

## Overview of PaNOSC key achievements in the 1<sup>st</sup> implementation period (18 months)

Coordinator:



Partners:



EUROPEAN  
SPALLATION  
SOURCE



**PROJECT  
COORDINATOR**

Andy Götz (ESRF)

**PROJECT NUMBER**

823852

**WEBSITE**

[www.panosc.eu](http://www.panosc.eu)



PaNOSC has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852

PaNOSC is a European project financed by the INFRAEOSC-04 call for making FAIR data a reality in six European Research Infrastructures (RIs) for photon and neutron science, developing and providing services for scientific data, and connecting these to the European Open Science Cloud (EOSC).

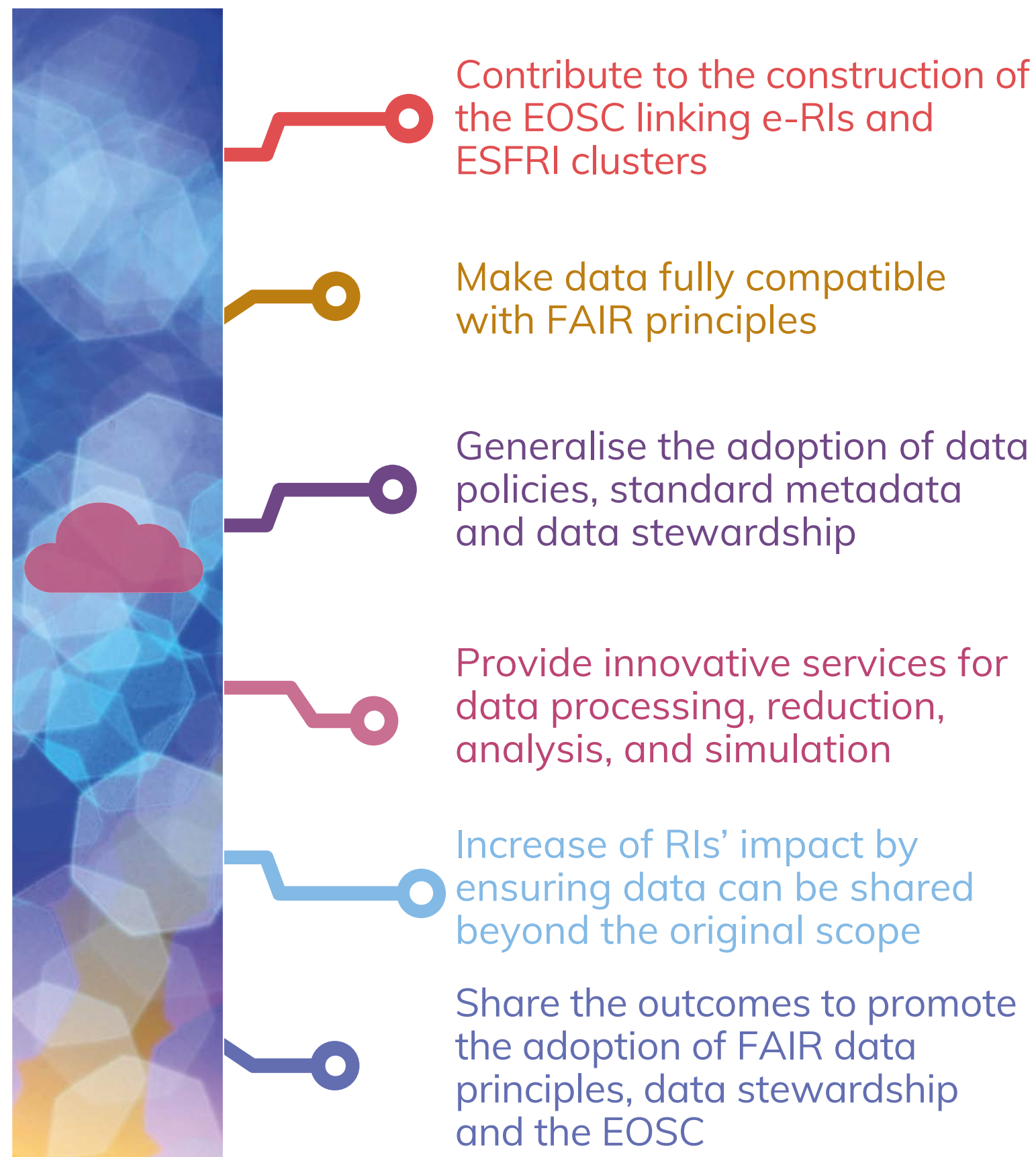
## Background and Scope

Photon and Neutron (PaN) facilities are essential research infrastructures for the understanding of matter and its properties. Together these facilities produce petabytes of data, which can give us a more complete picture of the world around us.

PaNOSC (and its sister project, ExPaNDS) will make the data produced easily accessible to the users and the public, by providing scientific data management for enabling Open Science.

Data will be managed according to the FAIR principles. This means data will be curated and made available under an Open Data policy, and be findable, interoperable and reusable.

## Main Goals and Objectives



## Overall Concept

The overall approach implemented implies to make the data available, as well as the data analysis software, contributing to Reproducible Science and FAIR data, and increase the ability to find and inspect the data interactively.

Based on this approach, PaNOSC will provide:

- **Data analysis services to complement data and metadata;**
- **Open Source Software to share data analysis knowledge and expertise;**

The analysis software that has been used on the data set, and the expertise required to create it, is made available.

- **Inform and train scientists;**

To include software that has created the results shown in the manuscript as part of the publication, and include a description of how the software needs to be executed to compute the results as part of the publication.

