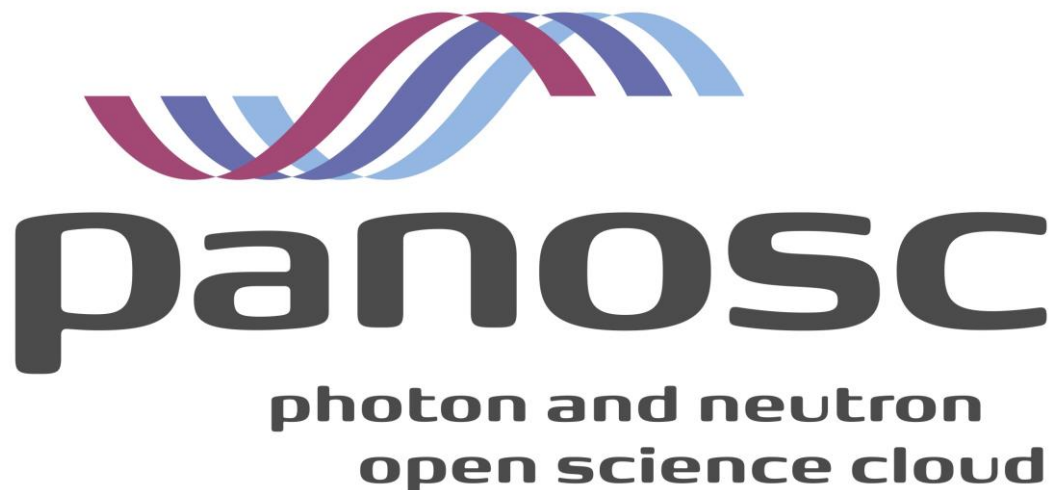


**PaNOSC**

**Photon and Neutron Open Science Cloud**

**H2020-INFRAEOSC-04-2018**

**Grant Agreement Number: 823852**



**Deliverable:**

**D7.4 PaN EOSC Sustainability plan**

# 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:	D7.4
Deliverable Type:	Report
Dissemination Level	Public
Contractual Delivery Date:	30/11/2022
Actual Delivery Date:	06/02/2023
EC project Officer:	Flavius Pana

## Document Control Sheet

<b>Document</b>	Title: PaN EOSC Sustainability plan
	Version: 1
	Available at: <a href="https://github.com/panosc-eu">https://github.com/panosc-eu</a>
	Files: 1
<b>Authorship</b>	Written by: Ornela de Giacomo, Angela Zennaro
	Contributors: Dariusz Brzosko, Teodor Ivanoaica, Florian Glickson, Andy Götz, Fabio Dall'Antonia, Jean-François Perrin; Tobias Richter, Erwan Le Gall, Juncheng E.
	Reviewed by: Andy Götz
	Approved: Jordi Boderà

## 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

Table of Content	4
1. Executive summary	6
2. Introduction	7
2.1 Drivers of FAIR and open data	7
ENVRI-FAIR for environmental research	7
EOSC-Life for life sciences	7
ESCAPE for astronomy and particle physics	7
PaNOSC for multidisciplinary scientific analysis	7
SSHOC for social sciences and humanities	7
2.2 Provision of FAIR data in the current context of Research Infrastructures	8
2.3 Why is the sustainability of the Photon and Neutron EOSC important?	10
2.4 Interaction with stakeholders	11
3. Factors affecting the long term sustainability of the PaN EOSC	12
3.1 Uptake and adoption by PaN facilities	12
3.2 Uptake and adoption by researchers	21
3.3 Alignment with EOSC and interoperability	21
3.4 Technical sustainability	22
3.5 Financial sustainability	23
3.5.1 Costs of development and upgrade	23
3.5.2 Costs of adoption or implementation	23
3.5.3 Costs of operation and maintenance	24
3.5.3.1 FAIR and open Data Policy harmonization	25
3.5.3.2 Standardized metadata and data catalogues	25
3.5.3.3 Open Data portal	26
3.5.3.4 Software catalogue	27
3.5.3.5 Community AAI UmbrellaId	27
3.5.3.6 Remote Data Analysis Services	28
3.5.3.7 The Virtual Neutron and x-ray Laboratory (ViNYL)	28
3.5.3.8 PaN learning and training platform	29
3.5.3.9 Communication and marketing	30

3.6 Governance	30
4. Strategies for long-term sustainability of the PaN EOSC	32
Scenario 1: Identification of essential services to be adopted by all PaN RIs	33
Scenario 2: Discussion of the single outputs independently.	34
5. Conclusions	35

# 1. Executive summary

Since the launch of the European Cloud initiative including a “research open science cloud”, the implementation of the European Open Science Cloud (EOSC) was supported with substantial resources from the EC. The call for “Connecting ESFRI<sup>1</sup> infrastructures through Cluster projects” allowed the Photon and Neutron (PaN) community cluster to receive co-funding by the European Commission (EC) to implement data policies based on FAIR principles<sup>2</sup> and connect to the EOSC. As the end of PaNOSC approaches, the sustainability of the developments achieved during these will need to be developed. Sustainability is influenced by many aspects, the most relevant ones are briefly described in this document.

A sustainable PaN EOSC should have common tools and services adopted by most PaN facilities and researchers. It should have a high degree of interoperability with EOSC and with other science clusters and communities. It should be technically sustainable, maintaining its functionality over time and must have enough resources to operate, develop and provide high quality services to researchers of the PaN community and, in a vision of EOSC, to any user interested. The resources PaN facilities count on are mostly from their ordinary budget, with some limited project funding available for new developments and networking activities. However, none of these resources can cover the costs to operate tools and services for a wider community. PaNOSC partners have identified the LEAPS<sup>3</sup> and LENS<sup>4</sup> initiatives as the Initiatives where the communities will continue their discussions in the future. Finally, a sustainably PaN EOSC should have an appropriate governance that ensures inclusiveness and participation, but at the same time provides the direction.

The sustainability in the long term can be achieved through different strategies. During the project, the authors of this document proposed a holistic approach that considered for the PaNOSC and ExPaNDS Key Exploitable Results (KERs), the vision for the PaN EOSC, the requirements and dependencies on external factors. This approach was considered however difficult to implement and sustain because it would require creating and managing a new organisation without having new resources. The option chosen by PaNOSC and ExPaNDS<sup>5</sup> to sustainability was to leverage the LEAPS and LENS initiatives, adopting the open source collaboration approach and where necessary to negotiate Memorandums of Understanding (MoUs) for the outputs individually. At the closure of this project, two MoUs are in advanced negotiation, one (AAI<sup>6</sup>) already signed by many PaNs, and a second one (VISA) In an advanced stage of preparation. Other KERs like H5Web viewer, Search API, OASYS and VINYL simulation, pan-learning.org training platform will be maintained as open source projects. The PaN Data Commons will be consolidated through LEAPS+LENS and the EOSC.

---

<sup>1</sup> European Strategy Forum on Research Infrastructures: <https://www.esfri.eu/>

<sup>2</sup> Findable, Accessible, Interoperable and Reusable: <https://www.nature.com/articles/sdata201618>

<sup>3</sup> League of European Accelerator-based Photon Sources: <https://leaps-initiative.eu/>

<sup>4</sup> League of Advanced European Neutron Sources: <https://lens-initiative.org/>

<sup>5</sup> European Open Science Cloud Photon and Neutron Data Service project: <https://expands.eu/>

<sup>6</sup> Authentication and Authorisation Infrastructure

## 2. Introduction

### 2.1 Drivers of FAIR and open data

In 2014, European Commission President Jean-Claude Juncker identified as priority n°2 for the next European Commission *A Connected Digital Single Market*. With a global economy rapidly becoming digital, it was clear that there would be immense opportunities for innovation, growth and jobs. However, it was acknowledged that the full exploitation of the Digital Single Market would raise challenging policy issues for public authorities, which would require coordinated EU action<sup>7</sup>. In “**A Digital Single Market Strategy for Europe**” Published in 2015<sup>8</sup> it is announced that “the Commission will launch a European Cloud initiative including cloud services certification, contracts, switching of cloud services providers and a **research open science cloud**”. The implementation of the EOSC was supported with substantial resources from the EC. The Horizon 2020 framework programme introduced the Research and Innovation action “Enabling an operational, open and FAIR EOSC ecosystem (INFRAEOSC)” that helped to develop the “Web of FAIR Data and Services” for Science, involving the main actors. In particular, in the call INFRAEOSC 04-2018 aimed at Connecting ESFRI infrastructures through Cluster projects. There were five successful cluster projects awarded funding<sup>9</sup>:



ENVRI-FAIR for environmental research



EOSC-Life for life sciences



ESCAPE for astronomy and particle physics



PaNOSC for multidisciplinary scientific analysis



SSHOC for social sciences and humanities

<sup>7</sup> Extract from the Political Guidelines for the next European Commission – A New Start for Europe: My Agenda for Jobs, Growth, Fairness and Democratic Change (15 July 2014)

<sup>8</sup> file:///Users/ornela/Downloads/DSM\_communication.pdf.pdf

<sup>9</sup><https://cordis.europa.eu/search/it?q=contenttype%3D%27project%27%20AND%20programme%2Fcode%3D%27INFRAEOSC-04-2018%27&p=1&num=10&srt=/project/contentUpdateDate:decreasing>

These ESFRI cluster projects have been working from 2018 to connect with the EOSC. Together, they've implemented interfaces to integrate computer and data management solutions, created cross-border and open cooperation spaces and promoted clouds via the EOSC portal for a larger user community. They've met regularly to exchange good practices and confer on common challenges. In terms of sustainability of the KERs and the community, some cluster projects adopted an MoU as the primary solution. In the case of the PaN cluster the LEAPS and LENS initiatives existed already as representatives of the PaN facilities in Europe.

With the cluster projects coming to an end, the sustainability of the developments achieved during these years thanks to the seed EC funding provided to PaNOSC will need to be developed. The PaN community has very good reasons to keep working together. The negotiation for the signature of MoUs is in progress, but is not straightforward. The sustainability of the Photon and Neutron Open Science Cloud (PaNOSC), intended as the cloud connecting all PaN facilities, has many aspects to be considered beyond the financial sustainability. These aspects have been discussed extensively during the execution of the project; this document presents a summary.

In addition to the cluster projects, the EC issued a call for a Research and innovation action with the topic INFRAEOSC-05-2018-2019 - Support to the EOSC Governance, where the ExPaNDS project was granted support.



The aim of the EU-funded ExPaNDS project is to enable photon and neutron research infrastructures at national level by making the majority of their data 'open', following FAIR principles, and harmonising their data catalogues and data analysis services through the EOSC. The work of PaNOSC and ExPaNDS has been defined together, since PaNOSC started a year earlier. Both projects have worked closely together and it is natural that the sustainability of the outputs is considered for the large community of ESFRIs and ERICs (PaNOSC) and national infrastructures (ExPaNDS).

## **2.2 Provision of FAIR data in the current context of Research Infrastructures**

One of the main difficulties when dealing with linking Research Infrastructures (RIs) to EOSC is the great variability of the landscape. There is no "one size fits all" for the management of data between RIs. They can be single sited or distributed, discipline specific or multi-disciplinary, large or small, digital or not. A part of this variability is represented in the cluster projects, although not all of it.

In this context, PaNOSC facilities are mainly particle accelerators, free electron lasers and



neutron sources. Until few years ago, the main objective of these facilities was to operate the best machines in the world, to enable science that is otherwise not possible. The investments to build these machines are huge and, in most cases, require international cooperation to be feasible, and consequently also the operation costs are substantial. In addition to the science they enable, they are wonderful examples of complex engineering and the societal return of these investments is much broader than pure science. They foster innovation, create jobs and capabilities. Although they produce data as the main output, data processes such as analysis and storage have been mostly left to the scientists operating the instruments. Therefore, in the same facilities, there were (and still are) different standards and procedures.

The project PaNOSC has allowed 5 ESFRI facilities, 1 ERIC and 1 E-Infrastructure to work together for a period of 4 years. During the project's lifetime, these RIs agreed on a framework data policy to be implemented, with the objective to adopt practices compliant with the FAIR principles, making data produced Findable, Accessible, Interoperable and Reusable. Achieving the implementation of FAIR principles implied several changes at the level of RIs, that went from the modification of their control systems to acquire relevant metadata, the adoption of standard formats for data, the use of electronic logbooks, Jupyter notebooks<sup>10</sup> and other tools and workflows for data analysis. This important step supports the reproducibility of scientific experiments and scientific results, interdisciplinary research, and, by promoting the reuse of data it improves scientific experiments, the evolution of common data culture, common training and preparation of new generations of researchers, thus maximizing the value and impact of research facilities' data by putting FAIR data in the foreground.

This said, the main objective of PaNOSC facilities is still to operate the best machines, but data management has entered into the culture of the RIs and their management, becoming a core aspect of the operations. The management of these facilities has committed, at the beginning of the EU project PaNOSC, to adopt the FAIR principles for data management and to provide sustainable services. However, the technical and governance complexity may have been underestimated at the time, due to the little experience of most of the facilities in the provision of open data and services.

Therefore, while many of the RIs in the other four ESFRI clusters (ESCAPE, ENVRI, EOSC-LIFE and SSHOC) have been dealing with open data and digital assets for decades as their core activity, the situation is different for the PaNOSC RIs. In other clusters, the data providers are third parties and the sustainability of some of the RIs strongly depend on the sustainability of their data providers. In their case, they behave as "mini EOSCs", delivering services for interoperability, findability and accessibility of data, and helping the data providers to agree on the standards that allow the interoperability of data. In PaNOSC, the RIs are the data providers, and so far, none of the PaN facilities have delegated the data

---

<sup>10</sup> <https://jupyter.org/>

management activities to other legal entities. This is both a strength and a weakness: a weakness because data standards and FAIR data management, including preservation, are a relatively new activity for most of the PaN RIs with few exceptions. Actually, until not long ago these were not considered as core activities, so there is a know-how and competence that needed to be created and will need to be developed in the future. However, being the producers, managers and custodians of the data gives the full power to these RIs to follow good practices, decide the quality of the services provided and to enforce any practices and policies they consider better for the exploitation of the data.

## 2.3 Why is the sustainability of the Photon and Neutron EOSC important?

- To increase **credibility in science**: The EOSC pursues this objective as a primary objective. To be able to link more effectively with the EOSC, the PaN community should develop and maintain the PaN Data Commons. The PaN Data Commons is all open data from Photon and Neutron sources in Europe accessible from a single web portal. Only through the adoption of the FAIR principles and common standards for the PaN community, will results become reproducible.
- To provide a better service for **users of our facilities**: The provision of FAIR data is in first place a service provided to users. This allows them to easily retrieve, handle, analyse their data and share it with the other members of their team, without the need to copy or move their data. With the improvement in detectors and high throughput instruments PaN facilities are meeting the condition where the volume of data produced in an experiment makes it impossible or non-practical to transfer the data from the facility. The tools developed by PaNOSC allow the analysis onsite, without the need to transfer data. They also allow to store safely any meaningful derived data and assign them a persistent identifier, for data to be found by others, cited in publications, etc. Moreover, through the adoption of common standards and tools, users don't need to learn a new way every time they visit a new PaN facility. The ambition is that members of the PaN community can be trained to use a common set of services, making **the most efficient use of time and resources** for them and the facilities.
- The provision of **FAIR data is a requirement** for Horizon Europe projects. The adoption of the FAIR principles by the facilities participating in PaNOSC will make it easier for them to participate in projects co-funded by the EC.
- Countries are developing and/or implementing **National data policies**, so RIs will not be exempt from adhering. In this sense, having PaN commons can help steer the process at the national level and contribute to the alignment of national policies.
- There were **substantial investments in PaNOSC** developments (€12 M from the EC, plus in-kind from the facilities), the EOSC Future project and others. This provides the opportunity for others to benefit from this experience and the project outputs, but it also gives the responsibility to PaN facilities to commit the effort to maintain these outputs in the future and further develop them.

## 2.4 Interaction with stakeholders

ExPaNDS and PaNOSC have worked on developing and supporting the adoption of the building blocks of a data commons by the Photon and Neutron scientific community. These data commons, ranging from policies and data services structuring the support of the management to common data standards adopted, is maintained and continuously improved by a scientific community which is now starting to also share services for computing and processing FAIR data from the PaN RIs.

