

# PaNOSC position paper on the EOSC

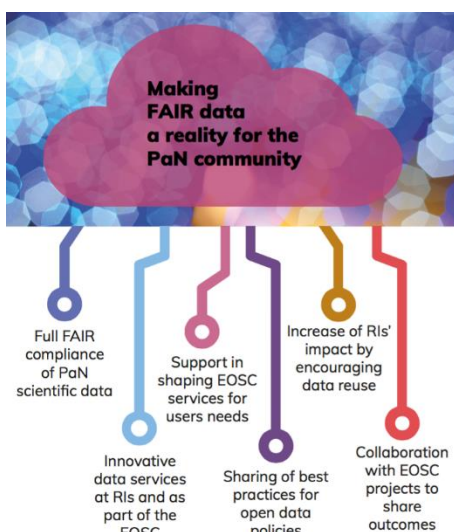
endorsed by the PaNOSC Executive and Management Board

Version 1.0 on the 22 November 2019

## Introduction

PaNOSC is an INFRAEOSC-04 cluster project of 6 European Photon and Neutron sources on the ESFRI roadmap [1]. The science done at the research institutes represents a huge variety of scientific disciplines using photon and neutron sources to study almost any kind of material on a wide range of length scales – from angstroms to microns – and time scales. The ESFRI and national sources (an additional 15 RIs involved in the ExPaNDS project [2]) represent a large user community of roughly 30 000 scientists annually who use the Photon and Neutron sources in Europe for their research projects. The science carried out at these institutes addresses at least five out of the seven societal challenges defined by the EC [3] which range from climate change, renewable and efficient energy sources, to drug discovery.

## PaNOSC Objectives



- *Participate in the construction of the EOSC by linking with the e-infrastructures and other ESFRI clusters. **Concretely** PaNOSC is working with EGI and GÉANT, participating in the Architecture Working Group and EOSC meetings. PaNOSC is developing a shared vision with the other INFRAEOSC-04 cluster projects to identify common areas of collaboration.*
- *Make scientific data produced at Europe's major Photon and Neutron sources fully compatible with the FAIR principles. **Concretely** PaNOSC is updating the PaNdata data policy framework to be FAIR compliant and implementing Data Management Plans (DMPs).*
- *Generalise the adoption of open data policies, standard metadata and data stewardship from 15 photon and neutron RIs and physics institutes across Europe. **Concretely** PaNOSC partners are updating/adopting their data policies.*
- *Provide innovative data and simulation services to the users of these facilities locally and the scientific community at large via the European Open Science Cloud (EOSC). **Concretely** PaNOSC is making it possible to run bespoke services for the PaN community from Jupyter or through remote desktops and providing Jupyter and remote desktops as a service as part of their local infrastructure close to the data and on the e-infrastructures.*
- *Increase the impact of RIs by ensuring data from user experiments can be used beyond the initial scope. **Concretely** PaNOSC partners are working on providing richer metadata by extending the photon and neutron community metadata standard Nexus [4] and adopting electronic logbooks for capturing user experiments. The goal is to enable open data to be used*

by the scientific community at large. PaNOSC will also provide training for scientists to understand FAIR and adopt FAIR practices for their data.

- *Provide training in the use of the involved facilities, services developed in PaNOSC, and data stewardship as an important tool for dissemination.* **Concretely** PaNOSC partners are extending the scope of the e-learning platform [e-neutrons.org](https://e-neutrons.org) to cover the domains represented by all the PaNOSC partners (i.e. including the photon sources), and developing courses and training material for that platform. PaNOSC will hold workshops and a summer school.
- *Share the outcomes with the national RIs who are observers in the PaNOSC project, the community at large and the EOSC, to promote the adoption of FAIR data principles and data stewardship.* **Concretely** PaNOSC is collaborating closely with the ExPaNDS project which is implementing FAIR data for the national photon and neutron RIs. PaNOSC is providing training and engaging with users at User Meetings to explain and train.

## EOSC Minimum Viable Ecosystem

PaNOSC sees EOSC as an opportunity to generalise the adoption of FAIR data practices at the 6 Photon and Neutron research facility partner institutes and eventually at all photon and neutron sources. Adopting FAIR data will enable data sharing across a wider community and the provisioning of services for remote data analysis. In order for these objectives to be realised the **EOSC must provide** the following services:

1. A common way of identifying, authenticating, and authorising users (**AAI**) across Europe. The EOSC should operate and sustain AAI as part of the EOSC infrastructure. The EOSC AAI should support the AAI features PaNOSC is implementing on the UmbrellaId AAI [6].
2. A service for **transferring and downloading data efficiently** (distributed and high bandwidth);
3. A solution for **long-term archiving of large quantities of open data** (petabytes) coupled to high-performance storage and compute resources for the (re)analysis of open data;
4. A **federated search capability** for searching and finding scientific data in a wide variety of domains;
5. A set of **services for data simulation and analysis** ranging from generic services like Jupyter notebooks to domain specific applications per scientific application in the PaN software catalogue [5]. These data services could be remote from the data source if the data are easy to move but should be available close to the data if the data are difficult to move.