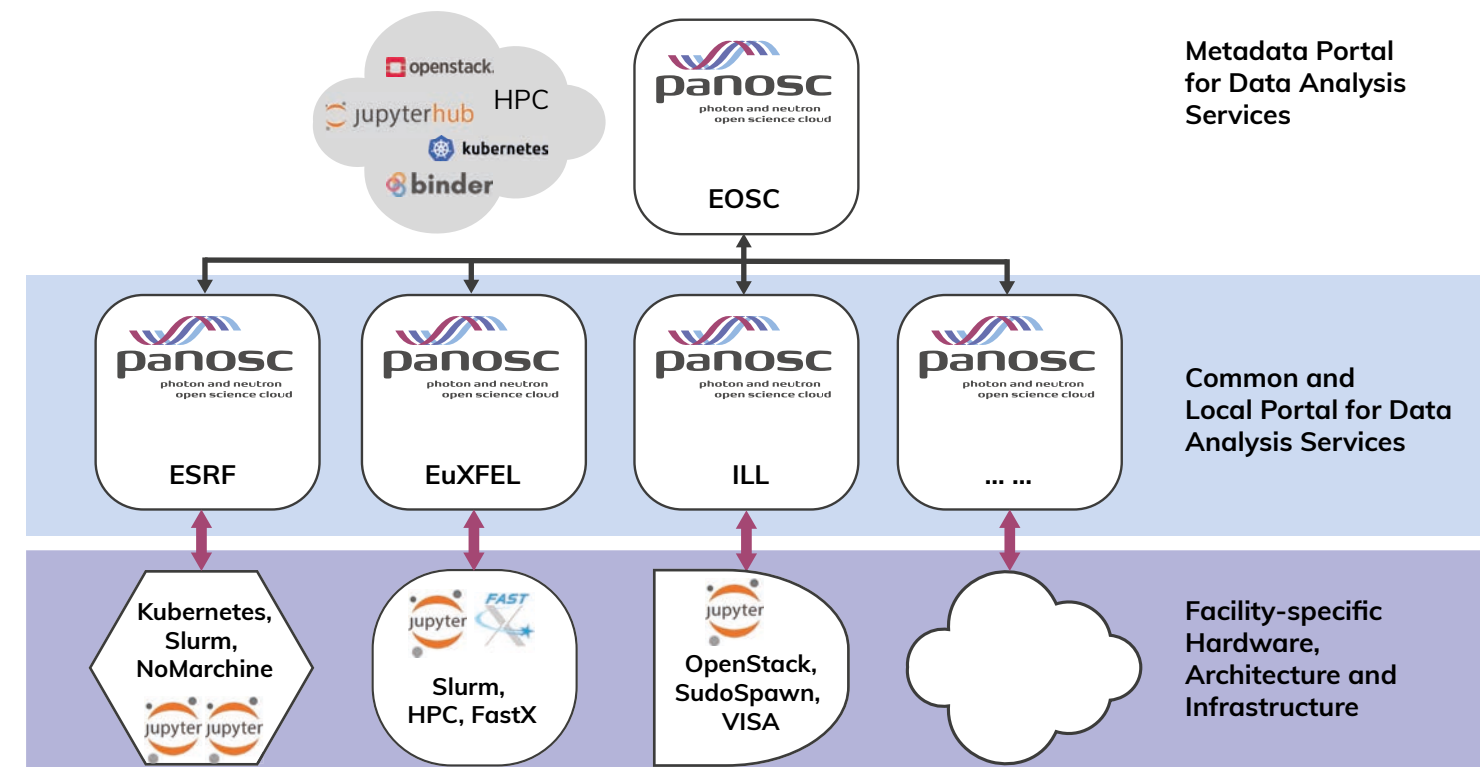


## Strategy

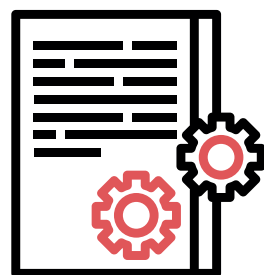
PaNOSC aims to create an analysis environment with analysis software available through a **data search portal** and **data analysis portal** connected to the facility-specific services, such as authentication, metadata catalogues, file location information and remote analysis services.

**Remote data analysis** will be possible via the data analysis portal with Jupyter Notebook, or remote Desktop technologies.

The **user experience** should be similar at all facilities. Data sources and services that are used as back-end to this portal will be facility-specific.







## Data Policy and Stewardship

WP leader: Andy Götz

One of the main objectives of PaNOSC is to make FAIR data a reality for research data produced by the photon and neutron facilities involved in PaNOSC and its sister project ExPaNDS.

To this aim, PaNOSC updated the **PaNdata data policy framework to be FAIR<sup>1</sup>**. The framework will be adopted by all partners to ensure they have FAIR data policies in place.

The work towards this achievement included the following steps:

- Share the lessons learned on adopting and implementing data policies;
- Analyse and compare existing data policies at the partners;
- Update the PaNdata data policy framework to comply with FAIR principles;
- Engage in discussions with scientists and management at the facilities to adopt the FAIR data policies;
- Adopt or adapt existing data policies to be compatible with the PaNOSC data policy.

The updated framework has been then compared with the RDA FAIR Data Maturity Model (FDMM)<sup>2</sup> to evaluate the level of FAIRness.

<sup>1</sup>PaNOSC FAIR Research Data Policy Framework: <https://doi.org/10.5281/zenodo.3862701>

