D2.3: Final Data Policy Framework for Photon and Neutron RIs

Document Control Information

Settings	Value
Document Identifier:	D2.3
Project Title:	ExPaNDS
Work Package:	WP2
Work Package Lead	UKRI
Document Author(s):	Abigail McBirnie (UKRI), Brian Matthews (UKRI), Brigitte Gagey (SOLEIL), Carlo Minotti (PSI), Daniel Salvat (ALBA), Frank Schlünzen (DESY), Andrei Vukolov (Elettra)
Document Contributor(s):	Alun W Ashton (PSI), Uwe Konrad (HZDR), Rolf Krahl (HZB), Sophie Servan (DESY)
Doc. Version:	1.0
Dissemination level:	Public
Date:	20/08/2021

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Abstract

This report considers how data policies for research infrastructures providing photon and neutron facilities (PaN RIs) should be framed, in particular in the light of supporting open science and FAIR data. It builds on and revises the initial policy framework presented in ExPaNDS deliverable *D2.1: Draft extended data policy framework for photon and neutron RIs*.

Over the period February – May 2021, ExPaNDS undertook a series of detailed consultation meetings with all ten of our partner PaN RIs. This report collects the themes that emerged most often and most strongly from those conversations. Based on the consultation feedback, we present a revised version of the ExPaNDS data policy framework, which now proposes a reduced set of 21 elements for RIs to consider when formulating data policy.

The consultation provided an opportunity to engage deeply in discussion with ExPaNDS partners about data policy and policy making. The level of engagement with the consultation process highlights ExPaNDS partners' strong interest in data policy as a powerful and important tool for national PaN RIs. There is clear consensus that having an appropriate data policy is fundamental to supporting good research data management and that flexibility in a data policy framework is vital to accommodating the diversity found across ExPaNDS facilities.

The ExPaNDS data policy framework is already having an impact on facilities' plans for future work, especially around ensuring that key implementations are in place to support FAIR data. Further work within ExPaNDS will provide guidance to support these activities.

Licence

This work is licenced under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit creativecommons.org/licenses/by/4.0/ .

Document Log

Version	Date	Comment	Author/Partner
Complete Draft	02/08/2021	Internal review version	Brian Matthews (UKRI), Abigail McBirnie (UKRI)
1.0	20/08/2021	Final version for submission	Brian Matthews (UKRI), Abigail McBirnie (UKRI)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Abbreviations and acronyms

CC	Creative Commons
DESY	Deutsches Elektronen-Synchrotron
DLS	Diamond Light Source
DMP	Data Management Plan
DP	Data Policy
EC	European Commission
EGI	European Grid Infrastructure Foundation
EOSC	European Open Science Cloud
EU	European Union
ExPaNDS	European Open Science Cloud (EOSC) Photon and Neutron Data Service
FAIR	Findable, Accessible, Interoperable, Reusable
FE	Framework Element
GDPR	General Data Protection Regulation
HZB	Helmholtz-Zentrum Berlin
HZDR	Helmholtz-Zentrum Dresden-Rossendorf
IPR	Intellectual Property Rights
ISIS	ISIS Neutron and Muon Source
ORCID	Open Researcher and Contributor ID
PaN	Photon and Neutron
PaNOSC	The Photon and Neutron Open Science Cloud
PI	Principal Investigator
PID	Persistent Identifier
PSI	Paul Scherrer Institute
QoS	Quality of Service
RI	Research Infrastructure
ROR	Research Organization Registry
STFC	Science and Technology Facilities Council
TFiR	Turning FAIR into Reality
UKRI	UK Research and Innovation
WG	Working Group
WP	Work Package

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Executive Summary

This report, ExPaNDS deliverable *D2.3: Final data policy framework for photon and neutron RIs*, considers how data policies for research infrastructures providing photon and neutron facilities should be framed, in particular in the light of supporting open science and FAIR data. It builds on and revises the initial policy framework presented in ExPaNDS deliverable *D2.1: Draft extended data policy framework for photon and neutron RIs*.

Deliverable D2.1 includes an overview of FAIR research data policy, a discussion of the European and national policy landscape, a review of the current status of data policy within photon and neutron research infrastructures, and a synopsis of policy recommendations emerging from the EOSC. Further, D2.1 discusses the relationship between the work of ExPaNDS and parallel work within the PaNOSC project. These sections remain valid, and the final version data policy deliverable we present here should be read in conjunction with this earlier draft deliverable.

Deliverable D2.1's key output was a set of thirty elements that provide a framework to inform the development of data policy for photon and neutron research infrastructures. In particular, these elements focus on enabling FAIR data. In essence, they are those elements about which research infrastructures need to make choices — both on the level of commitments that they themselves are prepared to make as well as with regards to the obligations that they will place on users.

These 30 elements were provisional, based on the analysis presented in the earlier sections of the D2.1 deliverable. Our intention was to use these draft principles and recommendations as a basis for further discussion with ExPaNDS partners to review and revise the elements into a final version of the data policy framework.

Discussion with ExPaNDS partners took the form of a formal consultation, undertaken over four months (February – May 2021) with the aims of:

1. Obtaining feedback from ExPaNDS partners on the draft data policy framework (i.e. with the intention of producing a final version framework)
2. Reviewing the current data policy situation at partner facilities and learning what future work is being planned
3. Engaging on a deeper level with ExPaNDS partners on matters related to data policy and policy making.

This report provides further details on the consultation approach, including about the consultation process itself and the qualitative analysis of the consultation feedback, and presents the detailed themes that emerged from the consultation discussions. This feedback is then synthesised to draw out ExPaNDS partners' summary perceptions of the framework.

Elements of a Data Policy Framework

The consultation offered a valuable opportunity to engage deeply in discussion with ExPaNDS partners about data policy and policy making. The ExPaNDS data policy framework has been revised to reflect well the themes, priorities, and needs that emerged most often and most strongly in those conversations. This has resulted in a reduced set of 21 elements in the revised

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

framework, accompanied by a detailed commentary on their scope and application. These 21 elements are as follows:

1. RIs should openly publish a data policy, including the period in which the policy is in force and when it is planned to be reviewed. A PID should be used to refer to the published version of the policy.

2. The RI's data policy should specify the commitment of the RI to ensure that experimental data is made available and reusable, including enabling access to the data beyond the experimental team as appropriate.

3. A RI's data policy should comply with their national research funder's policy. The data policy should be sufficiently flexible to accommodate the data policies of international funders, such as the EOSC, and those of users' funders.

4. RIs should seek to align their data policies, within the constraints of divergent national funder policies and legal frameworks.

5. RIs should specify the rights and responsibilities of particular classes of actors involved in the experimental process.

6. RIs should seek to clearly define the terminology used within the data policy and use this terminology consistently within the policy and, if possible, elsewhere in the RI's policies and practices.

7. The RI's data policy should seek to cover all classes of experimental data which are generated, stored, and analysed using the facility's resources (e.g. instruments, compute infrastructure, software, and staff) in the course of a user experiment.

8. The RI's data policy should cover the metadata used to document experimental data identified in the policy's definitions.

9. The data policy should specify the retention policy for each class of experimental data, with a minimum retention period and criteria for deletion. As this includes auxiliary data, this also includes software and tools.

10. In the event that data are deleted, the facility should retain a record that the data existed. This could constitute a (metadata) record of their essential characteristics or a method to allow the reconstruction of the data. The facility should support as much as possible the provenance and validation of published research results in such circumstances.

11. The RI's data policy should include commitments to enable the experimental data in scope to be FAIR. This may include the following commitments:

- The RI should provide the globally unique identification of experimental data via the association of an appropriate globally unique PID that conforms to the EOSC PID Policy.
- The RI should annotate data with metadata in conformance to publicly available community and domain standards.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

- ❖ The RI should support standard protocols for accessing data.
- ❖ The RI should provide data in formats conformant to publicly available standards.
- ❖ The RI should provide sufficient contextual metadata and auxiliary data.
- ❖ The RI should provide access to experimental data and associated metadata via human and machine-readable interfaces.

12. The RI's data policy should specify a licence under which the data are made available.

13. RIs should specify a decision making process for determining the grounds for restricting access to particular experimental data.

14. RIs should specify the time limit (an 'embargo period') for which users are allowed exclusive access to experimental data. This should also specify who can access the data (e.g. facilities staff), who can determine who should be given access rights, including who can lift the embargo to allow early publication, and the appeals process established to alter the embargo period.

15. The RIs must comply with the relevant national legislation, notably that under the GDPR framework, in the handling of personal and sensitive data.

16. The RI's data policy may consider the extent to which it commits to providing infrastructure to support the retention and distribution of FAIR data, for example:

- ❖ a storage and curation service to keep experimental data for the specified retention periods;
- ❖ a data discovery service to keep experimental data or its record findable;
- ❖ a data and metadata access and movement service to allow users to interrogate the experimental context and access experimental data.

17. The policy should specify the requirements on users to participate in the facilities data management planning activities, including whether the experimental team is responsible for preparing a DMP.

18. The RI's data policy should promote the recognition and citation of the use of facilities.

19. Users may be requested to report on compliance for previous experiments when applications for further access to the facility are received. RIs might consider that non-compliance may be a contributing factor in the refusal of further access.

20. RIs should have regular audits of their data management implementation and practices to evaluate compliance to the data policy and, in particular, the FAIR data principles.

21. Changes or termination to the data policy will be given in sufficient time for PIs to take alternative action to provide alternative provision to comply with their funders' data policies.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Impact of the Data Policy Framework

The consultation process allowed a deep engagement with RIs, with a thoughtful dialogue and exchange of ideas. It is clear from the consultation feedback that the ExPaNDS data policy framework is already having an impact on facilities' data policy planning. ExPaNDS partners plan to undertake future work in relation to a range of elements set out in the draft data policy framework. While not all ExPaNDS partners have set out the same future work plans, a degree of overlap in intentions is evident, and high priority areas relate to FAIR data, including persistent identifiers for data, metadata, data analysis, and software to support FAIR. Additionally, FAIR certification and data management planning are on a future roadmap; further work within ExPaNDS will provide guidance to support these aspects.

It is also notable that ExPaNDS partners saw the consultation process itself as influential in spurring facilities on to reconsider their data policies, shifting their thinking. Nevertheless, diversity on data policy will remain a feature of the photon and neutron research landscape as each RI reflects it priorities within its own national research landscape.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Table of Contents

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 857641.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

1. Introduction

1.1 Background to the deliverable

1.1.1 Draft version of the ExPaNDS data policy framework.

In our previous data policy deliverable, D2.1: Draft extended data policy framework for photon and neutron RIs (September 2020),[1] ExPaNDS considered how data policies for national photon and neutron research infrastructures (PaN RIs) should be framed, especially in terms of supporting open science and FAIR (Findable, Accessible, Interoperable, Reusable) data. As well as overviewing the national and European policy landscape and the current state of data policy at ExPaNDS partner facilities, the deliverable reviewed a set of FAIRsFAIR policy recommendations[2] and produced a parallel set of related recommendations for PaN RIs.

The report's key output was a draft framework of thirty elements that should inform data policy for PaN RIs. These thirty elements represented those elements about which we proposed that research infrastructures would need to make choices — for example, around level of commitments that they themselves are prepared to make as well as with regards to the obligations that they will place on users. Our intention was to use these draft principles and recommendations as a basis for further discussion with ExPaNDS partners to develop a final version of the data policy framework.

1.1.2 Consultation with ExPaNDS partners

Our discussion with ExPaNDS partners took the form of a formal consultation, undertaken over four months (February – May 2021). Section 2.1 provides further detail on the consultation approach, including the aims and purpose of the consultation, the consultation process itself, and the analysis of the consultation feedback. Section 2.2 presents the themes that emerged from the consultation discussions. Chapter 3 then synthesises these to draw out ExPaNDS partners' summary perceptions of the framework.

Overall, we found the consultation a very useful opportunity to engage deeply in discussion with ExPaNDS partners about data policy and policy making. It is our hope that the revised version of the ExPaNDS data policy framework (see Chapter 4) reflects well the themes, priorities, and needs that emerged most often and most strongly in those conversations.

1.2 Overview and content

This deliverable report comprises five main parts:

[1] Matthews, B., McBirnie, A., Vukolov, A. et al. (2020). D2.1: Draft extended data policy framework for Photon and Neutron RIs. https://doi.org/10.5281/zenodo.4014811
[2] Davidson, J., Grootveld, M., Whyte, A. et al. (2020). D3.3 policy enhancement recommendations. https://doi.org/10.5281/zenodo.3686901

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Chapter 1 provides background to the development of the ExPaNDS draft data policy framework and explains how we have engaged with ExPaNDS partners around the development of the final version of the framework.

Chapter 2 describes the consultation process (section 2.1) and presents the key themes and suggestions for revisions that emerged from the consultation feedback (section 2.2).

Chapter 3 synthesises the consultation findings presented in Chapter 2, asking a series of questions to draw out ExPaNDS partners' summary perceptions of the data policy framework.

Chapter 4 presents the revised, final version of the ExPaNDS data policy framework.