Some of the developments towards Open Science at facilities were co-funded by the European Commission through EOSC projects (e.g. PaNOSC, ExPaNDS), while others were covered with the regular funding of facilities who had committed to provide open and FAIR data, integrating this as one of their core activities, aligned with the mission of their Research Infrastructures (RIs).

With the projects coming to an end, the sustainability of the developments achieved during these years thanks to the seed EC funding provided to PaNOSC and ExPaNDS will need to be developed. The PaN community has very good reasons to keep working together and to invest in joint developments, since many of the services are useful for a large part of the facilities. There is also a common interest to make a more efficient use of the very limited resources (even more after the energy crisis that was generated after the aggression of Russia to Ukraine). Although important, the financial sustainability is only one of them. It is clear that after a funding stream ends, the question of how the operation and development costs be covered is natural, but that is only part of the problem. The governance of all these outputs is a crucial aspect as well. While the project structure assigns responsibilities and resources, for the execution of a well defined work plan, the end of a project marks the transition to a phase of self-organisation, that requires institutions to agree on a strategy and on the commitment of resources in the medium and long term. The commitments required are not always equally distributed between the institutions. For example, one of the facilities hosts and maintains the training portal operation for most of them. For this reason, facilities need to negotiate thoroughly between them, to agree on a fair compensation for the one hosting or maintaining a service and making it available to all the others. But again, the aspect of the governance is crucial to ensure the interest of all the institutions supporting a development is taken into account.

Most of the IT professionals and scientists working in PaNOSC have interacted during the project with users and other scientists, so they received their direct feedback. Work packages leaders have been exposed to them as well, and in addition have the overview on the implementation of the description of work, the obstacles and the direction. This valuable information contributed to our analysis and the preparation of scenarios for the

sustainability of the PaN EOSC and the outputs of PaNOSC.

We have run two interviews with the work package leaders and the project coordinator. The first one took place towards the beginning of the second project year, to start formulating hypotheses on sustainable models. The second one in the third project year, to get an update and account for any change that had happened in the meantime. The interviews were based on the following questions:

- What outputs of your WP are expected to last after PaNOSC?
- What are the necessary resources, per partner and per year, to maintain such outputs?
- Does the sustainability of these outputs depend also on the uptake by ExPaNDS facilities?
- Are there other external factors that may condition the sustainability of your WP outputs? (for example, do you rely on software/hardware developed by other communities?)

During the interviews, WP leaders provided their estimations on the resources needed but also their opinion on the best way to maintain these products or services after the end of the project. Their replies were very valuable for exploring scenarios for a sustainable future and also for identifying threats and pondering risks.

### **3. Factors affecting the long term sustainability of the PaN EOSC**

The sustainability of the achievements of PaNOSC is influenced by many aspects. This session will briefly describe some of the most relevant ones.

In the project we have identified the following elements as those impacting substantially in the sustainability of PaNOSC outputs and the PaN commons:

1. Uptake and adoption by PaN facilities
2. Uptake and adoption by researchers
3. Alignment with EOSC and interoperability
4. Technical sustainability
5. Financial sustainability
6. Governance

#### **3.1 Uptake and adoption by PaN facilities**

It is important to remember that PaNOSC worked alongside with ExPaNDS, but in order to optimize the outcomes, these projects had a different focus and therefore produced different outputs, that are however connected. This is relevant for the sustainability of PaNOSC

because for commons to be such they need to be adopted by all the community. This implies that there is an expectation for PaN facilities that were not partners in PaNOSC, to adopt PaNOSC (and ExPaNDS) outputs and contribute to their further development, upgrade and maintenance. For some of these outputs, it is critical that most of the PaN facilities adopt them. To provide a concrete example, the value of a search tool is connected to its performance in terms of whether it can deliver the search results expected but also to the number of registries composing the database searched. A researcher looking for all the existing datasets related to their own research would not use a PaN search tool unless it delivers information from most of the PaN facilities. In case only few facilities decided to implement the search tool, then the sustainability of this tool would be threatened because its added value would be questionable or not high enough to justify the investment by the facility, if competing services like B2Find or Open Aire would offer a less targeted search but would search across a higher number of facilities producing this kind of datasets.

These considerations were discussed in different phases of the project to evaluate the different approaches to the sustainability of PaNOSC outputs. During the third year of the project, the coordinator proposed to launch a survey between LEAPS members, presenting an estimation of the resources needed for the adoption of different PaNOSC/ExPaNDS outputs. The time estimates are approximate based on the assumption that facilities build on existing solutions developed by PaNOSC and ExPaNDS and do not start from scratch. The figures are lower boundaries and will vary from site to site depending on local resources available and priorities. Almost all the outcomes require that facilities have a Data Manager dedicated to scientific data management, this is still not the case at most facilities which makes it more difficult if not impossible. In the table below the effort for adopting PaNOSC and ExPaNDS outputs is shown, and the following abbreviations have been used: DM = Data Manager, DP = Data Policy, PM = Person Months.

Outcome	Action(s)	Resources required
1. FAIR data policy	Adopt or modify existing data policies to be FAIR.	Management support to prepare and present DP to governing bodies. 2 PMs of a DM to prepare the DP.
2. Data Management Plans (DMPs)	Implement DMPs for users based on outcomes of PaNOSC + ExPaNDS.	6 PMs of a DM to implement DMP solution and integrate it into the UO workflow.
3. FAIR assessment and common PID framework	Implement Digital Object Identifiers (DOIs) for data. Setup a WG to assess FAIRness of data.	12 PMs of a DM to setup a contract with Datacite and implement a data repository for registering data. This assumes adopting one the existing data repositories + having access to an infrastructure for archiving data.

Outcome	Action(s)	Resources required
4. Standardised metadata (Nexus/HDF5, PaN ontologies)	Adopt Nexus/HDF5 and produce data following the Nexus conventions.	12 PMs of a controls engineer to add support for Nexus/HDF5 to the control system. 2 PMs per technique for data scientists to add support for Nexus/HDF5.
5. Federated search API for PaN data catalogues	Implement the PaNOSC search API as a service for the local data repository.	6 PMs of a DM to implement the search API if the local repository is not a standard one.
6. Open Data portal for searching + downloading data	Implement a metadata catalogue and data repository to upload, search and download open data.	24 PMs of a DM + 12 PMs of an IT infrastructure engineer to adapt an existing solution locally. 10 k€ per year per petabyte of archived data for long term storage (rough estimate).
7. Community AAI Umbrellaid	Implement the PaN community AAI Umbrellaid based on eduTEAMS.	6 PMs of a DM to implement the latest version of Umbrellaid for data services.
8. JupyterLab notebooks and HDF5/NeXus files visualisation	Implement a Jupyterlab instance for remote data viewing and analysis.	6 PMs of a software engineer and 6 PMs of an IT infrastructure engineer.
9. Remote data analysis with VISA + data analysis pipelines	Implement the VISA remote analysis service.	12 PMs of an IT infrastructure engineer and 6 PMs of a DM.
10. Simulation software for simulating experimental data (SIMEX)	Adopt and install the SIMEX simulation software.	6 PMs of a software engineer.
11. PaN-learning platform (pan-learning.org + pan-training.org)	Use and upload new training material to the PaN-learning platform.	1 PM of a trainer to prepare training material per new topic.

Table 1: Estimation of the costs to adopt the main PaNOSC/ExPaNDS Outputs

After presenting these estimation, the LEAPS/LENS IT working group representatives together with their management assessed the feasibility of the adoption of the outputs,

based on the estimated effort. The result was that many of the PaN RIs declared that they were willing to do so. Some could not come with a reply or commitment from the management since at this point in the project all the context around EOSC and open data was not completely clear. The results of this first survey, dated October 2021 are presented as table 2:

FACILITY	FAIR data policy	DMPs	DOIs	Nexus HDF5	Search API	Open Data Portal	AAI	JupyterLab	VISA	SIMEX	Pan-learning/training
ALBA	P	P	WIP	WIP	P	P	U	P	U	U	U
DESY	WIP	P	P	Y	WIP	P	WIP	Y	U	N	WIP
DIAMOND											
ELETTRA	Y	WIP	Y	Y	WIP	WIP	Y	Y	WIP	Y	WIP
ESRF	Y	WIP	Y	Y	WIP	WIP	Y	Y	WIP	Y	WIP
EuXFEL	WIP	WIP	Y	WIP	WIP	WIP	WIP	Y	WIP	Y	WIP
FELIX	Y	P	WIP	U	U	WIP	U	U	N	N	U
HZB	Y,N,P <sup>1</sup>	P	WIP <sup>2</sup>	Y	P	Y	P	U	U	U	U
HZDR	WIP	WIP	Y	N	N	WIP	WIP	WIP	P	N	Y
INFN	U	U	U	U	U	U	U	U	U	U	U
ISA*	U	U	U	U	U	U	U	U	U	U	
MAX IV											
PSI	WIP	WIP	Y	WIP	Y	Y	WIP	WIP	N	N	N
PTB	Y	WIP	Y	WIP	N	Y	N	N	N	N	N
SOLARIS #											
SOLEIL	U	P	WIP	Y	P	P	WIP	U	WIP	U	Y

Table 2: First survey on the adoption of PaNOSC/ExPaNDS outputs run between LEAPS members in October 2021.