<sup>2</sup>FDMM: <https://doi.org/10.15497/RDA00045>



## Data Catalogue Service

WP leader: Tobias Richter

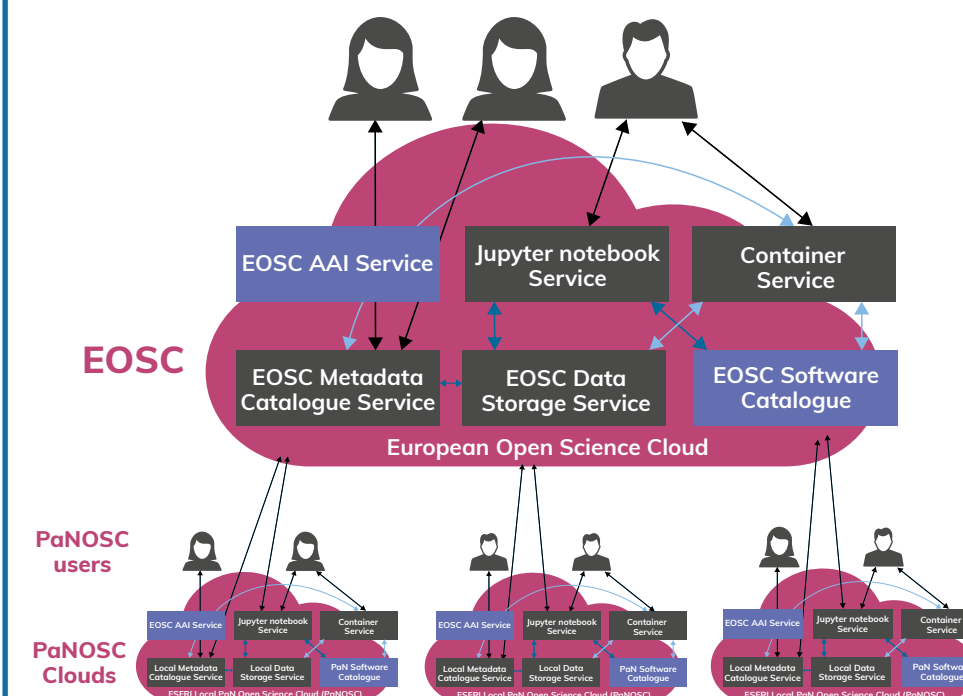
To make data **Findable** and **Accessible**, a **search API** has been defined and developed, to enable domain-specific searches across the PaNOSC data repositories.

All sites have also implemented the **OAI-PMH protocol** for indexing metadata and data by OpenAIRE/re3data and B2Find.

To make data **Re-usable**:

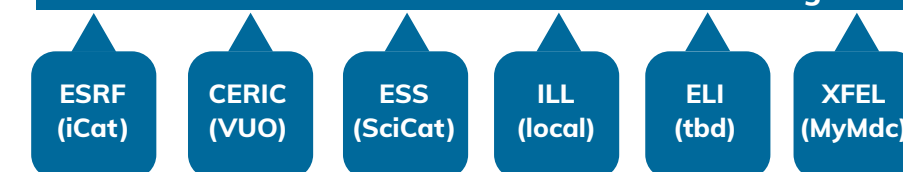
- Metadata harvesting endpoints have been deployed at all partners;
- **NeXus** has been promoted as community metadata standard for PaN sources;
- Electronic logbooks have been developed to capture what happens during experiments at few sites.

## EOSC + PaNOSC Clouds



## Common API for 6 catalogues

Common API to search across all PaNOSC catalogues



## Ongoing Activities

- Development of search API;
- Development of PaN specific metadata schemes and search dictionaries.



## Data Analysis Services

WP leader: Hans Fangohr

PaNOSC has been developing the **Common Portal for Data Analysis Services** to facilitate starting a data analysis session after a dataset of interest has been collected.

The Portal aims to provide access to both **remote desktop environments** and **Jupyter Notebooks**, enabling users to **remotely analyse data** from PaN facilities.

### Milestones reached

- Existing data analysis requirements and solutions from all partner sites (including ExPaNDS) have been surveyed;<sup>3,4</sup>
- All sites now provide remote desktop analysis services or remote Jupyter Notebook analysis services in a variety of states** (some in production with large user numbers);
- Provision of a citizen science prototype environment for remote and reproducible data analysis of COVID 19 infection data **OSCOVIDA** (<https://oscovida.github.io>).