Chapter 5 outlines ExPaNDS partners' plans for future work around data policy, offers reflections on the consultation, and considers next steps, especially in relation to upcoming work in the ExPaNDS project that will support the data policy framework.

2. Consultation on the ExPaNDS Draft Data Policy Framework

2.1 Consultation Approach

The sections below describe the aims of the consultation and how it was undertaken.

2.1.1 Aims and purpose

Purpose of the consultation

The purpose of the ExPaNDS WP2 data policy framework consultation was three-fold:

1. To obtain feedback from ExPaNDS partners on the draft data policy framework (i.e. with the intention of producing a final version framework)
2. To review the current data policy situation at partner facilities and learn what future work is being planned
3. To engage on a deeper level with ExPaNDS partners on matters related to data policy and policy making.

Specific aims of the consultation

More specifically, the consultation aimed:

- To capture any changes in the current data policy situation across ExPaNDS partners (i.e. since the December 2019 Landscape survey and analysis of data policy and practice for

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

ExPaNDS partners and the overview summary of that survey presented in section 3.2 of D2.1: Draft extended data policy framework for Photon and Neutron RIs)[3,4]

- To identify current data policy elements that support/hinder FAIR data practice
- To identify similarities/differences (and reasons for these) between ExPaNDS partners' data policies
- To compare ExPaNDS partners' current data policies with the proposed 'Thirty elements that should inform a data policy' (i.e. based on Chapter 6 of ExPaNDS D2.1) (See Appendix 1)
- To capture information on 'outside FAIR' matters (e.g. data retention/deletion, GDPR) that could have in/direct effects on FAIR data policy
- To understand what data policy aspects seem to have the most influence on scientists'/users' practice
- To explore which aspects of data policy are most likely to positively/negatively influence scientist/user behaviour.

2.1.2 Methods and framing questions

Consultation meetings and write ups

The consultation took the form of separate online meetings with the 10 ExPaNDS PaN RIs.[5] Write ups of these conversations were then passed back for accuracy checking to the colleagues who attended the meetings. Some partners also submitted additional written feedback.

Use of framing questions

A semi-structured interview approach, guided by a set of seven framing questions (see Appendix 2), ensured the meeting conversations addressed a set of common themes but also allowed discussion to move into other/additional areas of interest or concern to individual partners.

Qualitative analysis

Analysis of the consultation feedback was undertaken using qualitative coding approaches, specifically, simultaneous coding and descriptive (i.e. topic) coding.[6] A range of specific and wider themes emerged from the analysis. Section 2.2 presents those themes most relevant to the revision of the data policy framework and to understanding data policy at ExPaNDS partner RIs.

[3] Matthews, B., McBirnie, A., Vukolov, A. et al. (2020). D2.1: Draft extended data policy framework for Photon and Neutron RIs. https://doi.org/10.5281/zenodo.4014811
[4] Ashton, A., Da Graca Ramos, S., Matthews, B. et al. (2019). ExPaNDS data landscaping survey. https://doi.org/10.5281/zenodo.3673811
[5] The 10 ExPaNDS PaN RIs are: Deutsches Elektronen-Synchrotron (DESY), Paul Scherrer Institute (PSI), Helmholtz-Zentrum Dresden-Rossendorf (HZDR), Diamond Light Source (DLS), MAX IV, Elettra, ALBA, SOLEIL, Helmholtz-Zentrum Berlin (HZB), and ISIS Neutron and Muon Source (ISIS). Additional ExPaNDS partners include: UK Research and Innovation (UKRI), Science and Technology Facilities Council (STFC), and European Grid Infrastructure Foundation (EGI). ExPaNDS (2020). Partners. https://expands.eu/partners/
[6] For more detail on these qualitative coding approaches, see: Saldaña, J. (2009). *The Coding Manual for Qualitative Researchers*. Sage; pages 62–65 and 70–73.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

2.1.3 The consultation meetings and data collected

Quantitative information on the consultation feedback

The consultation feedback comprised around 100 pages of text. The responses ranged from just over 1000 words to almost 8000 words in length. In some cases, as well as responding to the framing questions, the feedback addressed each of the thirty elements of the draft data policy framework in turn, resulting in highly-relevant comments and specific suggestions for revisions.

All 10 ExPaNDS PaN RIs participated in the consultation. A subset of ExPaNDS WP2 colleagues organised and facilitated the consultation meetings. These same colleagues also produced the consultation write ups.

How the consultation meetings were undertaken

Where possible, consultation meetings were held with relevant colleagues from the facility all in attendance at the same time. This served to ensure that the conversation was as natural as possible for those participating and that multiple views could be collected. In a few cases, data was collected over the course of several meetings. Most meetings were at least an hour long, and many were longer.

Who participated in the consultation

With the aim of collecting feedback from multiple perspectives, a variety of ExPaNDS partner colleagues were approached and asked to participate in the consultation, including User Office staff, Senior Management, IT and research data managers, data librarians, instrument scientists, legal specialists, and other roles likely to participate in data policy making. Although the consultation did not involve facility users directly, it is important to emphasise that user needs are at the forefront of ExPaNDS partners' thinking. As such, users will have had an indirect impact on the feedback, particularly through the instrument scientists who participated.

2.1.4 Limitations of the data and analysis

Limitations

It is important to recognise that, despite the strong engagement of ExPaNDS partners with the consultation, there are limitations inherent in the approach adopted and the data collected.

Accuracy checking by participants

While the write ups strove to capture the essence of the conversations — verbatim, where possible — the interview discussions were not necessarily transcribed word for word. Hence, the reason that the checking of the write ups for accuracy by the participants was crucial.

Redundancy in the data analysis

Qualitative data analysis, no matter how systematic, will always be subjective to some degree. Being so close to the data as a qualitative researcher can be helpful (i.e. in achieving a deep understanding of participants' perspectives) but also requires care (i.e. around ensuring that bias is limited as much as possible).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Reaching the stage that high levels of redundancy appear in the data is one way of checking that qualitative analysis has been thorough and systematic. The analysis of the consultation feedback adopted this approach by allowing codes to emerge as the analysis progressed (i.e. rather than defining a set of codes in advance). Eventually, the number of codes stabilised and new codes no longer emerged.

Peer review

Additionally, in the case of our consultation, ExPaNDS WP2 was fortunate in being able to turn to peer review (internal and external) to flag up any concerns with the analysis and the resulting conclusions.

Use of quotes

Throughout section 2.2, which presents the themes that emerged from the analysis, we include quotes taken directly from the consultation write ups. Whenever such an approach is used, it is important not to reference the quote out of context, i.e. as this can change the essence of what the participant meant.

We have done our best to ensure the comments included remain true to the original context in which they appear in the consultation write up texts. However, we recognise that it is always possible, for example, through misunderstanding or lack of clarity, for quotes inadvertently to have been used out of context.

2.2 Consultation Feedback

2.2.1 Introductory comments and context

Focus on themes

The sections that follow present the themes that emerged from the analysis in relation to the framing questions as well as additional themes that emerged from the consultation discussions.

Anonymisation of quotes

To relay ExPaNDS partners' perspectives, we include a range of comments quoted verbatim from the consultation write ups. Names of facilities are removed in line with standard practice in qualitative analysis; however, due to the nature of ExPaNDS partner facilities and the nature of the data, it is not always possible to truly anonymise the comments. In this regard, it is important to emphasise that all ExPaNDS partners have agreed that their consultation write ups can be made publicly available. We include the full write ups of the consultations in a Supplementary Document.[7]

[7] McBirnie, A., Matthews, B., Gagey, B. et al. (2021). Supplementary document to ExPaNDS D2.3: Write ups of the ExPaNDS data policy consultation meetings. 10.5281/zenodo.5211847.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Diversity of ExPaNDS partners' data policies

In the details of any analysis, it is possible to see many trees but miss the forest. What is sometimes but not always obvious as we examine one particular theme or element in the discussion below is the diversity that exists across ExPaNDS partners when it comes to data policy.

Some ExPaNDS partner data policies only address raw data, while others cover a much broader range of data, including results data. Not all policies incorporate the concepts of open data or an embargo period. Some facility data policies apply only to internal staff; others focus on external users. While research data curation and data lifecycle aims clearly underlie some policies, for others, the focus, both historical and current, is more on data storage and access.

There are many reasons this diversity features across ExPaNDS partner facilities. We have taken the view in ExPaNDS that one size very much does not fit all, and our challenge is a data policy framework that can encompass all of this diversity. For us, this diversity is our 'forest', and we must keep our bearings as we navigate through it. As you the reader move through the trees below, we ask that you also aim to keep that wider forest in mind.

2.2.2 Framing Q1: Broad, thematic areas

Policy Scope

The nine themes of the draft policy framework

Responses were broadly favourable concerning the nine themes covered by the ExPaNDS draft data policy framework.

> *"All of the nine themes do seem relevant for a data policy."*
>
> *"The thematic areas are well chosen, and broad enough to cover all important topics."*
>
> *"[We] share the general drivers and principles expressed."*

Not all themes need to be covered by the data policy itself

While ExPaNDS partners generally consider the themes relevant — with several already included in current facility data policies — one partner emphasises that although all nine may have relevance to discussions around data policy, this does not equate to a need to cover all of the themes in the data policy itself, especially where other avenues, such as training, may offer alternative, and perhaps more effective, ways to address a theme.

> *"Most of these items are certainly useful for the general discussion. But many of them are rather not something to put into the data policy."*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

While other partners comment less specifically on this concern, they nonetheless echo the sentiment that the data policy framework should offer recommendations, but with an understanding that these may or may not be taken up formally in a facility data policy.

> *"The policy framework here can only provide recommendations"*

Proprietary data and implementation details are out of scope

In terms of other matters related to scope, ExPaNDS partners agree strongly on two points:

1. Proprietary data[8] would not normally be covered by the data policy framework
2. Implementation details have no place in the data policy framework.

Focus on commitments rather than actions

Regarding the second point, there is a clear concern that including implementation details in the framework (or, indeed, the data policy itself) could lead to the guidance quickly becoming obsolete. The feedback is that the data policy framework should focus on commitments and that the inclusion of actions should be minimised.

> *"But [to] say "specify the actions" or "specify the commitments" of the RI? Specify the actions can lead [one] to go into details which have no place in a data policy. Too much detail can make the data policy quickly obsolete."*

> *"The rule of thumb is to consider what we need do in order to provide FAIR data. Then, we need to ask whether we are allowed to do that without having it mentioned in the data policy. If the answer to this question is yes, then we don't need a regulation in the data policy."*

Harmonisation

Details reduce flexibility

The question of what is included in the data policy framework also links to the issue of harmonisation. While ExPaNDS partners recognise the importance of data policy alignment across PaN RIs, they balance this with the need to retain flexibility for their individual facilities, not just around implementation matters, but also in relation to policy making itself. There is the general sense that the more detailed the data policy framework, the more challenging harmonisation becomes.

> *"Aligning policies should not mean that all the RIs must have the same policy. We agree that we need to check that policies are aligned in the sense that they don't create any obstacles for researchers who want to visit different institutes or who want to combine datasets from different facilities."*

[8] 'Proprietary data' is sometimes also referred to as 'commercial data'. The term refers to data that is the result of experiments undertaken through the commercial or paid- access routes offered by ExPaNDS partner facilities.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Check for obstacles to alignment

This idea of checking for obstacles to alignment, i.e. rather than making alignment a formal element of the data policy framework itself stands out. PaN projects have the ongoing, inherent effect of focusing PaN RIs on the issue of harmonisation. Does/should this change how and to what extent harmonisation needs to be addressed formally in PaN data policy making?

> *"[Harmonisation] reinforces the usefulness of common projects like PaNdata, PaNOSC or ExPaNDS which allow [PaN] to have common bases. As these bases can then be declined differently at each RI, [could we] include alignment verification as part of each internal process to review a RI data policy? But, should the need for this alignment be written in the RI data policy?"*

Other policy alignments may take priority

More broadly, facilities have additional internal and external policy alignment needs to consider. When it comes to prioritisation, ExPaNDS partners note that harmonisation across PaN may not necessarily trump local and national policy alignment needs.

2.2.3 Framing Q2: Specific policy elements

Classes of data

State what data are in scope

The draft ExPaNDS data policy framework defines several classes of data that a data policy should cover. While there are extensive differences in how ExPaNDS partners define classes of data, there is a clear consensus that a data policy must state what classes of data are in scope.

Experimental data

As to the classification of experimental data into separate types (raw, reduced, processed, analysed, etc.) and the definitions of these, ExPaNDS partners feel this is neither something that can nor should be harmonised across facilities. The need for internal policy alignment provides an example of a situation where the adoption of a harmonised set of external definitions may not be possible.

> *"Currently the only existing definitions [in our policy] are "Raw Data" and "Metadata", but this can change in the future, but [we make] no commitment to use each of the terms suggested by the framework… For now, these two definitions should be broad enough to contain the other terms outlined ..."*

> *"What we actually have now is a common set of definitions that applies across all of our [internal] policies, i.e. as we found we often used the same terms in different policies. So, if we rewrote the data policy now, we could simply link out to that common set of definitions. We felt it was very important that the same definitions are used across all our policies, whether those policies focus on user access or data or safety."*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Definitions

Several partners also consider the definitions themselves confusing and unhelpful. There is even the overarching question of whether the ExPaNDS data policy framework should refer to the broader term 'scientific data', i.e. rather than 'experimental data'.