Reference a: legend for table 2

Yes, already adopted (Y)
Not Planning to be adopted (N)
In progress of being adopted (WIP)
Planned to be adopted (P)
Under evaluation (U)
No answer



Meaning of outcomes and their adoption status: The following table explains briefly what the PaNOSC and ExPaNDS outcomes are and what the adoption status codes (Y, WIP, P, N, U) represent for each outcome.

Outcome	Meaning	Meaning of progress adoption status
1. <b>FAIR data policy</b>	Adopt or modify existing data policies for Open Data specifically mentioning the FAIR principles	Y = Data Policy specifically mentions the FAIR principles; WIP = Data Policy has been updated to include FAIR and is in the process of being accepted; P = FAIR Data Policy is planned; N = do not have a FAIR Data Policy and not planning to update to support FAIR principles; U = still evaluating adopting a Data Policy
2. <b>Data Management Plans (DMPs)</b>	Implement DMPs for users based on outcomes of PaNOSC + ExPaNDS.	Y = DMPs are generated for each experiment but are not necessarily mandatory for users to fill in to get beamtime; WIP = DMPS are being implemented; P = DMPs are planned to be implemented; N = DMPs are not planned; U = still evaluating DMPs
3. <b>FAIR assessment and common PID framework</b>	Implement Digital Object Identifiers (DOIs) for data. Setup a WG to assess FAIRness of data.	Y = an institute DOI is generated for each experiment automatically and scientists are requested to cite the DOIs; WIP = an automatically generated institute DOI is in the process of being implemented; P = an institute DOI is planned to be implemented; N = institute DOIs will not be implemented; U = still evaluating DOIs



Outcome	Meaning	Meaning of progress adoption status
4. Standardised metadata (Nexus/HDF5, PaN ontologies)	Adopt Nexus/HDF5 and produce data following the Nexus conventions.	Y = Nexus/HDF5 is supported as the main data format for raw and (most) processed data; other formats can be converted to and from Nexus/HDF5; Nexus is the main ontology for metadata; WIP = in the process of supporting Nexus/HDF as supported format; P = plan to adopt Nexus/HDF5 as preferred format; N = not planning to adopt Nexus/HDF5; U = still evaluating Nexus/HDF5
5. Federated search API for PaN data catalogues	Implement the PaNOSC search API + scoring as a service for the local data repository.	Y = the PaNOSC search API and scoring are implemented and included in the search portal ( <a href="https://data.panosc.eu">https://data.panosc.eu</a> ); WIP = the PaNOSC search API and/or scoring are in the process of being implemented; P = the search API and scoring are planned to be implemented; N = the search API and scoring will not to be implemented; U = still evaluating the search API + scoring
6. Open Data portal for searching + downloading data	Implement a metadata catalogue and data repository to upload, search and download open data.	Y = a metadata catalogue and data repository for Open Data have been implemented and are publicly accessible; WIP = a metadata catalogue and data repository for Open Data are in the process of being implemented; P = a metadata catalogue and data repository for Open Data are planned to be implemented; N = a metadata catalogue and data repository for Open Data will not be implemented; U = still evaluating a metadata catalogue and data repository for Open Data;

Outcome	Meaning	Meaning of progress adoption status
7. <b>Community AAI Umbrellald</b>	Implement the PaN community AAI Umbrellald based on eduTEAMS.	Y = the Umbrellald AAI is implemented at the facility as one of the authentication options for one or more services; WIP = in the process of implementing Umbrellald AAI at the facility; P = planning to implement Umbrellald AAI; N = Umbrellald will not be implemented; U = still evaluating Umbrellald
8. <b>JupyterLab notebooks and HDF5/NeXus files visualisation</b>	Implement a Jupyterlab instance for remote data viewing and interactive analysis.	Y = a JupyterLab+h5web instance is installed as a service for users; WIP = in the process of installing a JupyterLab+h5web instance as a service for users; P = planning to install a JupyterLab+h5web instance as a service for users; N = will not install a JupyterLab+h5web instance as a service for users;; U = still evaluating Jupyterlab+h5web
9. <b>Remote data analysis with VISA + data analysis pipelines</b>	Implement the VISA remote analysis service.	Y = a VISA instance is installed as a service for users; WIP = in the process of installing a VISA instance for users; P = planning to install a VISA instance for users; N = will not install a VISA instance; U = still evaluating VISA
10. <b>Simulation software for simulating experimental data (VINYL)</b>	Adopt and install one or all of the simulation software packages which are part of VINYL i.e. SIMEX, OASYS, or McStasScript	Y = installed and using one or more VINYL simulation packages for one or more beamlines ; WIP = in the process of installing one or more VINYL packages; P = planning to install one or more of the VINYL packages; N =not planning to install any of the VINYL packages; U = still evaluating VINYL

Outcome	Meaning	Meaning of progress adoption status
<b>11. PaN-learning platform</b> (pan-learning.org + pan-training.org)	Use and upload new training material to the PaN-learning platform.	Y = developed training material on pan-learning.org; WIP = in the process of developing training material for pan-learning.org; P = planning to develop training material for pan-learning.org; N = will not develop training material for pan-learning.org; U = still evaluating pan-learning .org

Reference b: PaNOSC and ExPaNDS outcomes and meaning of the adoption status codes

The same survey was repeated towards the end of the project, this time including also LENS. The results were collected in an internal paper of the LEAPS/LENS IT working group, with recommendations on how to continue with the collaboration in the frame of this working group, also in view of the future calls of Horizon Europe. The paper will be published as part of a special issue of the EPJ+ journal on LEAPS strategy. The table with the replies of LEAPS/LENS facilities this time, and including also the PaNOSC partners CERIC-ERIC and ELI-ERIC (not members of LEAPS or LENS) are presented in table 3 below.

FACILITY	FAIR data policy	DMPs	DOIs	Nexus HDF5	Search API	Open Data Portal	AAI	Jupyter Lab	VISA	VINYL/O ASYS/Mc Stas	Pan- learning / training
ALBA	P	P	WIP	WIP	WIP	WIP	P	Y	WIP	N	U
BNC											
DESY	WIP	WIP	WIP	Y	WIP	P	WIP	Y	U	Y	WIP
CERIC- ERIC	Y	WIP	Y	WIP	Y	Y	Y	Y	Y	Y	Y
DIAMOND											
ELETTRA	Y	WIP	Y	Y	Y	Y	Y	Y	Y	Y	Y
ESRF	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
ELI-ERIC	Y	Y	P	Y	Y	Y	WIP	Y	Y	Y	Y
ESS	Y	Y	Y	Y	Y	Y	Y	WIP	WIP	Y	Y
EuXFEL	Y	WIP	Y	WIP	Y	Y	WIP	Y	WIP	Y	Y
FELIX	Y	P	WIP	U	U	WIP	U	U	N	N	U
HZB	Y	P	WIP	Y	P	Y	P	U	U	U	U
HZDR	Y	WIP	Y	N	UN	Y	Y	Y	P	N	Y
ILL	Y	WIP	Y	Y	WIP	Y	Y	Y	Y	Y	WIP
MAX-IV	WIP	U	Y	Y	Y	Y	Y	Y	U	U	U
HZDR	Y	WIP	Y	N	UN	Y	Y	Y	P	N	Y
ILL	Y	WIP	Y	Y	WIP	Y	Y	Y	Y	Y	WIP
MAX-IV	WIP	U	Y	Y	Y	Y	Y	Y	U	U	U
PSI	Y	WIP	Y	WIP	Y	Y	WIP	WIP	N	N	N
PTB	Y	WIP	Y	WIP	N	Y	N	N	N	N	N
SOLARIS											
SOLEIL	Y	WIP	WIP	Y	WIP	WIP	Y	WIP	WIP	U	Y
SESAME	Y	U	P	Y	P	WIP	P	P	N	Y	N

Table 3: Results of the survey run by the coordinator between LEAPS and LENS facilities, including PaNOSC partners CERIC-ERIC and ELI, in October 2022.