### Ongoing Activities

- Developing standard data analysis notebooks for specific techniques;
- Providing tools used in the Notebook-based data analysis at the facilities, and contributions to open source data analysis tools that are used in PaNOSC and elsewhere, specifically, h5py, h5glance, hdf5plugin;
- Developing web-based viewers for HDF5 files: h5nuvola and h5web;
- Providing an infrastructure (e.g., JupyterHub, or Jupyter-Slurm) so that notebooks can be executed remotely on the computing and data infrastructure of the facility;
- Exploring the use of software packaging managers to deploy versioned software at HPC installations and provide the same software in a portable container to support remote and cloud-based analysis software environment provision.

<sup>3</sup><https://www.panosc.eu/wp-content/uploads/2019/12/D4.1-Report-Data-Analysis-Capture.pdf>

<sup>4</sup><https://confluence.panosc.eu/display/wp4/Task+4.1+-+Extended+Survey+Summary>

### Typical workflow for Common Portal usage

User authentication using Umbrella ID

Search for datasets (for which the user has been a proposal member, thus requiring authenticated access)

Selection of an analysis environment (in either a remote desktop, or Jupyter Notebook environment)

Spawn the chosen analysis environment and link it to the chosen datasets (if the analysis environment is not physically located with the data, data transfer protocols will be active)

Access the environment via the Portal and perform data analysis

### Common Portal - Achievements

- Possible **use cases of the Portal** have been listed<sup>5</sup>;
- Definition of the Portal Architecture** by adopting a microservices approach (foundation services, user services and compute services<sup>6,7</sup>), for more flexible integration into site-specific infrastructures.

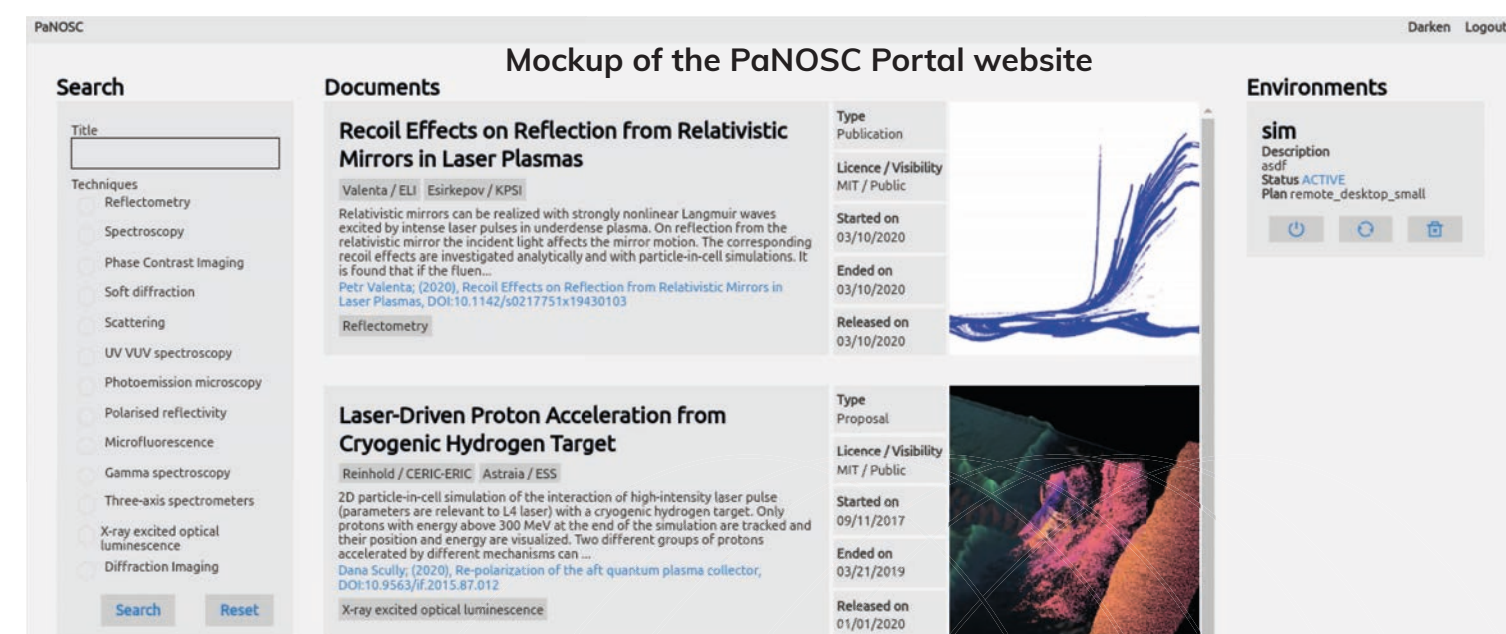
After initial deployment at facilities to provide remote analysis services to local data, the Portal will be deployed as part of the EOSC to provide federated data analysis of data across the facilities.