> *"It is said that "reduced data" is "generated from raw data produced using facilities' compute and software resources" but it is also true for the raw data itself. And what is said for processed data is also true for raw data. So what is really the difference between the three sentences? In the end, the distinction between raw, reduced, and processed data is a philosophical question that may be discussed in the context of a particular experimental workflow, but that cannot be solved at the abstract level of a data policy."*

> *"The definition of the classes of experiment data are naturally very disputable. The definition of "raw data" varies already a bit. Some colleagues wouldn't accept lossy compressed data as raw data, some would. The definitions for data the policy applies to appear partially too narrow (FE 12), and a bit too broad (FE 11). What all have in common that the policy will cover scientific data, and we prefer the notion of a scientific data policy framework."*

Distinctions relate to implementation matters

Indeed, distinctions between different types of classes of experimental data are seen by many as really only important in relation to implementation matters, particularly the right (or not) of the facility to do something.

> *"In the policy, we only need to make the distinction between classes of data if we want to apply different regulations respectively. But … at [our facility], we will treat raw, reduced, and processed data all alike, as far as regulations are concerned. We need only to distinguish raw data and results. Raw data becomes openly accessible after the embargo, results do not."*

> *"We expect to apply different policy measures, such as storage time, depending on the class."*

> *"There is this question of where this line is between raw and processed data, and then, what's generated by the facility and what's generated by the researchers within a project, albeit it may be generated on [our facility] resources. We don't know where that line is or have clarity on this at the moment to say that just because data is generated at [our facility] and processed at [our facility] on our systems means that we have the right to make all of that open."*

Metadata

Facilities recognise that metadata plays a significant role in producing FAIR data. They feel that there is much work to be done in this area, and that both the user and the facility have a role to play. In particular, there is the need to make it clear to users what their responsibilities are in relation to providing metadata.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

> *"There is no standardised way to document the dataset … and to provide necessary metadata to guarantee the accessible data are reusable, too."*

> *"Making data FAIR is not only depend[ent] on the facility, but also on the users (completing metadata, establishing a DMP, etc.)."*

> *"There is, anyway, a general agreement on the need to state in the policy what the obligations of the users are in terms of completeness of metadata…"*

Auxiliary data

ExPaNDS partners express considerable interest in auxiliary data, a class of data first introduced formally in the PaNOSC data policy framework.[9] Most see this class of data as important for FAIR and feel that it should be included somehow in the data policy. However, many questions remain around legal and implementation matters.

> *"Auxiliary data is a wide category and has some different considerations to take into account, for example, around rights in relation to third party data and such. However, this kind of data (e.g. samples in the life sciences) can be very useful in providing the contextual information needed to understand the experiment. … but it is also one of the hardest aspects to get to grips with. This kind of data is really important for FAIR but it is also really messy – for a start, who owns it, can we use this (i.e. given that it is not facility generated data)? It is very, very difficult to see how this will work but it is also very important. Another issue is that if you have a policy that covers only facility generated data, then how does auxiliary data sit within that?"*

Regarding terminology, for some, the term 'auxiliary data' is unclear, with "third-party data" or "user-provided data" suggested as possible alternatives.

Embargo

Embargo period

Most ExPaNDS partners do apply an embargo to experimental data. The most common embargo period is 3 years. A few facilities have an embargo period of 5 years. One facility does not have open data and, therefore, does not have the concept of an embargo.

Embargo extension

All facilities that have an embargo period also offer the possibility to extend the embargo period. Requests for extension often need to be made at director level (or equivalent); however, it seems that such requests are rarely, if ever, made in practice.

Outstanding questions

ExPaNDS partners do have outstanding questions and concerns related to the concept of embargo. These relate to matters such as the publishing of data linked to a scientific paper,

[9] Gotz, A., Perrin, J., Fanghor, H. et al. (2020). PaNOSC data policy framework. https://doi.org/10.5281/zenodo.3862701

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

whether data and metadata are both covered by the embargo period, and to what extent the facility itself should have the right to restrict access to data or decide that an embargo does not apply.

> *"I see an increasing need for opening the data as soon as a related paper is published…"*

> *"…there should not be situations where you even have to ask for an extension of the embargo period. There are also situations where the PI may not have a say, i.e. because the data policy allows the facility itself to take the decision..."*

Persistent Identifiers (PIDs)

Current use

ExPaNDS partners express the most interest in PIDs for publications, data, and people. A few do already use PIDs for all of these, although no facilities yet mandate the use of PIDs for people, such as ORCID.[10] A couple of facilities mention the possible future use of ROR[11] to identify research organisations.

In general, the use of PIDs for instruments does not feature in the consultation feedback, even in relation to discussions around the citation of instruments. There is also very little discussion about the value of linking PIDs.

Level of granularity

Some facilities, however, are clearly grappling with the questions of what data and experimental components should have PIDs and at what level of granularity.

> *"PIDs should be assigned to things of value that you want to keep around 'forever'. In the context of an experiment, it can be difficult to decide what you want to keep in this way. How much of the whole experiment do you need to keep the context? At what level of granularity should you assign PIDs? … [Our facility] is currently discussing many of these sorts of questions."*

PID for the data policy

The draft data policy framework's suggestion for the assignment of a PID to the data policy itself seems well-received by ExPaNDS partners. No facility rejects the idea. Indeed, several facilities say they intend to adopt PIDs for their data policy (versions) in future, although it is not necessarily clear how this will be implemented.

> *"[Our facility] data policy is available at our external website, but without a PID. Indeed, it would be useful to have [it] properly published with a PID for every version. We agree that we should do that in the future."*

[10] See: Open Researcher and Contributor ID. https://info.orcid.org/what-is-orcid/
[11] See: Research Organization Registry. https://ror.org/about/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

"No PID associated with the data policy. There is a general interest in achieving that at some point, and probably one per data policy version. It might require some work, and how to achieve this is not yet defined."

2.2.4 Framing Q3: Impact of FAIR

FAIR in the data policy

High level commitment

ExPaNDS partners support incorporating the FAIR Principles[12] into their data policies. Some policies already refer to FAIR, and there is a clear intention to include FAIR in more facility data policies in future. Currently, where FAIR data is mentioned in facility data policies, the reference appears in the introduction or policy preamble.

It is notable, however, that in both current mentions and plans for future mentions, there is a strong desire to avoid going into actions or implementation details and to focus instead on expressing a commitment to FAIR and/or the aim or goal to achieve FAIR. This echoes ExPaNDS partners' widely-held sentiment that implementation details are outside of the scope of the data policy (see section 2.2.2).

"In introducing its data management policy, [our facility] explicitly encourages its Users and Scientists to produce FAIR (Findable, Accessible, Interoperable, Reusable) data according to the principles of Open Science and explains aiming to facilitate this with its data management services."

"We do [mention FAIR in the data policy], but only in the preamble - there is no specification of how the goal is to be achieved."

"Will add a commitment to the FAIR data principles to the introduction of the policy… Again, how to do that in detail in the practice, this is not something that should be regulated in the policy. Producing FAIR data is a matter of doing actions and making tools to put this in place in practice… But implementation details are out of scope of the data policy."

Separate principles and implementation

Indeed, ExPaNDS partners do seem to separate the ideas of principles and implementation when it comes to FAIR. Achieving FAIR data is seen as a moving target, for which a 'best effort' approach is needed with the understanding that implementation will develop and improve over time. The FAIR Principles themselves are seen as more fixed, and in this sense, more appropriate to reference in a data policy, which itself may change only infrequently.

"Principles are one thing; broad and scalable implementation is another."

"Creating FAIR data is a dynamic process requiring innovation and constant adoption to best practices, while a data policy is a static document written once

[12] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data*, **3**:1. https://doi.org/10.1038/sdata.2016.18

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

which takes a huge effort to pass through committees and all the entities in a facility. If you describe the process of creating FAIR data too much in detail, you end up with something rigid which may become an obstacle to implement the best practices and the state-of-art techniques."

Realistic approach to FAIR

Data should be 'FAIR when it leaves the facility'

ExPaNDS partners strongly agree with the concept of 'FAIR when it leaves the facility' as introduced in the ExPaNDS draft data policy framework. Facilities are very aware that they have little to no real control over what users do with data once it leaves the facility.

"Importance of the statement: 'Once the experiment is complete, then subsequent research actions are out of the facilities' control. However, facilities can request that users continue to keep derived results data available and reusable.'…"

"Data sharing within the scope of "at the point of leaving the facility" is perceived as the most realistic approach to implement FAIRness."

Data Management Plans (DMPs)

A topic of active interest

Many ExPaNDS partners are actively considering the role of DMPs to support FAIR.

"We are already having a discussion around this area."

"DMPs are very much something [our facility] is thinking about … because of the nature of the experiments that are going to be coming down the line."

DMPs in the data policy

While facilities see potential value in DMPs, they do not always agree that the data policy itself should address DMPs. Some feel that, while user support for DMPs should be provided, DMPs should not be an obligation placed on users by the data policy.

"Not sure if it [DMPs] should be part of the policy. Think it should be regulated somewhere else."

"…users and also our own scientific staff often need to submit DMPs to funders. We plan to provide assistance on writing DMPs. It should be more a service to provide them more than a requirement to put on them…"

It is also notable that where DMPs are currently mentioned in a facility data policy, this seems to be having little impact in practice.

"We do mention DMPs in our data policy, but this mention is more vague than very specific. It's more of a recommendation rather than an obligation, and this is why most researchers don't do them."

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

In the event that facility data policies were to cover DMPs as standard, questions of how and to what extent remain open.

> *"Data management planning is one area to consider. Thought this was taken as read that facilities are going to have this in future — but, talking to people at [facility X], they say they may not ask users for individual DMPs, and talking to people at [facility Y], they say they may only ask users for DMPs if they expect to generate very large volumes of data. This is quite surprising because data management planning is potentially something that really underpins the whole data policy."*

Value of DMPs for the facility

The question of what value DMPs bring to the facility is also an important one for ExPaNDS partners. As the comment above indicates, there is the suggestion that DMPs may be able to help with understanding of experimental data volume; however, feeling is not unanimous on this point.

> *"It is often argued, a DMP would be needed so that the users should tell how much resources they would need. But in fact, we know all that already, we probably know it even better than the user. Because for each of our instruments and for each experimental technique, we know very well how much data they generate in each measurement and how many measurements it can perform during a given amount of time. In most cases we also know what kind of analysis can be performed on the data created by our instruments."*

> *"Providing estimates on the storage … needed to process the data are not required in [our] data policy. It seems rather to be done in a collaborative way with the RI staff (scientists who know the possibilities of the detectors and instruments at the beamline, IT staff, etc.)."*

Being able to articulate the value of DMPs for the facility is perceived as important for convincing senior management to place what they might otherwise see as simply an additional burden on users. There is also the question of how to enforce the use of DMPs, especially if they become an obligation set out in the facility data policy.

> *"Senior Management is not keen on this. Users already come to [our facility] with a DMP from their institution. But even if [our facility] did start using DMPs, how would we enforce this? This is the argument of senior management. But, if this becomes something that all facilities are using, then it will likely become more palatable to [our facility]."*

DMPs to support research data curation

More broadly, ExPaNDS partners want to understand how they could use DMPs to best support FAIR research data management at their facilities. Where collections of DMPs do exist at a facility, the view is that they are currently not used to best advantage.

> *"One of the questions that is key is how can a facility use a DMP? What are the advantages for the facility?"*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

"…the DMP is not stored in a repository where we can all see and share and talk about it and any problems with it. Instead, it is stored only on an internal network. So, it is not possible to see all of the DMPs together, and that's a problem."

"There is no central point for storing these DMPs and for connecting them to the research data curation process."

Value of FAIR data

Publications still have the 'value'

Feedback from ExPaNDS partners suggests that publications still tend to dominate thinking around value.

"Publications are still seen as having the 'value', not data. Indeed, most researchers feel that the data is so specialised and specific, that it cannot easily be reused by others. Most researchers see their data as only counting for their own projects."

"Currently, there is limited incentive for researchers to publish their data and get credit for it. There are no indices or measures, meaning that data is not counted statistically in terms of giving researchers credit."

Other effects of the focus on publications

The focus on publications appears to have other, more subtle effects as well. Comments from ExPaNDS partners highlight two intertwined issues:

1. Historically at facilities, there has been a tendency to take the thinking and policy approach that works for publications and simply transfer this to data; however, ExPaNDS partners recognise that this has created some confusion, for example, in areas such as data publication and data for publications. Facilities would like to address these concerns, but are unsure how best to do so.

   "We have a clear process for publication of scientific results. Pretty much the same process should also apply to the publication of scientific data. The definition of FAIR data is also very clear. There is however a bit of confusion when it comes to "published scientific data". Are FAIR data always published? Do published data have to be FAIR? Are data with a PID already published, even if just the embargo expired (some scientists would refuse this idea)?"