The above services are considered as the **Minimum Viable Ecosystem** for the EOSC from the PaNOSC point of view. Only if the EOSC provided these services it would encourage and help many medium and small institutes, as well as individual scientific groups to make FAIR data available, to re-use FAIR data for new scientific findings, but as well to make appropriate use of the EOSC.

Once the implementation of FAIR data is standard practice, then it would be desirable for the EOSC to be extended to be more than a source of FAIR data. The EOSC could become the **GitHub of Open Science in Europe**. This means making it a platform for scientists to share their data analysis and workflows and link these back to open data and other workflows – either their own, or that of other scientists. To achieve this, it will be necessary to provide scientists with a personal space where they can create content (data analysis

recipes, workflows, publications), store analysed data and share their work with collaborators via a versioning system like git.

## PaNOSC provides

The PaNOSC Research Institutes will be an essential part of the EOSC as sources of data and providers of data services.

The PaNOSC partners aim to provide:

1. **Petabytes** of raw and processed data in a wide variety of scientific domains
2. **Meta-data** that will create **FAIR** raw and processed scientific data
3. **Software** for generic and specific data simulation and data analysis
4. **Workflows and expertise** for reducing and analysing data
5. **Reference training material** and **training platform** for understanding photon and neutron science and associated handling of data
6. Liaison with large **user communities** of photon and neutron sources and their expectations for services

A summary of the PaNOSC work packages can be found in the publication [6] and on the website [4].

## Feedback to the EOSC Executive Board

PaNOSC is a bottom-up approach to making data FAIR and making FAIR data sustainable for the Photon and Neutron community to help users do better science, be more efficient with the help of better data management and to make science more reproducible. The EOSC Executive Board can play an important role in this by bringing in a top-down approach in the following areas:

- provide a clear concise answer to the question “what is the EOSC?”. This should be incorporated in the architecture being developed by the EOSC working group dedicated to this
- define common standards for FAIR data so that the different scientific fields have a common approach and understanding, e.g., FAIRsFAIR could provide clear guidelines with examples on how to implement FAIR by different communities
- provide long-term sustainable plan for how the EOSC will be maintained and financed
- provide cloud resources for running data analysis workflows and simulations, ideally unlimited but at least enough to make a significant difference for users needing access to computing resources beyond what can be offered by the PaNOSC partners
- collaborate with publishers to generalise the requirement for citing data in publications and making open data a publication in its own right
- provide documentation and training material on how to join and use the EOSC
- do a cost-benefit analysis of what the EOSC provides to the Photon and Neutron communities and comparing the benefits with the cost thereof.

## Example data

Some of the PaNOSC RIs have databases of data collected over the last decades that are currently under-exploited e.g. paleontology data in the <http://paleo.esrf.fr> is an example of processed data which are not widely known or exploited yet. These data are ideal for cross-disciplinary applications and linking up with data from museums and other scientific disciplines. PaNOSC will provide raw and processed data with metadata of a far higher quality. In turn, this will allow opportunities for useful data mining to take place, to the benefit of the community and its wider stakeholders. The EOSC will offer an excellent opportunity to make such data more widely known and used by different communities.



*Picture 1: Synchrotron tomographic data (from the ESRF) used for the study of the Egyptian crocodile mummy 90001591 to establish the cause of death, its age at death, and its diet, demonstrating that it was a wild animal hunted to make a mummy. Data from original publication [7] can be downloaded from [8].*

Other examples of open data are the ILL and ESRF data portals, respectively <https://data.ill.fr> and <https://data.esrf.fr>.

## References

- [1] ExPaNDS project – <https://expands.eu>
- [2] <https://ec.europa.eu/programmes/horizon2020/en/h2020-section/societal-challenges>
- [3] Nexus metadata format - <https://www.nexusformat.org/>
- [4] PaNOSC website – <https://panosc.eu>
- [5] PaN software catalogue - <https://software.pan-data.eu/>
- [6] “Enabling Open Science for Photon and Neutron sources” by A. Götz et. al., ICALEPCS 2019 Pre-*Proceedings*, <http://icalepcs2019.vrws.de/papers/tubpl02.pdf>
- [7] Porcier S. M., Berruyer C., Pasqali S., Ikram S., Berthet D., Tafforeau P. « **Wild crocodiles hunted to make mummies in Roman Egypt: Evidence from synchrotron imaging** ». *Journal of Archaeological Science*, 1 October 2019. Vol. 110, p. 105009. DOI : <https://doi.org/10.1016/j.jas.2019.105009>
- [8] <http://paleo.esrf.fr/index.php?/category/2846>