PaNOSC has also contributed significantly to implementing services for remote experiments which are being generalised due to the COVID-19 pandemic.

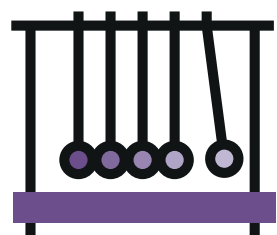
<sup>5</sup>PaNOSC Use Cases Confluence page: <https://confluence.panosc.eu/x/lwGm>

<sup>6</sup>Specifications for each microservice: <https://confluence.panosc.eu/x/lwCm>

<sup>7</sup>Source code on Github, for microservices' development: <https://github.com/panosc-portal>







## Simulation Data Services

WP leader: **Carsten Fortmann-Grote**

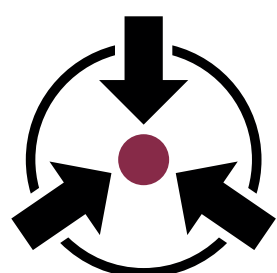
PaNOSC has been developing the “**Virtual Neutron and x-ray Laboratory**” (ViNYL), to offer services for simulation and modelling of PaN sources, beamlines and experimental instruments, as well as start-to-end simulations to describe entire experiments at PaN facilities.

### Milestones reached

- All simulation codes and frameworks were added to the PaNdata software catalogue;
- Domain-specific extensions published, to the simulation metadata standard openPMD for coherent wavefront data, molecular dynamics, photon and neutron raytracing.

### Ongoing Activities

- Harmonization of simulation APIs ongoing;
- A new API has started to be adopted in various simulation frameworks: SIMEX (start-to-end photon experiment simulations), McStas-Script (Neutron raytracing), OASYS (x-ray optics simulation framework);
- Developing a web based graphical tool to visualise and edit workflows.



## EOSC Integration

WP leader: **Jean-François Perrin**

PaNOSC has been working towards setting up a **federated Authorization and Authentication Infrastructure (AAI)** for the users of PaN facilities, which will allow seamless access to data and data services. In close collaboration with GÉANT, the UmbrellaID management formally approved the introduction of eduTEAMS in the UmbrellaID infrastructure, and services have been set up to accept authentication through it (e.g., [confluence.panosc.org](https://confluence.panosc.org), [pan-learning.org](https://pan-learning.org)).

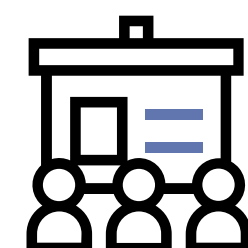


## Sustainability

WP leader: **Ornela De Giacomo**

PaNOSC has been constantly interacting with stakeholders, and contributing to shape the EOSC. This is necessary to ensure, through the proper definition of costs and metrics, realistic **business models** for the sustainability of the services to be developed during the lifetime of the project and made available through the EOSC.

