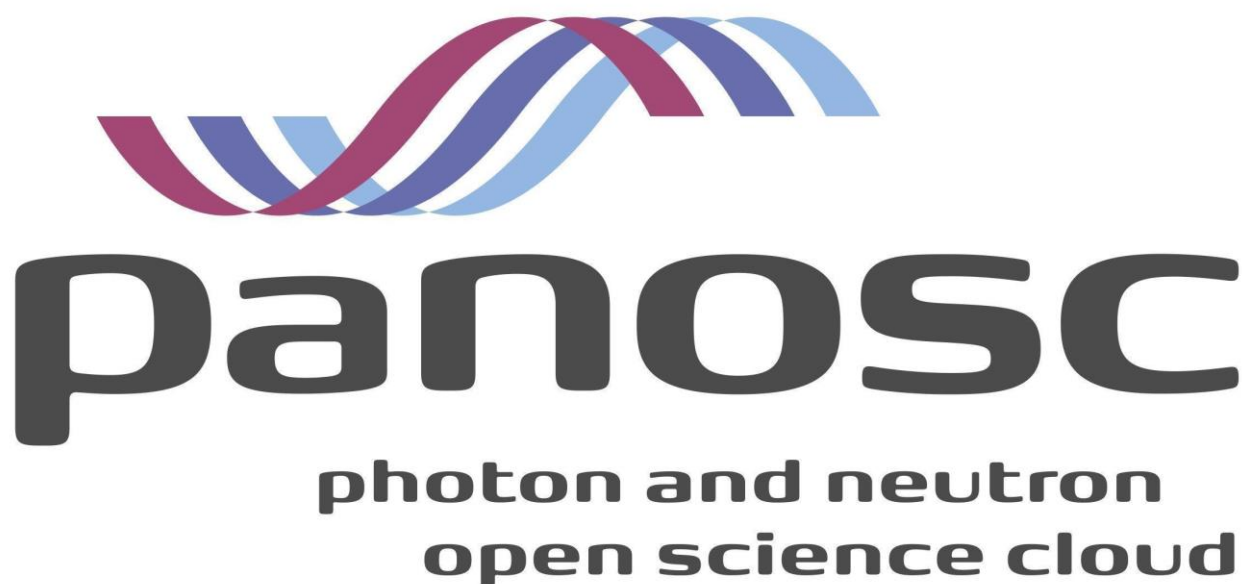


PaNOSC
Photon and Neutron Open Science Cloud
H2020-INFRAEOSC-04-2018
Grant Agreement Number: 823852



Milestone: MS2.2 - Adoption of PaNOSC DP framework

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.panosoc.eu
Deliverable No:	MS2.2
Deliverable Type:	Milestone
Dissemination Level	Public
Contractual Delivery Date:	01/11/2020
Actual Delivery Date:	12/02/2021
EC project Officer:	Rene Martins

Document Control Sheet

Document	Title: Adoption of PaNOSC DP framework
	Version: 1.0
	Available at: https://drive.google.com/file/d/1UVA810yReYjx-jw8cS5Z_DW3DWICK1tX/view?usp=sharing
	Files: 1
Date	15/02/2021
Authorship	Written by: A. Götz (ESRF), D.Roccella (CERIC-ERIC)
	Contributors: S.Caunt (ILL), E. le Gall (ILL), K.Wrona (EuXFEL), J.Malka (EuXFEL), J.Taylor (ESS), T.Ivānoica (ELI-DC), F.Gliksohn (ELI-DC), J-F.Perrin (ESRF)
	Reviewed by: R.Dimper (ESRF)
	Approved: J.Bodera Sempere (ESRF)

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 (EuXFEL)	Germany
4	The European Spallation Source (ESS)	Sweden
5	Extreme Light Infrastructure Delivery Consortium (ELI-DC)	Belgium
6	Central European Research Infrastructure Consortium (CERIC-ERIC)	Italy
7	EGI Foundation (EGI.eu)	The Netherlands

Table	of	Content
Adoption of PaNOSC data policy framework		5
ESRF		6
Changes		6
Process		7
EuXFEL		8
Changes		8
Process		9
ESS		10
Changes		10
Process		11
ILL		12
Changes		12
Process		13
Tools		14
CERIC-ERIC		15
Policy		15
Process		15
ELI		17
Policy		17
Process		17
Conclusions		18
Annex - Main Differences between PaNOSC and PaNdata data policies		19

Adoption of PaNOSC data policy framework

In May 2020 the PaNOSC data policy framework was delivered to the European Commission as a deliverable of the PaNOSC project. It was published on [zenodo](#) (DOI 10.5281/zenodo.3826040) and is also accessible via the [PaNOSC website](#). Now, after two years from the beginning of the project, ELI and CERIC are going to adopt the data policy and the other partners are adapting their existing policies to be compatible with the PaNOSC Data Policy framework. The scope of this document is to provide a clear picture of the status of this process and what the next steps are for each partner in order to get the data policies endorsed by their management.