   "We also do curated data publications in a separate workflow as a distinct service next to the regular data management workflow. It is not mentioned in the data policy and we need to check whether this requires some regulation. For instance, we need to give reviewers of a publication access to the data. That naturally happens before the publication. It is not very clear whether the policy allows us to do that."

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

2. With publications, it is usually clear what facilities store; for example, while a facility might store preprints and/or final versions of published articles, it is unlikely to store early rough drafts of publications. The decision on what to store may even be guided by a library collection management policy or a controlled vocabulary list of publication types in a facility database.

   With data, however, the situation is much less clear. In general, facilities appear to want to be guided by what is needed for FAIR data. Applying 'publications-thinking' to this situation (i.e. given that FAIR relates to data, not publications) may be less than helpful but can also be difficult to avoid given the lack of clarity (see point 1 above) on what 'counts' as data publication.

   "There is another interesting question here. Which data should be stored? Should all data be stored? In certain experiments, you create a lot of experiments that don't work (e.g. the samples aren't good enough, there are anti-pictures, etc.). In such cases, what do we really need to store?"

   "The issue with the processing also is that a lot of times users are exploring parameter space, and, very often, those parameters are incorrect. So, then, if you expose that out, would somebody confuse that with a properly processed result?"

Awareness of FAIR

Aside from noting these issues around data versus publications, the consultation feedback also emphasises that much better awareness of FAIR is necessary if the PaN community is to understand the value of FAIR data.

> *"A very key issue is the point that FAIR is not well known amongst the researcher/user community. For example, FAIR is normally not raised by or within the community. In this sense, researchers may not engage with the policy or, crucially, the change it hopes to bring about in terms of an increased focus on FAIR. There is still much to do to raise the profile of FAIR. The case needs to be made why FAIR is useful and why a researcher would want to engage with FAIR."*

2.2.5 Framing Q4: Consequences of legal and other external frameworks

Links to the wider policy and legislative ecosystem

National legislation and policy is a priority

ExPaNDS partners are aware of the need for the data policy to sit within a wider policy and legislative ecosystem. They consider national legislation and policy particularly important.

> *"There are quite a few national laws that we need to apply in relation to archiving data."*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

> *"We are a national facility, so if there was an overarching national policy, that would be the one we would [want] to adhere to."*

> *"… there is a commitment to accommodate funders' policy and it is also important for the users to have a reference to the underlying funders'/national policies."*

In acknowledgement of the wider ecosystem, facilities take a range of proactive steps:

- to examine the wider policy environment when developing policy;

  > *"[Our] current [facility] data policy does make provision for taking into account third party obligations. Anytime we develop a policy, we always make sure at the start that there will be nothing inconsistent with what might be in other relevant policies…"*

- to consult directly with funders on policy development;

  > *"Consultation with [national funder X] ensured that [our facility] is aligned with [our country's] largest research organisation."*

- to ensure users are aware of the data management policy requirements of the wider ecosystem.

  > *"In [our] data policy, [our facility] commits to bring help by sending out the available information to the main proposer if the funder or research institution or other organisation has particular data management requirements ..."*

Legislation must predominate

Several partners also stress the point that legislation and laws impose themselves on facilities – not the other way around. To this end, some felt the wording (e.g. FE18) in the draft data policy framework should be revised.

> *"It has to be considered that the legislation is predominant and imposes itself on us…"*

> *"We do not need to acknowledge a legal framework in the policy, but we need to act in conformance with it."*

> *"[Re FE18]: Laws don't need to be 'acknowledged', they need to be obeyed."*

Impact of changes to legislation

Other points ExPaNDS partners emphasise are that legislation can and does change and that facility data policies need to keep up with these changes. Indeed, changes to legal and external policy frameworks are one reason that facilities review their data policies. Despite their awareness of these points, however, facilities note that ensuring legal currency and compliance is an ongoing challenge that is not always straight forward.

> *"Legal frameworks are changing and sometimes difficult to comply with."*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

"…legal frameworks can evolve and will always be predominant."

"…old guidelines and frameworks are now being observed [in our current data policy]. This must … have an impact on a new version of the data policy."

Flexibility is an advantage

More broadly, ExPaNDS partners recognise the flexibility that a data policy framework can offer in terms of working in a dynamic and complex ecosystem. As more policies come into play, it is important to avoid conflicts between them.

"There are also new situations arising that can affect policy – for example, [our instrument X]. This is an instrument based [at facility Y], but where we are going to collect data as [our facility]. So if [our facility] has its own policy and [facility Y] has its own policy, those two policies need to be compatible. This could increasingly become an issue as experiments happen remotely or become more distributed and less (single) place-based. There will be more policies to take into account."

"A policy framework is not a model policy. Rather, it is meant to work in a complex policy environment, where you might need to bring in other internal and external policies. … we in PaN couldn't have data policies that conflict with those of our users. What would our users do in that case?"

Review of the data policy

ExPaNDS partners' awareness of the wider policy and legislative ecosystem is reflected in their views on the timings for reviews of the data policy. Most partners do not feel a scheduled regular review (e.g. annually) is necessary. Instead, many facilities prefer to review their data policies on an 'as needed' basis as and when the wider ecosystem changes.

"No regular date for review; rather, the policy is reviewed when that needs to happen."

"It is a good idea to review policies at some point but I do not understand why this should be scheduled so far in advance instead of having it as a natural reaction on changes in the 'world'."

This said, ExPaNDS partners recognise that it could be helpful to put some sort of internal processes in place so that these external changes do not go unnoticed by facilities.

"The need may depend on the evolution of the context. Rather than planning a date, it could be suggested to establish an internal process for data policy reviews."

"As national funders' policies may change, why not instead suggest including such a compliance verification as part of the internal process to review the RI data policy?"

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Licensing and data citation

Licenses in use at facilities

Not all facility data policies currently address to the topic of data licensing.

> *"…we need to specify what license we will use for data. Currently, we don't specify any license under which we will publish data — it's not even mentioned."*

Of those that policies that do address licensing, all except for one indicate the use of the Creative Commons[13] CC-by[14] license for data; the one facility that does not, applies CC0[15] to raw data.[16]

There is also an awareness of other types of licenses,[17] with some already in use at facilities. Several ExPaNDS partners raise concerns around the implication of these additional types of licenses for the data policy and for implementation matters.

> *"Should the restrictions on license selection be imposed by the Data Policy? How to define the priority list of the licenses? Also, no way is defined to succeed with the licence that is incoherent/contradictory with the Policy."*

> *In [our repository], there are options for other licences now, for example, for software. These new types of licenses, however, do not show up under the recommended default list, meaning the user has to know to search for them. CC0 is good but there are also some special licenses for data itself…"*

Data and copyright

The consultation feedback includes some comments on the question of whether or not data (and, especially, raw data) is subject to copyright. Most ExPaNDS partners acknowledge either that they have not really considered this question, i.e. because they have assumed that publications and data work the same way when it comes to copyright, or that they remain uncertain about the legal copyright status of data.

One partner, however, does present clear views on the copyright status of data and the implications of this for licensing.

> *"…data is not protected by copyright. That would mean that one cannot possibly put any license on it, because licenses are legally based on copyright."*

This same partner is the only one who applies CC0 to raw data. It is important to note that CC0 waives all copyright to the fullest extent allowed by law. As such, in practice, CC0 is a waiver

[13] See: Creative Commons. What we do. https://creativecommons.org/about/
[14] Creative Commons. Attribution 4.0 International (CC BY 4.0). https://creativecommons.org/licenses/by/4.0/
[15] Creative Commons. CC0 1.0 Universal (CC0 1.0) Public Domain Dedication. https://creativecommons.org/publicdomain/zero/1.0/
[16] For explanations and examples of the different Creative Commons licenses, see: https://creativecommons.org/about/cclicenses/
[17] Examples of other types of licenses include those from the Open Data Commons (https://opendatacommons.org/)and licenses designed specifically for software (e.g. see https://opensource.org/licenses).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

designed to remove restrictions rather than a license, which inherently seeks to impose restrictions.[18]

From a FAIR perspective, the partner argues that CC0 and its aim to remove restrictions has important advantages for data reuse.

> *"On the other hand, assuming for the moment that licenses could be applied to the data in the first place, they can create serious obstacles for those who want to use data in a proper way. Because if you want to combine several datasets you would have to obey to every single condition on every dataset you might have used. You can end up in a situation where two datasets have licenses with contradicting terms and you would not be able to lawfully combine them at all. Therefore, the less terms and conditions applied to the data, the better it is for the legitimate reuse that we want to encourage after all."*

Citation in the data policy

While ExPaNDS partners agree that citation of data is important, there is no consensus that this area should be covered formally by the data policy. Indeed, the feedback is highly varied on this topic.

Some facilities would like to see detailed formats and wordings for citation included in the data policy.

> *"The User Office does have a statement that tells users how we want them to cite the beamline when they do an experiment. There is no reason that we shouldn't have a similar statement in relation to data citation, i.e. so that it is a very specific set of words that we ask for."*

> *"As it could be demanding for the user to cite all deeply connected sources, could it be beneficial to specify a procedure/schema of citations?"*

Citation as an academic norm outside of the scope of policy

Other partners consider citation an academic norm; as such, they do not see the data policy as the appropriate tool for addressing citation, preferring instead training and user support.

> *"Of course, we agree that data should be properly cited. There is no doubt on that goal. In particular our library is very interested to educate the users to do this in a proper way so that we can trace the output from our instruments. But I don't think it will help a lot to put that as a requirement in the policy because it would have no impact on what people will do in practice."*

> *"What is more important is to improve the workflows and infrastructure and to do everything to make proper citation as easy as possible and then engage in training of users to tell them this is important please do that and tell them what they will gain if they do it properly."*

[18] For full details, see: https://creativecommons.org/publicdomain/zero/1.0/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Factors beyond the control of the user or facility

ExPaNDS partners also emphasise that citation is not necessarily even something under the control of either the facility or the user. For example, publishers also have a role. This situation has implications for the inclusion of citation as a formal requirement in the data policy.

> *"There are many factors beyond control of RIs and scientists when it comes to citations."*

> *"Another useful thing would be for the PaN community to do some lobbying with publishers to promote citation of data. Because the authors often do not have the choice on what and how things are cited in the paper. It's the journal editors that dictate the citation styles, often including instructions on what is eligible to be cited."*

Using CC-by to 'force' citation

Aside from the question of whether or not citation should be included as an element of the data policy, there is the matter of facilities linking licensing to citation. The consultation feedback suggests that at least some ExPaNDS partners recommend the CC-by license for data because they are aiming to 'enforce' citation of their data. In this sense, they are not relying on the academic norm of citation but rather the legal requirement of attribution inherent in the CC-by license.

> *"The issue of CC-by versus CC0 has come up before. The concern is around ensuring attribution."*

> *[Our facility] (which uses CC-by) would not suggest CC0 as a license due to the fact that a key incentive for researchers – receiving recognition through the citation of their data – would be lost as others could use the researchers' data without any acknowledgement."*

This situation raises some interesting questions, especially in relation to the issues discussed above around copyright and data. Licenses are a legal tool, but if copyright in data has no legal basis, then can a license truly enforce citation? The consultation feedback suggests that facilities may need to think more about this issue.

Other reasons for the use of CC-by

It is important to note, however, that the use of CC-by ExPaNDS partners is not only linked to citation.

> *"…for [our facility], the use of CC-By is not linked to the question of citation."*

Indeed, some partners are not able to explain why CC-by is used at their facility. There is some indication that, historically, the practice of using CC-by for open access publications was simply transferred to apply to data; however, partners cannot say this for certain, and the specific reasons for the historical choice of CC-by have now been effectively 'lost in time'.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Other legal aspects

Status of the data policy

It is notable from the consultation feedback that some ExPaNDS partners consider the data policy a legal document (contract) while others see the data policy as a statement of best practice principles or set of guidelines. Some also feel a data policy can be both.

> *"The data policy is not a legally binding document but rather a recommendation."*

> *"[The data policy] is … a legal document as it becomes a part of the contract that we conclude as a facility with our users."*

> *"It is possible for the data policy to do both (be guidance and also a contract)…"*

Partners are in agreement that legal aspects should be clearly flagged as such, regardless of the status (i.e. legal contract or not) of the data policy. However, for facilities that see the data policy as a legal document, there is the additional need to bear in mind that everything in the data policy could be subject to legal enforcement, either now or at some future time (i.e. as legal frameworks change).

> *"…there are legal aspects that have to be clearly legal aspects, but likewise, there is no reason that we shouldn't also put in aspirational aspects and highlight that these are aspirational aspects."*

> *"This legal status of the policy is important, because we need certain regulations in a legally binding form. … But as a consequence of this legal status, we must be very careful about what we put in the data policy, because everything may become subject to legal enforcement. Including everything that you believe what should be a good idea or technical instructions on how to create FAIR data might become a problem in the future."*

Avoiding certain claims

Furthermore, some ExPaNDS partners emphasise the need not to make certain statements or claims in the data policy that could result in litigation or that could be impossible to justify legally. Specific concerns raised in relation to the draft data policy framework include the matter of 'validation of research results' (FE13), placing expectations on re-users of data, and mandating the citation of data (FE26) if data is licensed as CC0.