The partners also identified possible areas of collaboration with other projects and initiatives.



## Staff & User Training

WP leader: **Thomas Rod**

The e-learning platform [e-neutrons.org](https://e-neutrons.org) has been migrated to ESS, where it is now operating under the domain name [pan-learning.org](https://pan-learning.org). It will be used to provide training resources for both staff and users of PaN sources.

Various solutions for integration of Jupyter in the platform have been identified, and work has started to integrate federated AAI.

PaNOSC and ExPaNDS will add new content, and workshops for both PaN staff and users will take place to get acquainted with the features and functionalities of [pan-learning.org](https://pan-learning.org).

The screenshot shows the pan-learning.org website. At the top is a navigation bar with links: FRONTPAGE, ABOUT PAN-LEARNING, FOR TEACHERS, SUPPORT, SIGN ME UP!, and COURSE LOGIN. Below the navigation bar, there are two main sections: 'Courses (login required)' and 'Science cases (login required)'. The 'Courses' section lists several courses, including 'Materials Science at a Virtual Neutron Facility', 'Introduction to Neutron Scattering', and 'Topics in Neutron Scattering'. The 'Science cases' section lists cases like 'cultural heritage objects', 'Monitoring solid state reactions', 'Characterising liposomes in suspension', 'Finding crystal structure', and 'Characterising magnetic order'. There is also a 'BRAGG EDGE IMAGING' section with a description of a module for investigating cultural heritage objects.



## Collaboration with ExPaNDS and the ESFRI clusters

PaNOSC has been working closely with its sister project ExPaNDS, to coordinate and harmonize joint actions in management and communications, as well as for the development of the PaN research data policy framework and the services for data catalogue, analysis and simulation.

Collaboration with ESFRI cluster projects and FAIRsFAIR has been increasing, to identify common challenges, share knowledge and tools, and identify common actions to better target user communities.

### Main Achievements @ month 18

- Active participation in the INFRAEOSC-O3 proposal with ESFRI clusters and European e-infrastructure;
- Publication of the ESFRI clusters' position papers on expectations and planned contributions to the EOSC;<sup>8</sup>
- Video release "The DOI for data"<sup>9</sup>, steered by PaNOSC, and resulted from the joint effort of PaN facilities in the PaNOSC and ExPaNDS projects, and LENS and LEAPS initiatives;
- Publication of a position paper on "EOSC – a tool for enabling Open Science in Europe" with the ESFRI clusters and European e-infrastructures;
- Contribution to EOSC Secretariat documents on EOSC sustainability (Strawman and Tinman documents).

### Ongoing Activities

- An action advocating scientific journals to include data DOIs in published papers is ongoing together with the LEAPS and LENS initiatives, and the FILL2030 project;
- Definition of common templates for the publication of use cases.

<sup>8</sup><https://doi.org/10.5281/zenodo.3675081>

<sup>9</sup>Video "The DOI for data" <https://bit.ly/The-DOI-for-Data>

## PaNOSC Partners

In collaboration with:



## What the European Commission says about PaNOSC

*PaNOSC has delivered exceptional results with significant immediate or potential impact.*

*PaNOSC has demonstrated a great understanding of EOSC ecosystem, objectives and challenges.*