ESRF

ESRF has done an analysis of the differences between its existing data policy and the PaNOSC data policy framework and identified the changes they would like to include in an updated ESRF data policy.



Changes

The main changes between the current and updated data policies are:

ESRF Data Policy (2015)	Updated ESRF Data Policy (2021)
FAIR concepts not mentioned	Explicitly mention FAIR and the objectives
Reduced or compressed data not mentioned	Explicitly mention the possibility of storing reduced/compressed data as raw data
Processed data are not included (mention is made of curation of results on a best effort)	Explicitly open the possibility of storing processed data covered by the data policy
Auxiliary data not mentioned	Define and mention auxiliary data
Electronic logbook not mentioned	Explicitly mention the electronic logbook as part of the metadata capture
ORCID not mentioned	Explicitly mention ORCID as a means of linking users to data DOIs
Termination of data custodianship not mentioned	Explicitly include data custodianship termination clause
Data format is not explicitly mentioned	Explicitly mention that the preferred data format is HDF5
Only metadata from ESRF software accepted	Allow the possibility of metadata from non-ESRF software especially for processed data
Persistent identifiers (PIDs) mentioned in general	Explicitly mention use of DOIs as PIDs
Granularity of PID is experiment and dataset	Explicitly mention PID can refer to a bespoke collection of datasets in addition to experiment and dataset automatically generated
No explicit mention of data under embargo being available for AI/ML use by the facility	Explicitly mention that data under embargo can be used for AI/ML by the facility
No formal process for making changes to the Data Policy	Define an internal process for making minor changes to the Data Policy

The changes are considered to be not major and should not require the full process of requesting the ESRF Science Advisory Committee (SAC) and Council to approve them. Therefore the following light process is envisaged.

Process

The steps to get the updated data policy accepted at the ESRF are as follows:

1. Prepare the new data policy document together with a document describing the changes of the updated data policy with respect to the existing one (February 2021)
2. Present both documents to the Data Management Working Group (DMWG) at the ESRF, which deals with issues around data with the scientists, for comments and feedback (February/March 2021)
3. Present the outcome of the DMWG to the ESRF Directors and request them to endorse the changes (March/April 2021)
4. If the Directors nevertheless decide that the changes need to be endorsed by SAC and approved by Council then submit the changes to SAC and Council
5. Otherwise inform the SAC, Council, and users of the new data policy and publish the new data policy on the ESRF web site.

EuXFEL

The estimated time to adopt an updated Data Policy is about one year. The adoption process requires consultations with user communities, internal stakeholders, Detector Advisory Committee, and Scientific Advisory Committee (SAC), followed by approval of the Management Board and the European XFEL Council.



European XFEL identified differences between the existing “Scientific Data Policy for the European XFEL Facility” and the recommendations in the “PaNOSC Data Policy framework” (main differences included in the table below).

Changes

The main changes between the current and updated data policies are:

EuXFEL Scientific Data Policy (2017)	Updated EuXFEL Scientific Data Policy (2021)
FAIR concepts not mentioned	Explicitly mention FAIR and the objectives
Data Management Plan not mentioned	Define the Data Management Plan as mandatory
Reduced data not mentioned	Explicitly mention the possibility of storing reduced data as raw data
Processed data are not included in the long-term storage	Consider storing processed data for long-term
Some particular auxiliary data were defined (geometries, calibration constants)	Define general auxiliary data term, expand on this data and introduce rules to manage it
Electronic logbook not mentioned	Explicitly mention the electronic logbook as part of the metadata capture
ORCID not mentioned	Explicitly mention ORCID as a means of linking users to data DOIs
Data format is not explicitly mentioned	Explicitly mention that the preferred data format is HDF5
Only metadata from European XFEL software accepted	Allow the possibility of metadata from non-European XFEL software especially for processed data

Persistent identifiers (PIDs) mentioned in general	Explicitly mention use of DOIs as PIDs
Granularity of PID is experiment and dataset	Explicitly mention PID can refer to a bespoke collection of datasets in addition to experiment and dataset automatically generated
Personal Data Protection included but GDPR not explicitly mentioned	Make explicit reference to the GDPR regulations
Few good practices included	Expand on the "Good practices" according to the new recommendations

Process

European XFEL has received positive recommendation from the external committee (appointed by the Management of the European XFEL to advise on detector and data related topics) to align with the PaNOSC recommendations and we are in the process of consulting with user community on its implications (a dedicated session on "FAIR Data Management" has been organized on the occasion of the annual European XFEL User Meeting). Further consultations with internal stakeholders are required (Instrument Scientists, User Office, Legal group). Then after receiving a positive decision by the Management Board, the Scientific Advisory Committee will be presented with the new data policy for final recommendation and if positive the European XFEL Council will be asked for the approval.