One LEAPS member (SOLARIS) and one LENS member (BNC) who did not provide answers, provide access to their instruments through CERIC-ERIC and are therefore following closely the implementation of PaNOSC and ExPaNDS outputs in CERIC-ERIC.

From the positive outcome of this survey, it emerges that the costs of adoption are acceptable and most facilities either have already adopted PaNOSC/ ExPaNDS outputs or are in the process. We can conclude that the costs of adoption are not a threat to the sustainability of the outputs.

### **3.2 Uptake and adoption by researchers**

Everything that is being developed makes sense if researchers perceive it as an advantage, a useful service more than an obligation. Researchers are aware that there is a concerted effort in Europe towards making science and scientific results open, but they struggle to find the time to take care of additional “bureaucracy” on top of their obligations. For this reason, PaNOSC delivered solutions that reply to the call of requiring the less possible manual intervention. The Data Management Plan tool adopted requires from no intervention (in the case of some facilities) to little in the case of others, but provides an instrument that help researchers to deliver a document that is required by funding agencies and by some institutions. For this reason, the uptake by the community has been very positive. Regarding metadata, it is evident that an effort from the community in defining the set of relevant metadata is necessary, but this discussion has been ongoing from years, independently on these projects. The use of electronic logbooks has helped to capture the input of users in terms of metadata (sample description) and although this is still in progress, many of the facilities have successfully adapted their control systems to capture as many relevant parameters as possible. Facilities that were still not in operation at the beginning of the project (e.g. ESS) had the advantage to adopt FAIR principles by design. The availability of rich metadata is seen as an advantage not only for researchers that want to reuse the data but for the same teams generating the data, since this allows also them to reconstruct their experiments in every detail, having the information stored safely. The simulation tools and online processing have also been well received by researchers. In the past, the usual procedure was for them to record their data and copy it in a drive, sometimes many of them. They needed to install the data analysis software in their own computers and deal with compatibility issues. Currently, the availability of local storage, simulation and processing tools online has made these issues a matter of the past. Moreover, during these years the simulation and data analysis software was improved in a concerted effort and users have better chances of finding the same software in different facilities, making it easier for them and reducing the time they need to employ to learn how to use different software for the same purpose.

The uptake and adoption by the community will be monitored in the future, and special attention will be put in developing solutions that are perceived as such by the researchers. We expect that a part of this monitoring process will be done directly in the EOSC interface, as most of the solutions were onboarded in the EOSC portal and will be accessible openly from there. The PaN community follow their own community closely, how this will be done will depend on the governance model chosen.

### **3.3 Alignment with EOSC and interoperability**

The main goal of the PaNOSC project was to link with EOSC. This has been done following

closely and participating actively in EOSC working groups (first phase) and the EOSC-A task forces (second phase). Partners have also been active in participating to consultations shaping the EOSC and in producing position papers, together with the other clusters. For example, the AAI solution used by the PAN community, UmbrellaID, was developed to become EOSC-ready from the technical point of view, through the adoption of the AARC blueprint architecture. Another example are the FAIR data catalogues. In this case, a search API was developed to create an interoperability layer that supports meta-data harvesting and discovery by OpenAIRE or B2FIND as well as interoperability in the EOSC ecosystem.

The aspect of the alignment with EOSC has remained a priority and many services were onboarded to make them available through the EOSC portal. However, at the time the project is ending, the interoperability layer of EOSC has not been defined, so we expect this work to be carried out in the following years, after the end of PaNOSC.

The interoperability of PaNOSC tools and services is a goal since the adoption of the services by a broader community will contribute to their sustainability. In this sense, PaNOSC has maintained a strong interaction with the other science clusters to find common services and practices, and was involved in the EOSC Future project through the participation of the coordinator, who regularly updated PaNOSC partners on the developments of the project.

The future co-funded activities, currently projects in discussion, are in fact activities that explore the use of common tools across disciplines, in collaboration with the other clusters.

### **3.4 Technical sustainability**

The services developed will need to be maintained and updated or upgraded in the near future. Just to provide an example, the reproducibility of an experiment requires that also the analysis workflow is available. Facilities have developed software for data analysis and adopt best practices for versioning. However, with a change of the operating system the same software may no longer work. All these aspects will need to be kept under control, to allow the reproducibility and sometimes the reusability of data. The same applies to all the processes, software and Application Programming Interfaces (APIs) used in the implementation of the FAIR principles. Facilities will need to guarantee that enough personnel with the right qualification to perform these tasks is available over time.

Another consideration regards third party software. Some of the PaNOSC outputs rely on or are based on software developed by other communities or even companies. It must be ensured that these communities or companies maintain the development and support of these products in the long term. One example is the Nexus standard. Other examples are the B2Find catalogue and the Globus software for data transfer. Although we cannot decide on the sustainability of the software we rely on, our choices as a community can influence their own decisions. In the case of non-profit associations or communities, choosing solutions adopted widely, contributing to their sustainability by citing them properly, disseminating

the potential of the solution adopted and when necessary, contributing financially. In the case of companies, it is usually not possible to foresee the sustainability of a commercial solution. Companies make their choices based on profit, so in this case the only consideration that can be made is whether the commercial solution is the best possible one, and in many cases is because the service provided is of good quality and ready to use.

### **3.5 Financial sustainability**

This is one of the most obvious aspects of the sustainability. All the processes leading to the provision of FAIR and open data have costs of development, that have been in a good part co-funded with European and RI funding, costs of adoption of the tools and services, when the RI has not adopted them from the development stage, costs of implementation, operation and maintenance to keep them up and running, and finally costs for further developments and upgrades. The costs that need to be sustained after the adoption of the outputs are briefly explained in the following points.

#### **3.5.1 Costs of development and upgrade**

As mentioned earlier, the provision of FAIR and open data was not a common practice until some years ago in PaN facilities, with few exceptions. For this reason, some data management processes, infrastructure and services to link to EOSC had to be developed from scratch or at least involved upgrading the existing procedures and services. The costs of these developments are the typical object of competitive projects, where co-funding is made available to RIs to steer these processes. However, projects involve a limited partnership and those who are not lucky enough to become beneficiaries, need to sustain the full cost of their developments. As a good practice, European funding requires that the outputs are made available to the community, therefore we can assume that the costs for major developments usually benefit from co-funding.

For PaNOSC outputs, future developments if necessary will be explored in joint activities involving PaNOSC and ExPaNDS partners, mostly through LEAPS and LENS.

#### **3.5.2 Costs of adoption or implementation**

The outputs of projects or collaborations, even those made available openly and for free, usually require some technical work for their implementation. Plug and play solutions are rare. This is also the case for PaNOSC outputs. The costs of adoption of the outputs were mentioned as an independent point before, although they could be seen as belonging to this section. However, we believe that the commitment for adoption is fundamental and comes as the first step in the discussion for a set of sustainable outputs. If there was no commitment for adoption by many of the facilities, then a good part of the following considerations would make no sense. We can see the financial sustainability as all the investments that are required after the outputs are adopted, to maintain the same level of

service to the users.

The costs for adoption were described in section 3.1 above, and the conclusion was that after an estimation of the effort involved, most of the PaN facilities were either already in the process of adopting them or were willing to do so. With the consensus on the adoption by the facilities, we can focus on the other factors affecting the sustainability.