*The performance and ambitions have been outstanding.*

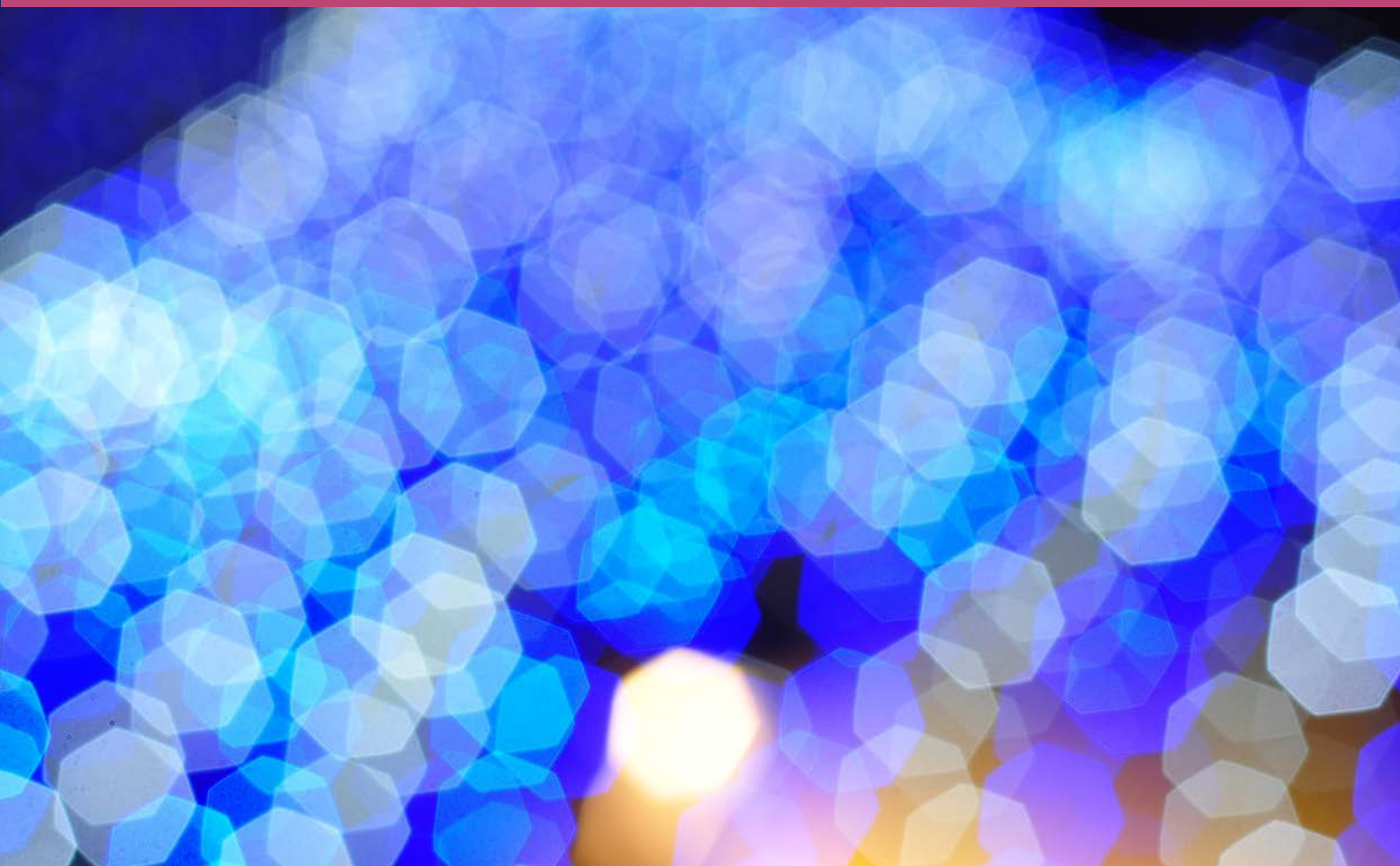
*Its major contribution is its impact on RIs, e-Infras, ESFRIs and national nodes in the PaN domain, promoting and facilitating the adoption of FAIR data principles.*

*The project has been very effectively managed, and also efficiently.*





Visit [www.panosc.eu](http://www.panosc.eu) for more information



**PUBLICATION DESIGNED  
AND EDITED BY**

Nicoletta Carboni  
(CERIC-ERIC)

**DATE**

October 2020

**WEBSITE**

[www.panosc.eu](http://www.panosc.eu)



PaNOSC has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 823852