ESS

The ESS policy for scientific data has been in place since approval by the ESS Council in 2017.

The data policy covers raw data & metadata and results from non-proprietary use of ESS neutron instruments. The policy describes privileged and open access to data and specifies the use of persistent identifiers for raw data sets.

In light of recent developments of data policy best practice from PaNOSC and ExPaNDS, the policy will be updated to reflect the movement towards FAIR and best practice for research data management.



Changes

The main changes between the current and updated data policies are:

ESS Scientific Data Policy (2017)	Updated ESS Scientific Data Policy (2021)
FAIR concepts not included	Explicitly mention FAIR and the objectives
Data Management Plan not included	Requirements of the Data Management Plan added and becomes a mandatory part of the policy
Processed data and metadata not included	Add policy for processed data and metadata.
Auxiliary data not included	Policy aspects for auxiliary data added
Electronic logbook not mentioned	Explicitly mention the electronic logbook as part of the metadata capture
ORCID not mentioned	Explicitly mention ORCID as a means of linking users to data DOIs
Termination of custodianship not mentioned	Explicit description of how ESS can formal cease to be data custodian for scientific data
Persistent identifiers (PIDs) mentioned in general	Explicitly mention use of DOIs as PIDs

Data policy written as a static instance	Mechanism for policy updates defined
GDPR compliance not explicitly included	GDPR compliance statement added
Policy defined without supporting implementation notes or guidelines documentation. This is an inflexible approach for implementation necessitating the formal change of the policy for any change to the research data management process.	Policy will be implemented with reference to supporting procedure guidelines documents to define ESS best practice and research conduct ethics consideration relevant for FAIR. This will provide a documentary process with flexibility for implementation and future process alteration without the need for formal changes to the policy.

Process

The ESS policy will be updated in 2021 to address the aspects described in the preceding table.

ESS will develop supporting guidelines and implementation notes to allow the organisation a greater degree of flexibility in further data policy developments. These will include the research code of conduct from [The European Code of Conduct for Research Integrity](#), requirements for open access publications and recommendations for FAIR data within the photon and neutron domain developed by the ExPaNDS project (10.5281/zenodo.4312825). This process is expected to take 12 months and depends upon the ESS governance cycle.

ILL

After analysis of the differences between the ILL's existing data policy (latest version July 2017) and the PaNOSC data policy framework, several changes have been identified that should be included in an updated ILL data policy. In the column "Updated ILL Data Policy (2021)" reference is made whenever possible to the precise clauses which require the update.

Changes

The main changes between the current and updated data policies are:

ILL Data Policy (2017)	Updated ILL Data Policy (2021)
Auxiliary data not mentioned	3.1.4, 3.6 Define and mention auxiliary data
Embargo period not explicitly defined	3.1.9 Explicitly define the meaning of embargo period
Data Management Plan (DMP) not explicitly defined	3.1.15 Explicitly define the meaning of DMP
FAIR concepts not mentioned	3.2.5 Explicitly mention FAIR and the objectives
GDPR is not mentioned	3.2.8 State that compliance of metadata and data at the facility is ensured by the facility
The non-liability of the ILL in the event of unavailability or loss of data is not mentioned	3.2.9, 3.2.10 Non-liability in the event of unavailability or loss of data will be stated
It is not mentioned that it is the responsibility of the PI to ensure collected metadata meets standards, only that the proposal number is correctly entered at the start of an experiment	3.4.3 Explicitly state that the PI should ensure metadata collected meets minimum requirements
There is no mention of data and metadata concerning user supplied equipment	3.4.7 Mention that data and metadata concerning user supplied equipment must be provided to the facility for curation
The archiving of data is not mentioned	3.4.8 Explicitly state that data will be copied to archival facilities for curation
Granularity of PID is proposal	3.1.17, 3.4.11 Explicitly mention PID is down to the level of a dataset and can refer to a