### 3.5.3 Costs of operation and maintenance

Once adopted, the outputs need to be operated. There are costs associated with licenses, computing capacity, network, software development and others. A more extensive list of costs is described in a previous deliverable, D7.2<sup>11</sup>. Projects usually aim at new developments, but once services and products are developed they need to be operated. It happens frequently that when entering a project, beneficiaries don't have a clear view of the future operation costs of the outputs, this is one of the trickiest points. In the case of PaNOSC, although some of the costs can be estimated (e.g. for running the services for the users of the facility), the demand by external users, more related to the EOSC, is completely unknown. For this reason, the efforts presented here are in most cases the pure **costs** (expressed either as effort in personnel -FTE- or costs) **of maintaining the outputs functional, not the full operation costs derived from them**. Deliverable D7.2 contains complementary information on the estimation of the total costs of operating services.

WP7 interviewed all WP leaders and the coordinator, to estimate the costs of operating and maintaining PaNOSC outputs. The estimation of the resources needed was made taking into consideration the following two categories:

- Personnel costs
- Other costs

We opted for expressing the **personnel costs** as effort in terms of full-time equivalents (FTE), due to the difference in the salary level across countries and facilities. The FTE was considered appropriate for the estimation of operation and maintenance efforts since they are planned as annual and generally constant, under the conditions explained above (no data available to estimate the efforts required to serve users outside the community). For some of the outputs, an effort is required to maintain the local instance running, while for others it is also required to maintain the federated service. In the case of local operation and maintenance, the efforts are expressed as FTE per year per facility. The common effort of all PaNOSC facilities to maintain certain outputs is expressed as "Common FTE per year".

The non-personnel costs ("**other costs**") usually relate to fees to service providers and maintenance of the computing hardware and software. For marketing, communication and dissemination, communication material such as brochures and videos, the publication of news in specialized magazines or social media campaigns, eventual implementation of

---

<sup>11</sup> PaNOSC D7.2 - Photon and Neutron EOSC metrics and costs model DOI 10.5281/zenodo.6010456.



subpages at RIs websites, participation to events and fairs, etc.

The estimation of the resources needed to sustain PaNOSC outcomes in the long-term follows. In all cases, the estimation corresponds to one year (yearly costs/efforts). The abbreviation NA means "Not Applicable", meaning that there is no cost associated, meanwhile the abbreviation TBD means "To Be Determined", meaning that at the current stage it's not possible to do an estimation.

### 3.5.3.1 FAIR and open Data Policy harmonization

Current Facilities' policies should be updated to better align with the FAIR principles. PaNOSC delivered a Data policy framework that can help facilities in this task. To adopt or modify existing data policies for Open Data, specifically mentioning the FAIR principles, facilities will need to undergo technical and management consultations. This process had started during the project, but if we intend to maintain the alignment of Data Policies at all PaNOSC sites, a periodic coordination meeting would be required. The same is true for Data management plans (DMPs). The current proposal is to have this alignment discussed in the joint LEAPS/LENS meetings, held monthly, and invite all non- LEAPS/LENS members to join the discussion. A monthly meeting of few hours would serve this purpose, therefore the effort involved is below 0.1 FTE.

Harmonisation of the FAIR and Open data policies and Data management plan		
FTE per Facility	Common FTE	Other costs
< 0.1	NA	NA

Once facilities have adopted Data policies for FAIR and open data, there will be a cost of implementation that is highly dependent on the status of the facility. These costs are not considered here but in "costs of adoption". What is included here is any small modification in the technical implementation that could be necessary to comply with the harmonization of the data policy and DMPs at PaNOSC sites, during time.

### 3.5.3.2 Standardized metadata and data catalogues

Tasks related to the creation and maintenance of ontologies, keywords and controlled vocabularies, as well as the data curation related to the catalogue will be necessary after the end of the project. Another activity to consider for future sustainability is pushing data out to third party repositories such as B2Find<sup>12</sup> or OpenAIRE<sup>13</sup>. B2find is putting the metadata under a "PaN community" heading, where in addition to the standard DataCite schema we can include information about instruments and make it searchable. The dialogue with them is in the beginning, since some PaNOSC partners only started to make their data

<sup>12</sup> <https://b2find.eudat.eu/site/about/>

<sup>13</sup> <https://www.openaire.eu/>

findable towards the end of the project. The advantage of relying on B2find is that it is a service that has existed for some time and make data visible outside the PaN community but has the drawback that the only data that can be pushed out is the one released from the embargo. The other problem is the richness of metadata, where the schema is not always relevant to PaN data. This restricts at the same time the search possibilities for PaN data through B2Find and OpenAIRE.

The ambition is to run PaNOSC search at the facilities, where there is also a federated service that is a one-stop shop to query all the databases at the facilities across the board. This requires data curation at the facilities, running the API and making sure that everything interacts upstream. The PaNOSC search API should work as a service for the local data repository.

In terms of effort per facility, it requires a significant amount of work to identify the search terms that make sense for the community and the facility (implementation costs), but for maintenance, 0.5 is enough.

We expect that after the end of the project, there will be the need to discuss periodically metadata schemas among PaNOSC partners and with third parties to continue pushing data with a meaningful schema, that allows for specific searches.

- 1) Maintenance of ontologies, keywords, controlled vocabularies and data curation;
- 2) Pushing data to third parties (Be2find/OpenAIRE), discussing metadata schemas;
- 3) Federated Search API for PaN data catalogues

FTE per Facility	Common FTE	Other costs
0.5		NA

### 3.5.3.3 Open Data portal

The Open Data portal would make it possible to search and download data. It requires an integration with AAI for the data under embargo: only the users authorized should have access to the data in the embargo period, while all data after the embargo should be accessible to all. For the data portal to work, it requires that all facilities have data curation and have implemented their local metadata catalogues and data repositories, where users can upload their data as well.

Open Data Portal		
FTE per Facility	Common FTE	Other costs
0.5	0.5 – 1*	NA

\* 1 FTE would be needed in case there would be a centralized data portal offering support. This is just for maintenance. For new developments, additional funding would be needed.

### 3.5.3.4 Software catalogue

The software catalogue was developed initially in PaNdata ODI<sup>14</sup>, a previous Horizon 2020 European project with a partnership of many PaN facilities. PaNOSC developed further functionalities for this catalogue for the PaN community, that collects an overview of software available for neutron and photon experiments and provides essential information, e.g. on their use with respect to instruments at experimental facilities. By registering and logging-in new software can be entered by PaN facilities, for it to be findable by users. This allows to create and enrich continuously the database beyond the single facilities contribution.

The success of this output depends on the engagement of all facilities in providing and maintaining the information about the software they develop. The effort needed to contribute at the level of the facilities can be considered limited, being around 1 PM/year for each. This output was not discussed in other project documents because it was not a deliverable, however the partners found it useful to maintain it. Towards the end of the project, when the EOSC portal started becoming functional and PaN facilities started onboarding their services there, the question emerged, whether the catalogue still provides added value or if it is a duplication of efforts. This aspect will become clearer when partners become more familiar with the onboarding process and the services provided by the EOSC. In the meantime, we include it as one of the potential outputs to be maintained.

Software catalogue		
FTE per Facility	Common FTE	Other costs
0.1	0.1 (overall)	NA

During PaNOSC the catalogue was hosted and maintained by ILL, but different agreements may be needed in the future.

### 3.5.3.5 Community AAI UmbrellaId

One of the main objectives of PaNOSC is to connect the Photon and Neutron RIs to the EOSC. In order to integrate the PaNOSC services into the EOSC a necessary step was to implement an AAI (Authentication-Authorisation-Infrastructure) for the community but taking into account the compatibility with the future EOSC AAI. Having a common AAI allows users of all PaN facilities to connect seamlessly to digital services with a single common ID. UmbrellaId was developed in previous European projects, starting in FP7 IRUVX<sup>15</sup>. During PaNOSC, Umbrella ID was migrated to EduTeams<sup>16</sup> (via GÉANT). After PaNOSC, it will be necessary to subscribe a service agreement for the maintenance and support. The support provided by GÉANT can be of different levels: we considered the costs of the basic service

<sup>14</sup> <http://pan-data.eu/PaNdataODI>

<sup>15</sup> <https://cordis.europa.eu/project/id/211285>

<sup>16</sup> <https://eduteams.org/>

and a more expensive one, where the support would be provided also e.g. during the weekends. The choice is with the facilities, but considering that most of them operate 24/7, it is likely that the most expensive option is the only one that covers the need of these facilities. The technical upgrades will be provided by GÉANT, as well as the helpdesk.

