists to find, re-use and cite data DOIs in publications. PaNOSC has already helped create domain-specific data portals, e.g. the [Human Organ Atlas](#). The Human Organ Atlas is a unique resource of Open Data for human organs with a resolution down to the cellular level for a wide range of pathologies.

Recommendations on follow-up priorities for the EOSC

The EOSC plays a strong role in promoting Open Science and the concept of FAIR data. Nonetheless the definition of the work involved in ensuring data is FAIR depends on the scientific community, and needs to be done by scientists who understand the domain. Recommendations where the EOSC could help:

1. Co-creation and co-financing of FAIR data scientists. These persons would be members of the scientific community hired by the clusters, and would help train scientists on producing FAIR data. They would produce guidelines and tools for journals and scientific communities to promote data curation for FAIR data.
2. EOSC could extend existing search engines to include domain-specific data searching,

<https://doi.org/10.5281/zenodo.3862701>

or co-finance communities who provide domain-specific search facilities, e.g. PaNOSC.

3. Funding to organise events to raise awareness of FAIR, aimed at researchers based at universities and research centres that use RIs, including the benefits of making their own data FAIR.

3. Technical, semantic, legal and organisational interoperability

PaNOSC and ExPaNDS partners have worked together on policy issues including:

- Development of a new FAIR Data Policy Framework;
- Adoption of the FAIRness assessment developed by RDA;
- Definition and implementation of Data Management Plans;
- Produced guidelines on open data licences based on the Creative Commons licences;
- Adopted Open Source Software licences for all software developed by PaNOSC;
- Re-used community solutions for data catalogues (ICAT + SciCat).

PaNOSC partners have participated in the EOSC association, as well as in the EOSC Working Groups, and will continue to do so until the end of the project and beyond. PaNOSC participates in national Open Science projects, e.g. "Science Ouverte" in France, by taking part in, and organising workshops to showcase scientific use cases, which profit from the services developed by PaNOSC. PaNOSC has been sharing outcomes with, and has been participating in national projects, such as DaphneNFDI and the Helmholtz Metadata Initiative.

ExPaNDS and PaNOSC are an opportunity for the partner RIs to enhance the interoperability of their data, adopting common community standards, such as NEXUS and HDF5. This has allowed the common deployment of tools for simulation and data analysis. However, considering the timeframe of these projects and the resources involved, it was not possible to address an extension of the interoperability to other communities that may benefit from PaN facilities' data.

An important activity in PaNOSC is the collection of a significant number of scientific use cases, and addressing them to ensure that the PaNOSC services address real needs of scientists⁵.

The PaNOSC policy framework allows partners to adapt to national rules. One of the partners, had to be compliant with the legislation of 8 countries wrt to the storage of data. Solutions developed in PaNOSC, benefitting from the OCRE procurement framework, needed to handle data management in agreement with national policies.

⁵ <https://www.panosc.eu/all-use-cases/>

Recommendations on follow-up priorities for the EOSC

EOSC does not currently have close contact with users. Bringing researchers to EOSC is one of the roles of the Science Clusters in general and in the EOSC Future project specifically. Without scientific use cases there is a very high risk the EOSC will not provide useful services for the scientific community. Creating a common environment for the scientific community is another. We therefore see the following areas which need addressing:

1. The EOSC should urgently create a database of scientific use cases for the long tail of science, and the solutions being developed for them in the EOSC.
2. The different services in the EOSC need to be streamlined so they are coherent and easy to use by different communities. The services must include training, data access and transfer, AAI, access to compute resources.
3. The technical interoperability requirements of the EOSC will evolve during its lifetime and require that service providers update their services accordingly (e.g., mandatory metadata for services). This needs sustainable funding for human resources.
4. A number of countries have not yet implemented a data strategy, EOSC data and service providers would benefit from an analysis and overview of the landscape updated periodically.

4. Stewardship of data

Before PaNOSC and ExPaNDS, the PaN community RIs had a wide variety of maturity levels with regards to data stewardship. Some of the RIs had adopted best practises as early as 2012, whereas others were still considering how to implement the best practises. One of the principal objectives of the two EOSC projects has been to bring the majority of the photon and neutron RIs to adopt and implement data stewardship in line with the EOSC advocated FAIR principles. To achieve this goal, the following 10 practical outcomes are being developed by PaNOSC and ExPaNDS:

1. FAIR data policy and DMPs
2. FAIR assessment and common PID framework
3. Standardised metadata (Nexus/HDF5, PaN ontologies)
4. Federated search API for PaN data catalogues
5. Open Data portal for searching + downloading data
6. Community AAI UmbrellaId
7. JupyterLab notebooks and Nexus/HDF5 files visualisation
8. Remote data analysis with VISA + data analysis pipelines
9. Simulation software for simulating experimental data (SIMEX)
10. PaN-learning platform (pan-learning.org + pan-training.eu)

A majority of PaN RIs (e.g., 11 of the 15 photon sources who are members of LEAPS⁶) have committed to implementing a majority of the 10 outcomes listed above. This is foreseen to happen by the end of PaNOSC and ExPaNDS. A working group has been set up to monitor and assist RIs to achieve this goal. Another very important goal is the creation of positions for and hiring of data stewards at each RI. This role does not currently exist at the majority of PaN RIs but through PaNOSC and ExPaNDS the PaN RIs are now well aware of the need to hire and train data stewards.