> *"When talking about validation of research results, care must be taken to ensure there is no misunderstanding: as [our facility] data policy points out, while we must do our best to support Users, we 'cannot be made liable for the consequences of any interpretation of the data' …"*

> *"[Our facility] does not have any contractual relation with persons using the data. We are simply not in the position to tell them what to do."*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Metadata and auxiliary data

Facilities recognise the need to consider legal aspects in relation to their use of metadata and auxiliary data. For example, feedback highlights GDPR[19] as a reason to specifically mention metadata in the policy.

> *"[Our facility] policy states that data is metadata is curated, processed and disseminated along with raw data. It always has been understood that this metadata includes information about the scientists that collected the data, e.g. personal information. But this has not been explicitly mentioned in the policy and that could be a problem with GDPR, so we should make it explicit."*

As mentioned already (see section 2.2.3), ExPaNDS partners highlight that there are many issues to address in relation to the legal aspects of auxiliary data. It is a complex area that is not straight forward for facilities or users. One partner provides an example of what their facility would need to put in place in relation to auxiliary data.

> *"…we will add to the policy that we encourage the users to provide additional information to improve the metadata. But we will need the licensing to be clear and probably put a statement that the users only be allowed to provide auxiliary data if the owner did waive the copyright of this data according to CC0. And we must add a provision that this auxiliary data must not contain any restricted material or personal information of any third party or anything that we would not be allowed to disseminate."*

Data ownership

ExPaNDS partners hold a range of different views in relation to the concept of data ownership. Some address the question of data ownership directly in their data policies; others specifically do not.

> *"[A section of our facility] data policy is dedicated to ownership of raw data and associated metadata, distinguishing those obtained via public research, covered by the data policy, and those obtained via proprietary research or specific agreements."*

> *"In [our facility data policy] there is no precise explicit literal clarification of who owns the data."*

Some facilities focus not on data ownership, but instead, on rights, such as the right to access data or to delete it.

> *"Access to the data is clearly defined in the data policy, while ownership is not clear in the data policy."*

[19] EU (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02016R0679-20160504

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

> *"Criteria of deletion: not specified in the current policy and there is no active mention in the policy of who the owner of the data is, thus who should be specifying that."*

At least one ExPaNDS partner argues that, from a legal perspective, the focus should indeed be on rights and not ownership.

> *…'ownership of data' is a problematic term which often creates lots of confusion. Because ownership is a legal term and ownership of data is something that strictly speaking does not exist. All this causes lots of discussion people say 'my data'. You need to explain that 'my data' has no meaning legally speaking. What is legally meaningful is the right of use."*

2.2.6 Framing Q5: Implementing the data policy framework

Facility infrastructure

Key to implementing data policy

For ExPaNDS partners, facility infrastructure is a key aspect of being able to implement the data policy framework. The consultation feedback references infrastructure issues in relation to many of the concepts and components discussed under framing questions 1 – 4 above (see sections 2.2.2 – 2.2.5):

- PIDs:

  > *"And questions of infrastructure also come into this – what do you have available, what can you actually do, and can you do what you want at the right level of granularity?"*

- Costs of FAIR data:

  > *"So in terms of infrastructure and costs, it is not so much that the infrastructure would have to change but more that the costs would be too prohibitive. You would have to buy and install the hardware, put a process in place, and there would also be the ongoing maintenance costs of ensuring the process is upheld."*

- Classes of data:

  > *"For raw data, [the infrastructure is] all there and this area is unlikely to become an issue. What is likely to be an issue is 'medium term' data. For example, for [our] users, there is no information about how long they can store their [medium term] data on the systems."*

- Citation:

  > *"What is more important is to improve the workflows and infrastructure and to do everything to make proper citation as easy as possible…"*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

- Alignment with the wider ecosystem:

> *"In [our country], we have, for example, an HPC infrastructure … and we really try to collaborate with that infrastructure. And it is in that interface that it is really important that we have similar approaches."*

Data access and analysis

Rights management and enabling access

ExPaNDS partners also highlight implementation issues related to data access. These include access rights management and having the appropriate protocols and systems in place to enable data access.

> *"… an authorisation framework (ideally, standardised) that could allow both beamline supervisor and Principal Investigator to grant access to the data to the selected persons without requiring them to pass a special registration procedure. This is, however, a difficult task for implementation…"*

> *"…access rights management is the responsibility of RI's Data Policy because it is system-dependent."*

> *"[Re FE21:] … it is easy to state that the data should be accessible using standard protocols. It's not so easy to implement that in practice: I'm not aware of any standard protocol to access data that is on tape."*

> *"[Re FE3:] There is room for improvement in the support to access … the data at [our facility]. … There is a data catalogue for the users to download their data, but we still need to integrate this into the workflows for each individual instrument."*

Data analysis environments

The draft policy framework makes several references to the analysis of data. However, at present, although at least one ExPaNDS partner does offer users a formal analysis environment, implementations related to data analysis remain a work in progress at many facilities.

> *"[Re FE3:] Still a lot of work to do to provide the users with analysis environments. Not very advanced in providing virtual environments for data processing. …And there are no data analysis capabilities connected to the data catalogue yet."*

> *"Concerning data analysis, it is also written in [our policy] that [our facility] aims at providing the means for reduction and/or processing of raw data. We have established a new group whose mission is data analysis. In a future version [of our policy], once [our facility] is more advanced in providing these means (thanks to ExPaNDS and other projects), it could be given the extent to which the facility will support the user using these means."*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Not all data policies address data analysis

While some ExPaNDS partners currently address data analysis in some form (i.e. including data reduction/processing) in their data policies, not all do. One concern is that implementations related to data analysis change rapidly, so any reference to support with software and tools would need to be fairly general.

> *"…still no mention to analysis. There is a general agreement for the policy to state the type of support provided. It might require a wording general enough to be applicable to the rapidly changing technology."*

Data storage and deletion

Storage period

ExPaNDS partners generally seek to commit to a minimum period (e.g. 5 years, 10 years) for the storage of raw data, i.e. rather than an open or indefinite period. For other types of data, where this is specified, the storage period may be shorter. Facilities are careful in their data policies to include caveats that can act as 'get out' clauses in the event the facility is not able to store the data as promised in the policy.

> *"Our current project is to create the infrastructure so that we can store the data for at least 10 years …"*

> *"[Our facility] data policy defines a global long-term retention period of up to 5 years (striving to 10 years) specifying that retention period could depend on the used beamlines/instruments…*

> *"[Our] current Data Policy does not oblige [our facility] to ensure all-time access to the data…"*

Data is not normally deleted

The consultation feedback indicates that, despite specifying minimum periods for how long they will strive to store data, to date, ExPaNDS partners have normally kept all of their raw data. Indeed, the topic of data deletion rarely features in facility data policies.

> *"…raw data … [our facility] keeps it all …"*

> *"…up to now, all data has been kept…"*

> *"[Re FE14:] The policy states that data and their associated metadata are kept for at least 10 years. This does not mean that everything is thrown away after 10 years. The criteria for data deletion are not yet defined at [our facility]."*

Exceptions

ExPaNDS partners do highlight a couple of situations where data is deleted:

1. When data does not comply with the data policy

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

"So far, [our facility] has not had to delete data, with the exception of cases which go against the data policy. In these cases which go against the policy, a landing page is left, but without metadata – rather, only a sentence explaining the reason that the record has been deleted."

2. In relation to other classes of data

"But a related discussed but as yet unsolved problem is around reduced data or processed data. This type of data is currently stored under … 'medium term' storage, and is deleted from time to time. When those data are deleted, there remain no digital record of them. Thus, provenance information is being lost. We imagine this is the same with every facility and it is difficult to imagine how facilities could actually do this..."

Keeping metadata is important for FAIR

As the second point highlights, ExPaNDS partners recognise that when data is deleted, there is the potential for important information needed for FAIR to be lost. This information is found in the metadata.

"…deleting data doesn't mean deleting metadata recorded in the data catalog. In our vision, in the event that data of an experiment are deleted, neither the landing page of the DOI assigned to the data nor the essential metadata of the data catalog should disappear."

At present, however, keeping the metadata of deleted data remains a theoretical issue, i.e. because ExPaNDS partners have not yet reached the stage of needing to delete data as standard practice. This situation is evident is facility policy making. Currently, even in those policies where data deletion is addressed, the issue of keeping metadata tombstone records for deleted data is not.

"The policy does talk about keeping data for 10 years, but it does not cover the creation of tombstone records for data that are deleted."

"[Our facility] data policy does specifically mention data deletion. However, the policy does not mention keeping digital footprints or tombstone metadata records."

Data deletion will become more important

This said, ExPaNDS partners recognise that data volumes are growing, especially for photon sources. In future, facilities are unlikely to be able to continue to store everything. As such, putting guidelines and systematic processes in place for data deletion and metadata retention are seen as important considerations.

"Re FE-14-15: Data deletion is an important issue."

"[Re FE15:] … There is the interest/thought in keeping the lifetime information of data and metadata even after deletion."

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

> *"At least if there is a systematic plan in place for what a facility will delete and why, then it is possible for the facility to trace back and say, 'well, yes, we did delete those data and these were the reasons'. At the moment, though, the assumption across facilities seems to be that they will be able to keep the data forever. Is this realistic? Certainly, people are looking to the future and at the data rates that are coming of detectors and thinking that the cost of just continually maintaining storage is so huge. There will come a point where this becomes an impossible task…"*

FAIR compliance

A new concept for facilities

FAIR compliance is an area of interest for ExPaNDS partners, especially as they look to the future.

> *"[Do you do any sort of audits of your data to check if it is FAIR?] Not at the moment, but it is on the wish list for the future!"*

However, it is also clear that it is a very new concept for facilities and that they have many questions about how FAIR monitoring and compliance can be implemented.

> *"…there is some FAIR validation to be done, but we are not sure how this validation is 'implementable'."*

> *"Re FE27: In what form and by whom? Internal or external audit? … Audits within a certification process?"*

> *"The FAIR monitoring and auditing system is still fuzzy…"*

Concerns around resources and costs

Particular concerns are the availability of tools and the costs involved.

> *"Monitoring the compliance of the FAIRness of the data might be very costly, and we are not sure…"*

> *"Re FE27: Certification as a trusted repository would be the best approach to do this. But we are by far not yet at the stage to do this now. We need to set up the repository properly and have everything work as we want…"*

> *"We are lacking several [resources for FAIR]: DOI and PID minting tools, FAIR metrics and implementations, data scientists and manager, budget ..."*

Important to retain flexibility

As with other elements of the draft data policy framework, ExPaNDS partners also raise concerns around ensuring flexibility is maintained for the facility, whether FAIR audits would have any legal aspects that require consideration, and if the data policy is the best mechanism for engaging with the issue of FAIR compliance.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

> *"[Re FE27:] … giving the freedom to the RI to define how to translate "regular" in practical terms is beneficial, as each RI will have different frequencies of audits."*
>
> *"[Re FE 27:] should the authorized legal entities be defined for audit explicitly?"*
>
> *"We understand [the goal] to establish the FAIR data principles as an obligation. We agree on this goal. But we are not so sure if facility's data policies are the best tool to achieve this goal. Other mechanisms might be more suitable. For example, the certification of the facility repository as a trusted repository might be a better tool to foster FAIR data principles: the certification process is more focused on the practices rather than the regulations and it is a dynamic process that more easily adopts to best practices rather than a static document that is just written once."*

Sanctions for non-compliance

On a broader note, concerning general compliance with the FAIR Principles, at least one ExPaNDS partner feels that lack of compliance should not affect users' access to RIs' resources.

> *"I see the use of compliance with [the] FAIR principles as a criteria to refuse access to RIs resources as a barrier."*

Compliance with policy

Compliance currently

The consultation feedback includes a mix of comments that relate to compliance with the data policy.

Firstly, there is the matter of whether compliance is addressed by the data policy explicitly, implicitly, or not at all.

> *"[Our] data policy specifies that 'deliberate infringements of the policy may lead to denial of access to data or metadata and/or denial of future beam-time requests at [our facility]'…"*
>
> *"[Our data policy] does not say much on compliance, and what it does say is more couched as a recommendation."*
>
> *"Policy compliance monitoring is not mentioned in [our] current [data] policy."*

Monitoring compliance

Secondly, even where facilities do mention compliance in the data policy, there are limited approaches in place for undertaking this monitoring. Most often, compliance is monitored through individual 'checks', e.g. by the proposal peer review panel, the user office, or by facility librarians.

Some ExPaNDS partners wonder what systematic process of compliance monitoring might be possible to put into place. Other partners feel that the current process of checks is adequate and

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

see no need to change their approach. Still others question the feasibility full stop of monitoring users' compliance with policy.