	bespoke collection of datasets
There is no mention that the facility does not guarantee the readability of user-generated processed data.	3.5.4 Explicitly state that the facility does not guarantee readability of user-generated processed data.
Metadata relating to the provenance of processing carried out is not mentioned	3.5.5 Explicitly state that the user must include metadata to describe provenance of processing
The interoperability of metadata for processed data is not mentioned	3.5.6 State that metadata for processed data should be interoperable
Termination of data custodianship not mentioned	5 Explicitly include data custodianship termination clause
No minimum duration of the curation is stated	Minimum duration of curation stated
Access to data via APIs is not mentioned	Access to data via OAI-PMH and PaNOSC search API explicitly specified
ORCIDs not mentioned	Explicitly mention ORCID as a means of linking users to data DOIs
Data format is not explicitly mentioned	Explicitly mention preferred data format is HDF5
Good practices of experimental team in order to reproduce results are not mentioned	4.2, 4.5 Mention that the experimental team should provide a complete log of the process carried out that can at least be linked to the data, and that they should link software used to obtain results and preferably make this openly accessible.

Process

1. Produce a draft update of the ILL Data Policy to include the changes relating to the PaNOSC data policy framework and produce a supplementary document highlighting the changes to the existing one.
2. The documents should be reviewed by the members of the DPP (Data Protection and Processing group) that includes representatives from relevant technical and scientific groups, as has been the case for previous updates to the ILL Data Policy.
3. The changes to the data policy and the outcome of the DPP review process will be submitted to the Management Board (MB) of the ILL.
4. If the MB agrees to the changes then the updated Data Policy will be adopted and published on the ILL website.

Tools

Certain changes to the policy impact the tools used to manage data at the ILL and technical efforts will be required to implement them. These include:

1. The creation of PIDs down to the level of datasets (and arbitrary collections of datasets).
2. The creation of PIDs related to software (including different software versions) to provide metadata relating to the processing of data.
3. A full implementation of the PaNOSC search API providing, at a minimum, access to open data.
4. The development of data archival procedures and mechanisms.

CERIC-ERIC

CERIC

CERIC-ERIC did not have a data policy before its involvement in the PaNOSC project. The first draft data policy, based on the PaNData model, was discussed by the BoD¹ in the beginning of February 2019. This first version has been applied only to the data generated by the pilots of an EU-funded project (ACCELERATE).

Policy

In the context of WP2 of PaNOSC, a new data policy framework has been developed and CERIC started to implement the second version of its policy by taking into account:

- the FAIR data management principles described in the OECD's "Principles and guidelines for access to research data from public funding";
- the PaN data guideline for a common Scientific Data Policy;
- the PaNOSC data policy framework;
- the National Data Policy of the country in which every Partner Facilities (PFs) is located (where present);
- the results of the interactions between the CERIC Directors and the CERIC IT team.

Process

At the end of March 2019, the new data policy has been presented to the BoD where it has been decided that:

- the custodian of data (raw data and associated metadata) will be the PF;
- there will be a transition period until the infrastructure will be developed. The DMP will need to address the open issues, e.g for long term storage;
- the new data policy will be enforced by the end of 2022.

With this positive result from the Board of Directors, the data policy was submitted for the approval by the General Assembly in June 2019. From that moment, CERIC worked with individual facilities, trying to assess their needs for the enforcement of the data policy, as PaNOSC WP2 developed the framework for the data policy, compliant with the FAIR principles as assess with the RDA's most recent draft of the guidelines for FAIR data maturity. CERIC is now completing an overview of the current standards and procedures adopted by the facilities for data acquisition and data management. This will contribute to identifying what are the next steps and the investments that CERIC has to do toward the harmonization of the data management policy and procedures.

¹ Board of Directors of the Partner Facilities (BoD): Is composed by one representative of every Partner Facility and, together with the Executive Director, is the executive body of CERIC.

This work is in progress and, together with the development of the draft, CERIC is annexing a technical document covering the requirements for every CERIC Partner Facility to adopt the new data policy:

- long term storage costs;
- compliance with the national data policy or national initiative (if existing);
- collection and storage of the metadata;
- standard data format (HDF5 is what CERIC, and more in general the PaN community, is suggesting);
- electronic logbook.

The result of the assessment of the requirements, together with a proposal for the implementation (budget, timeline) will be presented for discussion at the BoD as soon as it will be ready.

ELI

The Extreme Light Infrastructure (ELI) consists of complementary facilities located in the Czech Republic, Hungary and Romania. The ELI facilities, built as individual construction projects, are now coming together as an integrated organization, the ELI European Research Infrastructure Consortium (ELI ERIC), that will be in charge of their joint operations. ELI ERIC is expected to be established in the first quarter of 2021. It will initially include the Czech Republic and Hungary as host members; Romania will join at a later stage when accession conditions have been agreed.