The medium to long term goal of the PaN community is to provide a common data portal of all open data from PaN RIs which can be searchable and downloadable for verification and reuse. This goal is supported by the LEAPS and LENS⁷ organisations (representing all photon and neutron RIs respectively in Europe). PaNOSC has started this work (based on the search API and data portal developed in PaNOSC) but counts on LEAPS+LENS for long term sustainability.

Recommendations on follow-up priorities for the EOSC

1. Provide support for data stewardship through common guidelines, metrics, schools and training
2. Endorse common metrics/certifications for data stewardship and provide funding towards achieving them
3. Provide technical and financial support to help maintain common data portals for different EOSC scientific communities

5. Cross-cluster collaboration activities and achievements

PaNOSC, together with its sister project ExPaNDS, has been closely interacting with the other Science Clusters to align their activities within the projects as far as possible. The Science Cluster coordination became even more important during the EOSC Future proposal stage and now during the execution phase. Several common position papers have been published and PaNOSC, together with the other Science Clusters, has actively contributed to numerous

⁶ LEAPS (<https://leaps-initiative.eu/>), the League of European Accelerator based-Photon Sources, is a strategic Consortium whose primary goal is to promote and ensure the quality and impact of fundamental, applied and industrial research carried out at Synchrotron Radiation and Free Electron Laser (FEL) facilities in Europe.

⁷ LENS (<https://lens-initiative.org/>), the League of advanced European Neutron Sources, is a not-for-profit consortium working to promote cooperation between European-level neutron infrastructure providers.

events showcasing the project outcomes and promoting future directions of the EOSC.

Jointly with the other ESFRI Science Clusters and European e-infrastructures, PaNOSC provided feedback to the [Strategic Research and Innovation Agenda \(SRIA\)](#). Furthermore, the project contributed to two cluster position papers:

- [ESFRI cluster projects - Position papers on expectations and planned contributions to the EOSC](#), which was released in February 2020, and
- [ESFRI Science Clusters position statement on expectations and long-term commitment in Open Science](#), released in June 2021 and further discussed at the event "ESFRI Science Clusters' Long-Term Commitments to Open Science" held in the same month.

Despite the restrictions placed on travel for the last 2 years due to the pandemic, a number of inter-cluster actions around technical solutions have been started. Examples of cross cluster technical activities are (1) adoption of the TeSS training portal developed by ELIXIR (EOSC-Life), (2) sharing common technical solutions like Jupyterlab with SSHOC, (3) adoption of the HDF5 web viewer in Jupyterlab developed by PaNOSC by ESCAPE, (4) sharing of technical know-how on the data lake with ESCAPE, (5) cross linking of raw data in the Protein Data Bank (PDB) between EOSC-Life and PaNOSC. These action show the potential of cross cluster technical collaboration, we foresee more collaborations once travel is permitted.

In the fortnightly Science Cluster coordination meetings the project coordinators and other key persons exchange on the progress of the EOSC Future project and EOSC in general, information on strategic developments and initiate targeted technical meetings focussing on key technologies. The next workshop will be at the end of January and involve technical representatives of the project CS3MESH, ESCAPE, and PaNOSC to discuss synergies for Jupyter Notebook services, CVMFS, and EFSS.

Recommendations on follow-up priorities for the EOSC

1. Continue supporting the Science Clusters as providers of data and services for EOSC as a key stakeholder and interface to a large portion of the scientific user community in Europe.
2. Collaborate with the Science Clusters to coordinate actions to achieve increased interoperability of services and resources in the EOSC.
3. Address the needs of the science use cases by providing and operating core services like AAI, data transfer, compute resources.

Conclusion

To conclude, the PaNOSC cluster project has made major steps towards implementing the goals of the EOSC at the PaN ESFRI RIs. Together with the ExPaNDS project these outcomes are being extended to all national PaN RIs. The close collaboration with the other science clusters since the beginning of the EOSC has been essential in advancing towards the EOSC goals. The future of the clusters as part of the EOSC will however depend strongly on the added value of the core services of the EOSC. PaNOSC has made significant steps to bringing the EOSC to the PaN scientific communities and vice versa. The EOSC core services now need to improve in order for user communities like the PaN community to see the added value of the EOSC. This paper proposes a number of recommendations to work on to ensure the added value of the EOSC for the PaN community.