> *"What kind of process could be put in place to monitor compliance?"*
>
> *"[Re FE29], There are certain requirements on the users in the terms and conditions outside the data policy, such as providing a report and to publish the scientific results of a beamtime. This is monitored by the user office and indeed, non-compliance may have consequences on the future access to the facility. Beyond that, we don't think we need anything else."*
>
> *"[Re FE28]: … it is one thing to say [comply with policy], another to put in place what it needs to do."*

Restrictions on users for non-compliance

Thirdly, there are the questions of how 'heavy-handed' restrictions are/can/should be when it comes to users failing to comply with the data policy.

> *"[Re] FE29: RIs data policies should explain if and how the level of [compliance] contributes to evaluation of resource requests. 'Non-compliance may be a contributing factor in the refusal of further access' is way beyond what we would ever consider. Access to resources will always be based on scientific quality. Compliance or non-compliance might or might not be a minor factor in estimating scientific quality."*
>
> *"To what extent should restrictions be included in the event of non-compliance? This is a question that [our facility] has for ExPaNDS."*
>
> *"In a sense, [our facility] does use the data policy for compliance, for example, when it comes to getting users to cite data DOIs in publications. Generally, if a policy says something then people will do it. But, non-compliance has normally not been something that has been focused on in terms of individuals."*

Facility compliance with the data policy

Lastly, there is the issue of to what extent compliance with policy sits with the user and how much actually depends on the facility and its actions. In particular, ExPaNDS partners note that the facility staff, infrastructure, and implementation practices also need to comply with the data policy.

> *"What [our facility] has done is use the policy as a tool internally by people who create the infrastructure to guide development. In other words, the policy acts as a check for the facility. This has shown to be very useful at [our facility]."*
>
> *"[Re FE28:] Most of the workflows that implement the data policy are not performed by the users but by our own staff and facilities. It is data curation that happens at our instrument that is important. Moving data to the repository and applying access rules and all that are things that are done by ourselves and not by our users. So we don't need to monitor compliance by users, but we just need to put it in practice ourselves."*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

2.2.7 Other notable themes

Several themes not directly addressed by any of the framing questions also emerge from the consultation feedback.

Policy role

Purpose must be clear

ExPaNDS partners emphasise the need for the purpose of a data policy to be clear.

> *"What should be the main goal of a data policy? … It's a very central question for implementing a data policy."*

Perceptions of role differ

This said, there is also the recognition that ExPaNDS partners may hold different views on the role of a data policy, and that, even within the same facility, perceptions may differ.

> *"There seem to be differences in the conceptions on what is the role and purpose of a data policy and what it can achieve and what not."*
>
> *"Asking a bit around for the 'most important expectation from a data policy', answers were:*
>
> - *a clear definition of QoS [Quality of Service] for my datasets*
> - *a well-defined data publication process*
> - *I only need DOIs to reference my datasets*
> - *compliance with scientific best practice.*
> - *a clear commitment from management and scientists to make data FAIR for an unlimited timespan*
> - *maximize reuse of data, minimize duplication of research efforts, and minimize embargo periods"*

Articulation of policy role needs more attention

Aspects of the consultation feedback suggest that the question of the role(s) of a data policy is not adequately addressed by the draft data policy framework. In particular, there is some confusion around FAIR. Is the main purpose of the data policy framework to focus on FAIR data? What about other roles played by a facility data policy?

> *"There was a general discussion about what a Data Policy is, what it can do and what is its role: Some feel that there is a little bit confusion."*
>
> *"There are different aspects to be discussed: one thing is the concept of FAIR data and what we should do to make data FAIR data and all related things. Another thing is what to do with a data policy."*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Roles, responsibilities, and rights

Roles

ExPaNDS partners often refer to roles as being important to specify in the data policy. Some of the consultation feedback indicates that the draft data policy framework should take roles more into account. In some cases, ExPaNDS partners feel the framework should be more specific about roles; in others, more general.

> *"The classification of users appears to be centred around 'authorisation to do something with data'. That makes the policy framework a little nebulous. A more role-oriented classification might have been better (e.g. owner, curator, manager, consumer…)"*

> *"Who is responsible for documenting the metadata for this data – the beamline scientist, the PI, the user group?"*

> *"[Re FE25:] What is the meaning of PI? Maybe it would be better to give the freedom to specify the role of [being] responsible [for] the DMP, instead of allowing the PI to either be responsible or not."*

Responsibilities

Statements about with whom responsibility sits come up often in the consultation feedback. It is notable that these statements usually go hand in hand with references to roles.

> *"[Our facility] data policy is differencing obligations on Users (experimental team) and obligations on researchers who reuse data."*

> *"Having user responsibilities clearly set out could also be useful for areas such as compliance."*

> *"The data policy should also include the pattern by which responsibility for a data/software record is transferred to another person and how it can be ensured that no records exist without an owner…"*

> *"Re FE25:] The user group is responsible for the creation [of a DMP]."*

Rights

Rights also feature in the consultation feedback. As with responsibilities, rights are normally referenced in relation to roles. ExPaNDS partners often mention rights in a contractual or legal context — but not exclusively.

> *"The question is, in an open data policy context, how far along that workflow does [our facility] actually have the right itself to expose the data? There could be a point beyond which we would need to rely on the researcher to say 'actually, yes, we want to make this data open'."*

> *"About copying of the data belonging to the Principal Investigator: distribution of the copies is allowed explicitly for raw data. However… the contractual*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

obligations are implicitly indicated as a condition to distribute raw data derivatives. This can prevent the co-workers of the Principal Investigator to inherit the rights …"

"[Our national] law on the right to employees' inventions … contains an exception …. This exception is called … the professor's privilege, and … gives the employee the intellectual property rights to … the results from [their] research … It's … an ongoing debate whether research data is covered by this exception…"

"Access [to data] restricted to experimental team during the embargo period, except specific transfer or grant of rights by the main proposer…"

Make roles, responsibilities and rights clear

Some ExPaNDS partners make clear their view that a data policy should indeed pay particular attention to roles, responsibilities, and rights. One even mentions that these aspects were not deemed sufficiently addressed in the original PaNdata policy,[20] i.e. the policy framework on which ExPaNDS partners' current data policies are built.

"[Re FE22:] … that is what [our facility] data policy is all about: to specify rights, responsibilities, and obligations of all involved parties."

"[Re FE22:] …It was a recommendation of [our national funder's] scientific data experts who advised us for [our facility] data policy: it seemed to them that it was not highlighted enough in the PaNdata model. They suggested us to have a dedicated chapter [on] roles and responsibilities…"

Policy structure

As noted above, one ExPaNDS partner provides an example of how they were advised by external experts to structure their policy to make roles and responsibilities clear.

"During the long process that led to the adoption of its data policy, [our facility] gradually included different evolutions of form and content on the advice of [our national funder's] scientific data experts. For example, on the aspect of form, one of their recommendations was to organise the data policy document so that roles and responsibilities could be better identified."

Another ExPaNDS facility also expresses an interest in structuring their policies around roles and responsibilities.

"Once the revised content has been approved, it would be interesting to take that content and to try to restructure/reframe into a different document that might

[20] Dimper, R. (2011). D2.1: Common policy framework on scientific data. https://doi.org/10.5281/zenodo.3738498

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

be more readable for users. This document could also set out clearly who is responsible for what."

Policy audience

Two documents with two different audiences

The consultation feedback highlights the need to consider the audience for both the data policy framework and the data policy itself — and to recognise that these two audiences may differ.

> *"The policy framework should aim at persons involved in developing and implementing an actual data policy. These persons can be scientists (facility users), librarians, administrators, security or legal staff."*

> *"[Our facility] policy also sets out really clearly the purpose of the policy and also to whom the policy applies. .... Setting these points out clearly and early makes it easier for readers to work out what the policy is about and if it relates to them."*

Audience needs

ExPaNDS partners also provide specific suggestions on accommodating audience needs.

> *"The document is very much focussed on data scientists, and it might be worthwhile to make the recommendation a little less abstract and easier to understand for non-data-scientists."*

> *"Care should also be taken not to make the text too long or more complex, which could discourage people from reading it."*

Machine readable data policy

Especially when it comes to machine readability, there are questions about the audience for the data policy. ExPaNDS partners are not at all clear why a machine would need to read the policy or what the purpose of this would be.

> *"… [the] question of [a] machine readable policy leads to additional questions around who is reading the policy framework and what is its goal…"*

> *"But isn't the point of a policy to be understood by humans? Why would you want that a machine can read it?"*

> *"[Re FE9:] To be honest, I am not sure that I really understand what this is supposed to mean for a data policy. So, I don't know if this would be useful."*

2.2.8 Suggestions for revisions to the framework

Approach to the framework revision

The consultation feedback includes a large number of very specific suggestions for revisions to the draft data policy framework. It is not possible to present these all in detail here.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

As well, common to most consultations, we cannot and do not attempt to incorporate every suggestion into the final version of the data policy framework. Our focus is on those suggestions made most often by multiple partners. We also pay special attention to feedback that relates to issues we did not consider (i.e. sufficiently or at all) in the draft version of the framework. Appendix 3 lists the major changes incorporated into the final version framework.

A degree of pragmatism is needed when it comes to deciding which suggestions represent tweaks and which represent major rethinks, especially in light of what/how much can be done now in terms of revisions and what would require more time or, possibly, even further consultation. To this end, the 'final version' Framework presented in Chapter 3 represents current 'best effort'; it should not be seen as fixed in stone.

What to keep

As highlighted in section 2.2.2, ExPaNDS partners generally do see all nine themes of the draft data policy framework as relevant to thinking around data policy (i.e. even if partners do not always agree that all nine themes should feature in the data policy itself). To this end, partners specifically state that the nine themes should remain in the final version framework.

More broadly, ExPaNDS partners acknowledge the flexibility of having a framework to guide but not dictate their policy making.

Changes and clarifications

ExPaNDS partners put forth a range of suggestions for changes to the data policy framework. Redundancy across the thirty framework elements is a particular issue; for example, several partners highlight the similarities between FE16 and FE21 (i.e. both of which relate to enabling FAIR data). One partner argues that some elements can be combined for the sake of brevity.

> *"High-level it appears there are some repetitions in the elements and collapsing some of them together can make it shorter."*

ExPaNDS partners also pick up on specific terminology, for example, questioning the use of phrases such as 'compute power' (i.e. in Theme 1) and 'new user' rather than simply, 'user' (i.e. in FE16). The framework's use of 'PI' is another point of concern raised by multiple partners as is ambiguous wording, especially for lists (e.g. see FE11).

> *"There was the impression that PI in the framework is not always consistent and its meaning should be better specified."*

> *"[Re FE11:] RIs' data policies should apply to data which are 'generated, stored, and analysed using facility resources (e.g. instruments, compute infrastructure, software, staff)'. That's at least not very clear. I would interpret that as "data which are generated using facility resources AND stored on facilities storage AND analysed using facilities compute resources, but then the examples given don't really make sense. Alternatively it would be sufficient if one of the conditions is met for the policy to apply, which also doesn't make sense."*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

> *"It would help … to clarify if this 'generated, stored, and analysed using facility resources' wording is to be read as a series of ANDs or ORs."*

Along with terminology that is unclear, ExPaNDS partners find some elements of the framework unclear. Several partners highlight the definitions of classes of data (FE12) as confusing, and many are unclear about the new concept of auxiliary data. There is also very little understanding of FAIR compliance and what this might involve (FE27). Other references of FAIR in the framework (FE16) are seen as too abstract and worded in ways that do not make sense.

> *"FE-16 is very abstract … to me. The data policy should 'provide sufficient contextual metadata'?"*

As well, several partners comment on the meaning and use of the phrase 'digital footprint', questioning the framework's intention in relation to FE15.

> *"…the term 'footprint' seems to carry the wrong connotation because it is mostly defined as activity tracking."*

> *"[Re FE 15:] … 'digital footprint' is a bold term. It could be defined here as both reconstructions of the data from derivatives and recipe on how to re-acquire the data…"*

Consider removing

ExPaNDS partners do raise concerns about some elements of the framework that they see as impossible to implement. In particular, these relate to metadata, citation, and subsequent use of (open) data. Regarding the latter two, the argument is that the data policy cannot possibly enforce demands on either.

> *"…this requirement that the user should cite the data. It is impossible to implement this in the policy because we cannot enforce it."*

> *"[RE FE4:] … [our facility] does not have any contractual relation with persons using the data. We are simply not in the position to tell them what to do. It is important to understand that publishing something or putting it to Open Access means ceding control on it."*

In terms of the first point (i.e. around metadata), the concern is that the framework's implication is that the facility is tasked with providing full metadata, which is not possible in the case of user experiments.

> *"FE-21 similar to FE-16 sound to me like if the RI should provide all metadata which is not possible."*

One partner argues that creating a complete list of all grounds for restricting access to data (FE19) is an impossible task. Instead, it would be better to rely on the concepts of rights and responsibilities.

> *"[RE FE19:] … 'RIs should specify the grounds for restricting access to data…' I personally wouldn't phrase it that way. It's impossible to create a complete list*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

> *of grounds, so one will probably end up adding a statement like 'or any other good reason'. I would rather state that the policy should line out 'who has the rights to restrict data and who is responsible that access to data is compliant with legal regulations'."*

ExPaNDS partners also raise concerns around making statements on budget or financial plans, especially in any detail, and about the inclusion of costs, i.e. normally assumed to mean amounts (monetary value) paid or charged, in a data policy (FE23).