Under ELI ERIC, ELI will operate as a multi-site organisation with single management and single governance. Over the next 2-3 years, the ELI ERIC will work jointly with the ELI Facilities on the development of an integrated management system that should eventually allow ELI ERIC to directly operate the Facilities.

Policy

The General Assembly of ELI ERIC is in charge of approving the policies of the organisation, including the Data Policy. In this respect, Article 13 of the ELI ERIC Statutes reads as follows:

13(1) 'Data' refers to all information collected by users and the staff while performing scientific experiments under the access for users Policy and performing operations of the ELI Facilities.

13(2) Open Access to FAIR data sets and metadata stored in Open Access repositories shall be favoured for data collected as a result of the use of the ELI Facilities and, to the extent possible in case of software and computer programmes created by the ELI ERIC and the ELI Facilities; open source principles shall be considered.

The statutory policies, which form the cornerstone of ELI's future management system, are to be approved after the establishment of ELI ERIC. Policies are worked out by the management of the organisation, reviewed by advisory bodies and eventually approved by the General Assembly.

A first version of the ELI Data Policy has been drafted by the ELI management. It builds on the Data Policy Framework developed within the WP2 of PaNOSC and proclaims the commitment of ELI to abide by the FAIR principles in managing experimental data.

Process

The approval of the Data Policy requires the approval of the ELI ERIC General Assembly by qualified majority. As a first step, a proposal will have to be approved by the ELI ERIC Board of Directors, which will include the ELI ERIC Directors and Directors of the ELI ERIC Facilities. The opinion and recommendations of the Administration and Financial Committee (AFC) and the International Scientific and Technical Advisory Committee will be sought. The proposal will be finalised after this review process and submitted formally for approval by the ELI ERIC Director General to the ELI General Assembly. We expect the overall process to take a maximum of 6 months as of the establishment of ELI ERIC. The development of the Data Management Plan for all ELI experiments will be carried out in parallel.

Conclusions

The Milestone 2.2 of WP2 consisted of two steps:

1. Identifying the changes each partner wanted to make to their existing data policy to make it aligned with the PaNOSC policy framework and/or developing a new data policy in the case partners do not have a data policy.
2. Document and initiate the process for having the updated data policy ratified.

The PaNOSC WP2 members are not the decision makers, therefore the objective was to perform the analysis and prepare, start and follow up the process of adopting a new or updated data policy at each PaNOSC facility. As the WP2 members were the ones who prepared and produced the PaNOSC data policy framework they are well placed to do this work.

A large overlap in the envisaged updates at the facilities can be seen. This can be explained by the fact that the existing data policies were derived from the same original framework (PaNdata framework) and all members have committed to providing FAIR data.

The process for policy implementation (or changes to existing policies) for each partner is quite long requiring oversight from internal directors and the facility governance. . This is an indication of how serious data management and data policies are taken at the facilities. We hope that this exercise will serve to facilitate future updates at all partners sites. Data policies are not static but living documents that should be updated regularly to take into account changes in the facilities, communities and the scientific world at large.

The process at each partner site will be followed up with internal milestones to ensure that the new/updated data policies are all in place by the end of the project (end of 2022).

Annex - Main Differences between PaNOSC and PaNdata data policies

The following table summarises the main differences between the PaNOSC and PaNdata data policies:

PaNdata Data Policy Framework (2010)	PaNOSC Data Policy Framework (2020)
FAIR concepts not mentioned	Explicitly mentions FAIR and the objectives
Auxiliary data, beam time, users, embargo period, Data Management Plan, processed data and data object are not mentioned	Auxiliary data, beam time, users, embargo period, Data Management Plan, processed data and data object have been defined
The long term storage is defined as a period from 5 up to 10 years	The long term storage is not defined but the topic is mentioned in the implementation note (depending on the capacity/restriction of the facilities)
General principles not mentioned	The chapter "General principles" define the embargo period, the curation of the data, the licences and other good practices for users and facilities
DOI as Persistent Identifier not defined	Persistent Identifier defined (DOI) in chapter 3.3
The policy does not define who is the custodian and the curation of the datasets	The policy defines who is the custodian of raw data and associated metadata, processed data and associated metadata, auxiliary data and results. It also defines who is in charge of their curation
Electronic logbook not mentioned	Explicitly mention the electronic logbook as part of the metadata capture
Termination of data custodianship not mentioned	Explicitly include data custodianship termination clause