PaN community AAI Umbrella based on eduTEAMS.		
FTE per Facility	Common FTE	Other costs
0.1	NA	Subscription fee: 10 k€-25 k€

### 3.5.3.6 Remote Data Analysis Services

PaNOSC created and provided a data analysis services via VISA enabling users to remotely analyse data from PaN facilities during or after the experiment. Regarding the deployment of VISA service, CVMFS (CernVM-File System) has been tested as a scalable, reliable and low-maintenance software distribution service to deploy software on a distributed computing infrastructure used to run data processing applications. The goal is to have CVMFS shared federated among the RIs, in this way any facility can use software developed by other facilities, expanding exponentially the capabilities to analyse data. The use of CVMFS will have a minimum cost for the facilities. Currently, facilities' users will have access to local instances of VISA guaranteed, but It's desirable to have this service available to other users in the future though an expansion of the capacities. It's foreseen that this capacity could be provided by a third party such EOSC or EGI, but the costs have not been estimated because the demand and the support provided by EOSC are not known.

The effort needed to maintain VISA may change across facilities since it depends on the backend of the facility. Therefore, the estimation is provided as a range.

Virtual Infrastructure for Scientific Analysis (VISA)		
FTE per Facility	Common FTE	Other costs
1-2	1 *	

\*The common FTEs will depend on future agreements and will condition the future developments. In case this value is 0, only the local instances will be maintained.

### 3.5.3.7 The Virtual Neutron and x-ray Laboratory (ViNYL)

ViNYL offers 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.

The ViNYL<sup>17</sup> virtual lab consists on a series of tools that include upgrades of open source software such as McStas<sup>18</sup>, which is world leading in simulation of instrumentation for

<sup>17</sup> [10.5281/zenodo.6562105](https://zenodo.org/record/6562105).

<sup>18</sup> [10.5281/zenodo.4247598](https://zenodo.org/record/4247598).

neutron scattering and virtual neutron scattering experiments, OASYS<sup>19</sup> open-source graphical environment for x-ray virtual experiments and SimEx<sup>20</sup>, a python library to facilitate setup, execution, and analysis of simulations of experiments at advanced laser light sources. During the project, several adaptations or improvement of this software were made, also in the direction of making it more user friendly.

The service is ready, from the technical point of view, to be offered to any kind of users and to be onboarded in EOSC. However, due to the computing capabilities of the Facilities, currently it is only considered as a service for PaNOSC's facilities users. This may change in the future if the boundary conditions change. These eventual costs will add as operation costs, not maintenance.

The efforts needed to maintain the virtual lab were expressed in terms of cumulative FTEs. At the moment this deliverable it is written, it has not been defined if all facilities will contribute to further developments or just some of them. This effort can be split into many contributors, but it is advice that one of the facility takes care of the coordination, with a higher effort (at least 1FTE). Also, 4 FTE is the minimum effort to keep the current software versions running, but it is likely that upgrades and further developments will be needed.

ViNYL Virtual Neutron and x-raY Laboratory		
FTE per Facility	Common FTE	Other costs
*	4-7*	NA

\* 4 FTE overall (cumulative effort from all facilities) for maintenance. For any further improvements e.g. user friendliness, at least 3 FTE will be needed.

### 3.5.3.8 PaN learning and training platform

The PaN e-learning platform hosts free education and training for scientists and students on topics related to PaN science. It collects online interactive courses on the theory of photon and neutron science, how to use python code or software for data reduction and modelling, schools, data stewardship courses, etc.

Both Jupyter and OASYS have been integrated into the platform, which can be accessed via a single sign-on using the federated AAI, UmbrellaID.

The long-term sustainability of the platform depends on two factors: the facility hosting it and the collective effort to develop and upload training material. As with the other outputs, the efforts reported here are only these ones. Simulation and data analysis tutorials will require in addition computing resources, as part of the operation costs, but those are not included in the tables and may vary according to the demand. Just to give an indication, these variable yearly costs may amount to, according to the number of simultaneous users:

<sup>19</sup> [10.5281/zenodo.7371214](https://zenodo.org/record/7371214).

<sup>20</sup> [10.5281/zenodo.4249614](https://zenodo.org/record/4249614)

- 100 simultaneous users: 3.000 – 4.000€ (on-premises) / 7.500 – 10.000 € (cloud).
- 1000 simultaneous users: 30.000 – 40.000€ (on-prem) / 75.000 – 100.000 € (cloud)

These costs are not negligible and they clearly need to be supported by all PaN facilities. Maintaining and hosting of the e-learning training platform and its content will require an effort of 2FTE overall per year and few thousand euros for hosting and licenses.

PaN learning and training platform		
FTE per Facility	Common FTE	Other costs
1	2	5 k€

### 3.5.3.9 Communication and marketing

Communication, dissemination and marketing activities will be fundamental in the years following the project. When the project will end, all outputs will be complete and ready to be deployed, however, not investing the right effort in communication and dissemination may lead to a low response from researchers. This action is necessary for the long-term sustainability of PaNOSC outcomes, all partners will need to contribute to the development of materials, but ideally one facility should be coordinating the effort.

The effort estimated per year per facility is 1-2 FTE plus 1-2 FTE for the coordinator/s, who should also maintain the link with the other clusters. Other direct costs related to hosting the website, preparing and delivering material, licenses, organizing workshops and schools, etc. can be estimated in 50.000-100.000€, depending on the amount and type of activities.

Communication, dissemination and marketing		
FTE per Facility	Common FTE	Other costs
1-2	1-2*	50-100 k€

\*Ideally, one facility coordinating the effort

## 3.6 Governance

The governance is the most important aspect influencing sustainability. In many of the interviews, it emerged that beyond the resources required, a clue role would be that of the "coordinator" e.g. of the developments. The question "who will decide?" was a concern across work packages. During the execution of the project, partners followed the description of work and the lead of the Work package leaders, to execute the tasks. Many of the outputs were tested as pilots or proof of concepts, therefore there was also the possibility to volunteer or to step out from an implementation in an initial state. But with the project coming to an end, this context changes: the implementation of the PaNOSC solutions should be in full scale, it should involve all PaN facilities, there is no formal occasion to meet and discuss (e.g. periodic WP meetings), etc. All these changes and rising needs must be

managed, and the structure managing these decisions, steering the direction for new developments, etc. does not exist.

Create a legal entity that would manage PaN facilities' data like an ERIC<sup>21</sup>, AISBL<sup>22</sup>, etc. The possibilities are many. The ERICs have the advantage of having the states and the international organisations as members. The ERIC is an excellent vehicle to manage pan-European efforts, is inclusive (intergovernmental organizations and non-EU countries can be members as well). Having the management of data as core business would allow these facilities to operate efficiently and deliver results that would be hard for PaN facilities, whose core objective is to run the best machines. Examples of these ERICs exist already for other disciplines (EPOS<sup>23</sup>, CESSDA<sup>24</sup>) and the results they have achieved are a proof that this legal structure can provide a successful setup. Linking resources across communities and disciplines, a closer interaction with industry, higher interoperability, efficient use of resources through scale-up are some of the advantages of creating an independent infrastructure to manage data. There are other legal vehicles that allow to reach the same goal. ELIXIR<sup>25</sup> is an intergovernmental organization that brings together life science resources from across Europe.

These communities saw the opportunity in a new legal entity. This opportunity was mentioned but apparently is far from being adopted by PaN facilities, who are still keen to have the full responsibility for their data lifetime. For this reason, this option was discarded at the moment, but it may be reconsidered in the future, if the PaN commons will evolve.

The alternative to a new legal entity is the subscription of agreements between facilities. These agreements can be either legally binding (e.g. collaboration agreements) or not, like MoUs (Memorandum of understanding). To subscribe a legally binding agreement does not offer many more guarantees than a simple MoU because it is very unlikely that in case of breach, some of the PaN RIs will take legal actions against another.

MoUs, although not legally binding are based on the good will of the facilities to cooperate for a common interest. It will be necessary that together with the MoU, a consortium agreement is signed, explaining how the decisions will be made, how IP will be dealt with, the election of the Chair, voting majorities, etc. The MoU alone is not enough to regulate the collaboration.