> *"We are interested to know why infrastructure and costs is a theme in the framework? Costs, in particular, how do these fit into a data policy?"*

> *"I don't believe [our facility] management would be willing or even be allowed to make any statements on details of budget plans in the data policy."*

Potential additions

Several ExPaNDS partners highlight a need for the data policy to cover more than just photon and neutron sources. One partner suggests incorporating the term 'related laboratories'.

> *"[Our organisation] is bigger than PaN sources, how do we consolidate that the data policy could cover more than the PaN sources?"*

> *"Key message: not only the Photon and Neutron facilities should be mentioned, because [our organisation] for example, also ha[s] beamlines that can be categorized as neither Photon nor Neutron facilities. So the 'related laboratories' term suggested being mentioned."*

There is also a suggestion to address the issue of AI-based data access and analysis.

> *"It was proposed to consider at some point how to cover the access of AI based applications to data. By this we mean, some AI analysis can require to analyse multiple datasets collected at RIs. How should access regulation work in such a case? General agreement to cover this in the data policy."*

3. Summary Perceptions of the Draft Framework

This section brings together the themes discussed in section 2.2 to provide a summary overview of ExPaNDS partners' perceptions of the ExPaNDS draft data policy framework. We examine areas of consensus and difference, consider views on the strengths and weaknesses of the framework, and address the question of whether or not the framework meets the needs of ExPaNDS partners.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

3.1 The Overall Picture

3.1.1 Broadly positive

Partners are broadly positive about many aspects of the framework, and indeed, as section 5.1 will illustrate, the framework's impact is already being felt in relation to facilities' future work plans.

This said, differences do remain around some aspects of the framework, and indeed, around its overall role and purpose. As we saw in section 2.2.8, there are specific areas in the framework where ExPaNDS partners feel there is room for improvement. Some themes in section 2.2 also highlight other areas of the framework that may need to be rethought.

3.1.2 Strong interest in data policy

Certainly, however, the consultation feedback and the level of engagement with the consultation process itself stand as evidence of ExPaNDS partners' strong interest in data policy. There is a clear recognition that having an appropriate data policy is fundamental to supporting good research data management.

There is also the awareness that the wider ecosystem changes, and that data policy, where it makes sense to do so, should incorporate and reflect these changes. Given that the PaNdata model on which facility data policies are built dates from 2011, many ExPaNDS partners welcomed the opportunity the consultation offered to explore developing trends in research data management and FAIR and to consider how these could affect data policy.

3.2 Areas of Consensus and Difference

Is there consensus across ExPaNDS partners about the framework/ are views consistent? If not, where do they differ?

3.2.1 Consensus on overarching aspects

ExPaNDS partners express consensus in relation to several overarching aspects of the data policy framework. They are unanimous in their view that the framework's role is to make recommendations, with the understanding that these may or may not be taken up by facilities. Partners are also very clear that implementation details are out of scope, and that the facility infrastructure is key to being able to implement the data policy.

Importance of a high level commitment to FAIR

ExPaNDS partners agree that a high level commitment to FAIR data is important, provided that this commitment is tempered by the realistic aim that data should be 'FAIR when it leaves the facility'. Also related to FAIR, all partners recognise the importance of metadata for FAIR. Indeed, they express similar views on the challenges of capturing and collecting metadata in the PaN RI context.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Specific revisions to the framework

The consultation feedback indicates that ExPaNDS partners are also in consensus in their recommendations that certain framework elements are revised. In particular, these elements include several under Theme 1 of the draft data policy framework. For example, partners do not understand the relevance of the data policy being machine readable (FE9), neither do they feel that the purpose of the data policy is to specify actions (FE5).

Most partners also have concerns about linking the maximisation of scientific impact (FE5, FE13) or issues related to validation (FE13) specifically to the data policy, i.e. as these matters relate to much more than just data policy. There is also broad consensus that the framework's definitions of classes of data (FE12) are not especially helpful, and at best, unclear.

In terms of specific framework elements to keep, partners share the view that assigning a PID to the data policy (FE1) would be helpful, especially in keeping track of versions of policies and being able to confirm which version was in place at a given time.

3.3.2 Where views differ

ExPaNDS partners' views differ in relation to many areas of the framework. Indeed, aspects of the discussion in section 2.2 set out these differences in detail, and we do not attempt a repeat here. However, it is useful to highlight some of the most notable differences, especially where views are strongly held. In general, however, it is important to note that different views on the data policy framework are not surprising, given the diversity of facility data policies (see section 2.2.1).

Status and role of the data policy

One fundamental area of difference is ExPaNDS partners' understandings of the role of the data policy. How facilities see the status of their data policy is also related to this point: the data policy can be a legal document, a best practice set of guidelines — or both. In relation to policy role and status, there is no consensus across ExPaNDS partners.

Rights versus data ownership

The question of data ownership is also treated very differently across ExPaNDS facilities. Some data policies specifically refer to the role of data owner, others refer to rights in relation to data (i.e. rather than ownership), and still others remain vague on the matter of data ownership. Tied into this matter of data ownership, are the issues of data, copyright, and licensing.

Certainly, what is clear from the consultation feedback, is that some uncertainty exists around the concept of data ownership. Given this, many would prefer a focus on rights and responsibilities in relation to data instead.

Debate over citation

Citation was yet another area where different views were strongly held. Some ExPaNDS partners are keen for the framework to cover this area because, for them, enforcing citation (i.e. and, therefore, the formal acknowledgement inherent in citation) is an important concern for their facilities. For other partners, citation is an academic norm that should be addressed by means other than policy, for example, training or awareness raising. There is also the argument that

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

some aspects of citation are beyond the control of either the facility or the user, and that the need to comply with these external demands may well override any demands of the data policy, creating a potential compliance issue.

Views on DMPs

ExPaNDS partners express a range of views on DMPs. Indeed, views range from the highly positive to the strongly negative. Certainly, while some are keen to progress work on DMPs and include these as an element of data policy, others have no plans at present to bring in the use of DMPs.

3.3 Strengths and Weaknesses

What do ExPaNDS partners see as the strengths and weaknesses of the framework?

3.3.1 Areas of strength

Broad themes

ExPaNDS partners see both the flexibility and breadth of the framework as positive aspects. While not all partners would necessarily include all nine broad themes of the framework in their data policies, partners can see the relevance of the broad themes to data policy making. To this end, the themes are seen as useful prompts and also as a way of harmonising thinking around data policy making across facilities.

Specific elements

ExPaNDS partners also see the framework's inclusion of specific elements, for example auxiliary data, PIDs for datasets, and FAIR, as important in reflecting external science policy developments, especially around FAIR data. The framework's explicit acknowledgement that a facility data policy needs to sit within a wider policy and legislative ecosystem is also highlighted as an area of strength. Furthermore, some partners see the inclusion of newer developments, such as DMPs, as very welcome.

3.3.2 Weaker areas

Too action focused

There are some areas where ExPaNDS partners feel the data policy framework is weaker. One concern is the framework's references to actions; partners see implementation details as out of scope. Most would prefer discussion of commitments rather than actions.

Not enough attention to roles and responsibilities

Related to this, some partners also argue that the framework should focus more on presenting the thirty elements in relation to roles, responsibilities, and (where appropriate) rights. Some ExPaNDS partners specifically note that this point about roles and responsibilities is insufficiently addressed in existing PaN data policy models.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Clearly articulate purpose

The role of a data policy is another area that partners do not feel is well addressed by the framework. Articulating the purpose of a data policy is seen as a fundamental part of policy making. Indeed, ExPaNDS partners are not even sure about the purpose of the framework itself: does it seek primarily to address FAIR data or, more broadly, all issues related to data policy?

3.3.3 Revision and rethinking

Specific suggestions

ExPaNDS partners provide a number of specific suggestions for revisions. These relate to changes, clarifications, terminology, and additions. Aside from these more straight-forward revisions, ExPaNDS partners also highlight aspects of the framework that they feel need removing or that they feel would be impossible to implement.

Inclusion of citation

The inclusion of citation is one point of debate, with some partners arguing this is an academic norm that cannot be regulated by a policy. Partners also note that not all aspects related to citation are within the control of either the user or the facility, potentially making compliance with policy impossible.

To whom does the data policy apply

Other 'impossible' elements relate to placing conditions or expectations on re-users of data. In this case, the question is to whom does the policy apply? While applying the policy to internal staff or external users are both seen as reasonable approaches, attempting to apply the policy to secondary data users, especially in the case of open data, is seen as impossible to implement from a compliance perspective.

Remove machine readability element

In terms of framework elements to consider removing, partners are unanimous in their confusion about why a data policy needs to be machine readable and how this is to be accomplished. As ExPaNDS partners see it, the audience for the data policy is human, not machine.

3.4 Expectations, Needs, and Priorities

Can/how well does the framework meet the expectations, needs, and priorities of ExPaNDS partners?

3.4.1 Overall, the framework meets ExPaNDS partners' needs

The consultation feedback suggests that, overall, the ExPaNDS data policy framework meets the needs of ExPaNDS partners. This finding assumes, of course, that the feedback from the consultation process is sufficiently taken into account and acted upon in producing the revised version of the framework.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

This said, the question of ***how*** the framework satisfies ExPaNDS partners' needs, i.e. through meeting differing expectations and priorities, is more nuanced.

3.4.2 Expectations of the framework differ

It is clear that ExPaNDS partners have different expectations of the data policy framework. Some see the framework as a detailed guide for the sorts of content a facility should cover in its data policy; in other words, the data policy itself should incorporate many, if not all, of the framework elements.

Other partners, however, see the framework as supporting discussions around data policy making. For these partners, the framework is not so much a list of potential things to include in a data policy, but rather, a starting point for discussion on areas that need to be thought about in relation to research data management and FAIR. It then becomes a separate question as to which of the framework elements to address by means of the data policy. The consultation feedback suggests that, with this approach to the use of the data policy framework, far fewer framework elements would find their way into the actual data policy. Instead, many would translate into actions for the facility, i.e. as part of implementation activity.

3.4.3 Flexibility of the framework is a key priority

A key priority for ExPaNDS partners is retaining the flexibility to develop their data policies according to national and local needs. Related to this is a strong desire to avoid including implementation details in the data policy framework, i.e. as doing so is seen as limiting flexibility.

In relation to their views on the data policy framework, ExPaNDS partners' feedback is somewhat contradictory when it comes to these closely linked issues. On the one hand, ExPaNDS partners do seem to feel that the framework can offer the necessary flexibility. On the other hand, partners highlight concerns that too many of the framework elements focus on actions, i.e. which partners equate with implementation matters. Perhaps the most straight-forward conclusion that can be drawn from this conflicting stance is that ExPaNDS partners believe there is room yet for even more flexibility in the framework.

3.4.4 Commitment to FAIR data

ExPaNDS partners are unanimous in their desire to commit, as far as is reasonable, to data that is 'FAIR when it leaves the facility'. To this end, partners do support the fact that the data policy framework incorporates the concept of FAIR, i.e. a concept missing from the original PaNdata policy model.

However, most facilities also emphasise that they are likely to include reference to FAIR as a high level commitment in the introduction or preamble to their data policies. Again, the desire is to avoid focusing on actions, and therefore, implementation issues. In this sense, it the principles behind FAIR that are the seen as the priority. This said, many facilities do have work planned or already underway on a range of FAIR-related implementation matters, including PIDs for data, metadata, data analysis, and software to support FAIR (see section 5.1).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

4. Key Policy Elements within a PaN RI Data Policy Framework (Final Version)

4.1 Overview

A policy framework might be defined as:

> "A policy framework is [a] document that sets out a set of procedures or goals, which might be used in negotiation or decision-making to guide a more detailed set of policies, or to guide ongoing maintenance of an organization's policies"[21]

A data policy framework aims to provide a supporting structure around which a data policy can be built.[22] Three components come together to form this supporting structure:

1. Principles
    - Principles are propositions that set out fundamental beliefs, behaviours or requirements.
    - In the case of the data policy framework, what is the point of the framework? What is being proposed and why?
2. Underpinning ideas
    - Underpinning ideas represent the approach and key concepts that serve as a basis or foundation.
    - These need to be relevant to the context in which the framework will operate to ensure that the framework is grounded in and applicable to a 'real' situation.
    - In the case of the data policy framework, in what context will the framework sit and how will it be used in practice?
3. Elements
    - Elements are the essential or characteristic parts.
    - A framework can only capture limited information about 'real life practice'. The focus is on those elements that are likely to play the most significant roles.
    - In terms of data policy, what elements must the framework address, and which can it afford to leave out?

A key point is that a framework is a supporting structure designed to leave flexibility for a range of different implementations to be 'built' around the framework, and requiring deviations from the

[21] Wikipedia (2020). Policy framework. https://en.wikipedia.org/wiki/Policy_framework
[22] Cambridge Dictionary (2020). Framework. https://dictionary.cambridge.org/dictionary/english/framework

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

frameworks' principles to be justified. Thus, when a data policy framework is applied, it is entirely possible for a range of different data policies to be the result.