Having chosen the MoU as the more suitable way to proceed, during the lifetime of the project we considered different scenarios. Unfortunately, none of them could be agreed upon and shaped before the end of the project, but we expect that this will happen soon after. In the following chapter, we summarize the scenarios that we considered and provide a short analysis of their pros and cons.

---

<sup>21</sup><https://eur-lex.europa.eu/EN/legal-content/summary/european-research-infrastructure-consortium-eric.html>

<sup>22</sup> [https://justice.belgium.be/fr/themes\\_et\\_dossiers/societes\\_associations\\_et\\_fondations/associations/aisbl](https://justice.belgium.be/fr/themes_et_dossiers/societes_associations_et_fondations/associations/aisbl)

<sup>23</sup> <https://www.epos-eu.org/>

<sup>24</sup> <https://www.cessda.eu/>

<sup>25</sup> <https://elixir-europe.org/about-us>



## **4. Strategies for long-term sustainability of the PaN EOSC**

The long-term strategy for the PaN EOSC must factor all the strengths of working together for this common aim, the PaN commons, and the threats to the sustainability mentioned before. We discussed different strategies during the project, also in a broader context (e.g. ERF). We took strongly into account the opinion of work package leaders during the interviews. We asked them what would make more sense in order to maintain a federated service: a centralized maintenance (by one or few partners) running the service for all, or a distributed effort, equally divided amongst the partners. The replies vary depending on the outputs or services, but in general there was a tendency to prefer that a partner took over the maintenance of a federated service for all the partners.

The original proposal by WP7 was to develop a mid-to-long-term strategy for the PaN EOSC involving PaNOSC and ExPaNDS facilities, and any other relevant service provider like for example EGI. This strategy would have contained:

- The vision, or direction in which PaN facilities should be moving in terms of good practices for FAIR data management and preservation;
- The requirements for a PaN EOSC, interoperable with the EOSC, inclusive and user friendly;
- The dependencies on external communities and services and required actions towards third parties, to contribute to the sustainability of the required services;
- A list of the outputs to be maintained, including PaNOSC and ExPaNDS KERs, and any other useful tool or service;
- The proposal for strategies to be adopted to maintain these outputs.

Such a document would have provided a complete overview of the resources needed but also the new services adopted, and their added value. Without a comprehensive overview it is difficult for any manager to make decisions on the allocation of resources. The final goal was to achieve the signature of an MoU, subscribed by most of the facilities, even though they may not have contributed evenly. After discussion with the project coordinator and the PaN partners in LEAP and LENS it was decided to continue the work to sustain vision and the outcomes via the two organisations which represent all the PaN facilities in Europe. The two PaNOSC facilities (CERIC-ERIC and ELI-ERIC) which are currently not members of either of these organisation will be considered as strategic partners and will be invited to participate in all activities related to the KERs, FAIR data and the EOSC. The overall goal of producing FAIR data and contributing to the EOSC will be represented by the LEAPS and LENS initiatives. In November 2021 LEAPS published a Data Strategy white paper committing to this vision. By adopting a pragmatic approach to sustaining the KERs separately each facility can decide to sustain and contribute to the KERs which they are interested in. Where possible an MOU will be setup around a KER or group of KERs e.g. for

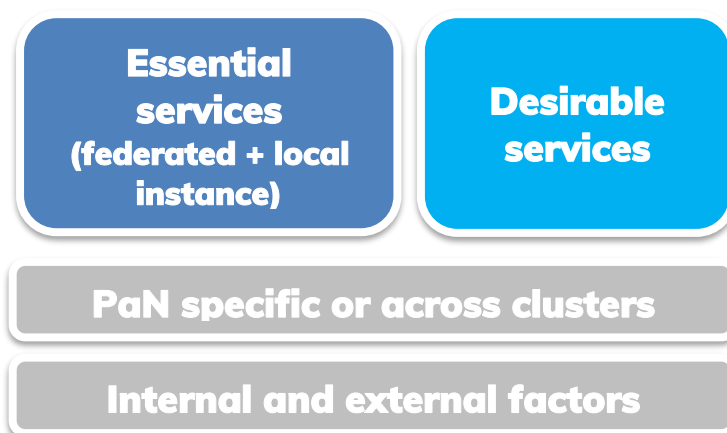


VISA, the remote data analysis platform.

We will summarize in the following examples the scenarios and the strategies that we identified.

### Scenario 1: Identification of essential services to be adopted by all PaN RIs

This scenario is based on the consideration that a wide adoption of some of the outputs by PaN facilities contributes to their added value, the most efficient use of resources and a higher adoption by users ("single tool" from the user's perspective).



In this scenario, it is required that the outputs are analysed and divided in the two groups "essential services" and "desirable services". Essential services are those that would need to be adopted by most if not all the PaN facilities, to maximize their value. These essential services would be the Open data portal with search API, the Community AAI UmbrellaID and the harmonisation of the FAIR and Open

data policies and Data Management Plans. Some tools and services could be considered in this category but based on a further consideration, like the cost of maintenance. In the previous section we presented these costs. The training and learning portal, ViNYL and VISA have very high costs for the facilities hosting or maintaining/developing them. From the point of view of the cost, for these outputs it is essential that an arrangement between facilities is found and that they are maintained by a large number of RIs, however it is not essential that all RIs or most of them adopt these tools and participate in the joint maintenance for them to be meaningful. For the training and learning portal, as well as for the data catalogue, it is fundamental that PaN RIs contribute to the contents, in the first case with training and learning material, and in the second declaring and maintaining up to date their list of software.

The distinction between essential and non-essential outputs should be made with wide consensus, since the essential services should be adopted and maintained by most or all PaN Facilities. Tools and services that are not essential should be maintained because they were developed with effort, but as long as a small number of facilities agree to do so, and that they can guarantee the resources to maintain them, then this should be good enough.

PaNOSC tools and services that have a dependency on external factors should include in their agreements actions towards these external communities, in order to secure or at least contribute to the sustainability of the services and tools necessary for the PaN EOSC. The

agreements should also consider if some of the services can be adopted by a wider community (outside PaN) and include a proper outreach to these.

### Scenario 2: Discussion of the single outputs independently.

The second scenario does not include any kind of concerted action. Whoever is the main driver of an output shall endeavor to persuade the other IT colleagues, and these in turn should reach out to their managers to obtain the necessary commitment. This is the approach that emerged naturally towards the end of the project, for some of the outputs. This approach has the weakness of providing a partial picture, without a comprehensive understanding of the level of service that should be offered to users by any pan-European ESFRI RI or ERIC (as PaNOSC partners). For the first MoUs it may seem easier because commitments are reduced, if compared to the commitments to maintain all essential and non-essential outputs. However, we can expect that in a context of limited resources, the first ones to come will be the best served. Another drawback is the complexity of managing a series of MoUs between different parties and with completely different conditions. As mentioned before, agreeing on resources does not solve the issue of governance, and for joint developments this is critical. The introduction of different configurations and conditions in many agreements only increase the complexity of the governance of these outputs, and does not lead towards convergence of services.

WP 7 would not recommend partners to follow this approach, but to base the proposal for sustainability on Scenario 1.

## 5. Conclusions

Different approaches can be adopted to make PaNOSC outputs sustainable. Based on the experience and exchanges in the Sustainability work package of this project, the strategy proposed was that in which all the outputs would have been considered together in a holistic approach for RIs to offer quality services to their users and a wider community, applying best practices and fulfilling their commitments and obligations as institutions operating mostly with public funding. This approach was considered difficult to implement and we could not achieve to start the discussion at the level of RIs management during the lifetime of the project. Some PaN partners proposed to strive for a more granular approach securing the sustainability of tools and services gradually. There is a MoU signed by several PaN RIs ensuring the sustainability of the AAI (UmbrellaID) and one at an advanced status of discussion for VISA. Core discussions are taking place in the frame of the LEAPS and LENS initiatives. Three of the seven PaNOSC partners are not members of neither these initiatives (EGI, ELI ERIC, CERIC-ERIC) but were invited to participate regularly in the meetings as observers. The further development of some of the outputs is being explored as part of future calls (Horizon Europe), however the question of operations, especially for generic users of the EOSC, still remains.

### Acknowledgement

WP7 members are most grateful to the WP leaders and coordinators of PaNOSC and ExPaNDS projects, that provided valuable information on the sustainability issues and considerations of their WPs several times during the execution of the project.