The aim of the data policy framework within the ExPaNDS project is to provide a common framework to guide the development of compatible data policies across facilities. In particular, the purpose of the framework is twofold:

1. To promote a common coverage and approach to data policy so that the policies of ExPaNDS RIs have a consistent scope, allowing users to compare policies and enable cross-facilities working and data sharing.
2. To recommend the evolution of RIs' data policies to enable the production and publication of FAIR data.

In the rest of this section, we discuss some key principles and elements that should be considered in relation to RIs' data policy making. These elements are those in which RIs need to make choices on the level of commitments which they are prepared to make and as well as the obligations that they require of users.

In particular, we have taken into account the needs of enabling FAIR data in the light of the recommendations above. We have been guided on the structure of the principles by a CODATA note[23] on best practice for research data management policies, which gives a useful categorisation of the topics that a data policy should consider.

4.2 General Drivers and Principles

RIs wish to maximise their scientific value and their broader societal impact, while at the same time serving the requirements of their user communities. The construction and operation of user facilities are significant long-term investments for public sector research, and therefore, those facilities are obliged to seek the maximum return for the public expenditure. This includes the selection of user experiments that have the maximum scientific value from the allocation of instrument time. This has been traditionally measured in terms of high-impact publications and other research outputs (e.g. patents and products).

Facilities science has become more data intensive: with the volume and complexity of research data increasing as beam intensity increases, more automation has been introduced, and more sensitive detectors have been developed. Thus, the value of the experiment is increasingly encapsulated in those data. At the same time, the volume and complexity has meant that user communities have found it increasingly challenging to store and process those data at their home institutions. Consequently, facilities themselves have taken a role to help manage the data lifecycle directly, with data storage capacity, specialised analysis software, and often, significant computing resources.

This additional responsibility of user facilities over data has meant that RIs have needed to clarify the rights and responsibilities of their facilities and their user communities, and thus the need has

[23] Hodson, S., and Molloy, L. (2015). Current best practice for research data management policies. http://doi.org/10.5281/zenodo.27872

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

emerged for data policies, as public statements of the approach to data of the RI, to be acknowledged as part of the agreement for user access to the facility.

1. RIs should openly publish a data policy, including the period in which the policy is in force and when it is planned to be reviewed. A PID should be used to refer to the published version of the policy.

The RIs' data policies should seek to address the following objectives.

a) For facility users, the policy should clarify for users the access rights to data collected and managed at the facility and specify the obligations on users in the subsequent use of the data.

b) For facility support staff, the policy should scope the responsibility of the facility to supply the user with support to store, access, and analyse the data, within the experiment and beyond for subsequent analysis.

Notes:

- We avoid considering *ownership* of data in this framework as this is a complex concept with differing legal interpretations in different countries. Rather, we focus on *rights* and *responsibilities* over data collected and managed on RIs' resources as part of user experiments.
- A review date may be flexible and in response to changing circumstances, but it is recommended that a maximum period is specified to avoid the policy becoming outdated through lack of review.
- A PID for the data policy allows the publication of a definitive reference version for the RI and aids its findability. Appropriate PIDs to use would be those used to publish technical reports or other grey literature.
- Clause (b) above sets out what data the RI will take responsibility for, without committing resources or particular implementations.
- Data policies form part of the agreement of use of facilities; different RIs within different jurisdictions interpret the formality of this agreement differently, from a statement of intent to a contractual obligation. The RI should be clear on the status of the data policy within its terms and conditions of use.

Further, the incentive for RIs to maximise the scientific value of the use of the facilities forms a strong motivation for publishing experimental data for reanalysis and reuse. As a rare and specialised source of data, the potential for reuse would require specialised expertise. It is also a reasonable expectation that the user should be in the best position to exploit the experimental data results. Nevertheless, it is of value for the data to be made available for others to reanalyse and validate the results and to reuse the data within their own lines of research. Thus the policy should further the following additional goal:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

2. The RI's data policy should specify the commitment of the RI to ensure that experimental data is made available and reusable, including enabling access to the data beyond the experimental team as appropriate.

Facilities are usually not funders of research, but rather most experiments have a dual funding regime, where the RI's funders resource the facility operations and staff while the users are supported by grants from other funders. For a particular experiment, the data policies of both funders need to be respected. In ExPaNDS, we are focussing on national RIs, funded via major national public sector research bodies, and thus national RIs need to accommodate their requirements.

3. A RI's data policy should comply with their national research funder's policy. The data policy should be sufficiently flexible to accommodate the data policies of international funders, such as the EOSC, and those of users' funders.

RIs may also want to consider the interaction of their policies with those of publishers that are of particular importance to the RIs' user communities. Further, if facilities services are to be included within the EOSC ecosystem, the policy should also take into account the EOSC Rules of Participation[24] and other requirements, such as on-boarding into the EOSC marketplace.[25]

One feature of PaN RIs is that there is a shared user community. The extent that PaN facilities share a common user base has been demonstrated in the PaNdata-ODI project[26] and remains the case. Users will use instruments at different facilities, taking advantage of the different characteristics of instruments and the different capabilities of neutron and photon sources and other user facilities. If different facilities have different approaches to managing and sharing data, this forms a barrier to the integration and sharing of data, where the user can bring data from different experiments, and then publish the data in a reliable and consistent manner.

4. RIs should seek to align their data policies, within the constraints of divergent national funder policies and legal frameworks.

Machine-readable policies are recommended within the Turning FAIR into Reality (TFiR)[27] report to make interpreting policies easier for machine-to-machine access to data. This is not widely done currently, and there is no consensus on what format such a machine readable policy should take and how it should be processed and used. Consequently, this should be left as an option for a future review of RIs' data policies.

[24] EOSC Executive Board Rule of Participation Working Group (2021). EOSC rules of participation. https://op.europa.eu/en/publication-detail/-/publication/a96d6233-554e-11eb-b59f-01aa75ed71a1/language-en/format-PDF/source-184432576

[25] Onboarding into the EOSC marketplace requires several formal steps, including registering as a provider and registering the resource. For further details, see https://eosc-portal.eu/providers-documentation .

[26] Bicarregui, J., Matthews, B. and Schluenzen, F. (2015). PaNdata: Open Data Infrastructure for Photon and Neutron Sources, *Synchrotron Radiation News*, 28:2, 30-35, 10.1080/08940886.2015.1013418

[27] EC Expert Group on FAIR data (2018). Turning FAIR into reality. https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

4.3 Roles and Responsibilities

The data policy should outline the rights and responsibilities, with respect to the data policy, of the actors involved.

5. RIs should specify the rights and responsibilities of particular classes of actors involved in the experimental process.

To facilitate this, the data policy should identify the different classes of user and their roles, with their accompanying rights and obligations. A proposed set of core actors is given below with recommendations on their rights and obligations under a facility data policy.

Actor	Definition	Rights	Responsibilities
RI	Large-scale centre owning and providing access to specialised instruments and other resources (including staff) for research purposes.	• Develop data policy and conditions on the access to facilities' resources and outputs. • Monitor the impact of the data policy, review and refine approach to data management.	• To maximise the scientific impact of the use of its resources for its user community and the wider research community. • To respect the data policy requirements of funders and users. • To allocate resources as deemed appropriate to data management practices and to provide tools for the stewardship and sharing of experimental data as outlined in the data policy.
Facilities support staff	Staff employed at the facility to support research. This includes for example: user office staff, instrument scientists, computing staff, data stewards, facility librarians	• Access experimental data and metadata and modify it with additional metadata for data curation and data sharing purposes and to improve facilities' processes and performance.	• To respect the data sharing restrictions on experimental data • To maintain the long-term access to and stewardship of data • To maintain FAIRness of data as is practicable.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Principal Investigator (PI)	The main proposer of an experiment, who undertakes the decision making for the conduct of the experiment and acts as the main liaison with the facility.	• To steer and control the collection of experimental data. • To determine who has access to experimental data during the embargo period.	• To agree with the data policy of the RI • To ensure that data management planning for the experiment is completed and followed.
Experimental Team	The PI and any other persons to whom the PI assigns access rights for the conduct and analysis of the experiment. The experimental team will often include members of the facilities support staff.	• Access to the experimental data • Add to the experimental data from additional runs and subsequent processing actions	• To comply with the RI's data policy and data management planning for the experiment. • To provide accurate information to maintain the FAIRness of experimental data
Data re-users	Third parties accessing the experimental data for further scientific purposes.	• Access to metadata describing experiments as soon as is practical after the experiment. • Access to the experimental data after any embargo period.	• Ethical use of the data. • Acknowledgement and citation of the RI and experimental team.

4.4 Scope

The data policy should define the scope of its coverage and what it excludes.

4.4.1 Definitions

RIs' data policy scope includes definitions of terms to ensure that there is clarity in the scope and coverage of the policy. This should include definitions of fundamental concepts of the experimental process and the data held.

6. RIs should seek to clearly define the terminology used within the data policy and use this terminology consistently within the policy and, if possible, elsewhere in the RI's policies and practices.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Fundamental to defining the scope of a data policy is the classification of the data in scope, and as an experiment is undertaken within a facility, different categories of data are generated and used in the process. Common terms used for these different categories include:

- *Raw or primary data* generated directly from the use of instruments and stored on facilities' storage resources.
- *Reduced, processed, and analysed data* generated from raw data produced using facilities' compute and software resources.
- *Auxiliary*, *third-party, or user-provided data* to provide contextual information, in so far as it is owned by the facility, processed using facilities resources, or submitted by users to provide supplementary information.

For ease of understanding and the harmonising of policies, it would be beneficial if RIs were to harmonise the definitions for these categories and to use them as consistently as possible. However, the wider variation in understanding and practice at facilities, even between different instruments and experiments, means that the exact definitions of terms such as 'raw data', 'reduced data', and 'processed data' and consensus on their meaning is difficult, and in a constantly changing environment, may not even be useful.

Instead, RIs should seek to communicate the meaning in their own terms, and in terms that are familiar to the user community, so that definitions can be understood and compared. Further, they should also refer to the wider definitions of terms, for example, as given in the emerging EOSC Glossary.[28]

Figures 1 and 2 give an *illustrative example* the definitions of data as defined in the SOLEIL data policy[29] as representative of the definitions given in data policy.

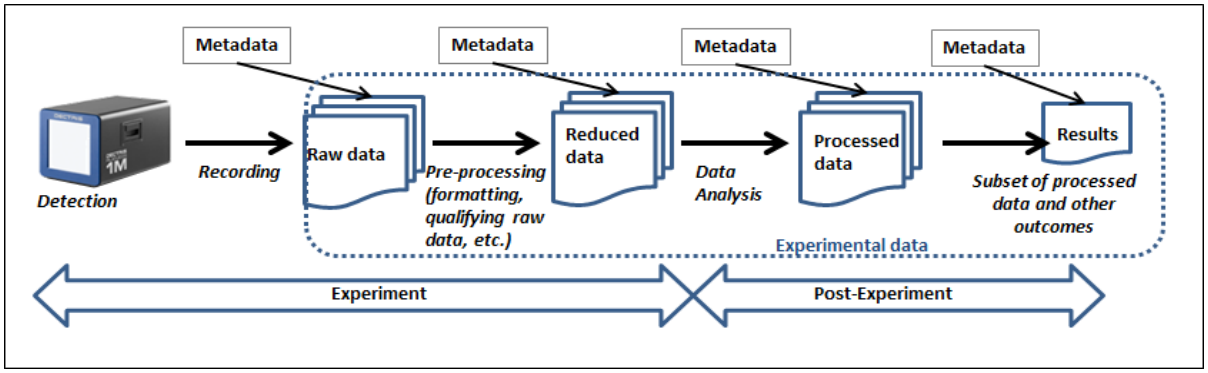

Figure 1: Simplified illustration of Classes of Experimental Data in the Science Life Cycle (from the Soleil Data Policy)

[28] EOSC Glossary Interest Group (2020). EOSC glossary December 2020. https://docs.google.com/document/d/1zcF95LChshSCv1biqS-AWG12VRyRyzZ5SKyMF5cyk3k/edit#heading=h.rzqwuch68sm . See also https://www.eoscsecretariat.eu/eosc-glossary

[29] Gagey, B. (ed.) (2018). SOLEIL data management policy. https://www.synchrotron-soleil.fr/en/file/11308/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

2.1. The term experimental data, see Figure 1, pertains to data collected from experiments performed on instruments. This definition includes (but is not limited to) data that are created automatically or manually by facility specific software and/or facility staff expertise to facilitate subsequent analysis of the experimental data.

2.2. The term raw data, see Figure 1, pertains to the experimental data that is recorded during experiments, as produced by the detection system, and cannot be derived from other persistent data.

2.3. The term reduced data, see Figure 1, pertains to the experimental data that is derived from raw data through pre-processing during experiments including (but not limited to) formatting and qualifying raw data and helping to decide on the continuation of the experiment.

2.4. The term processed data, see Figure 1, pertains to the experimental data that is derived from raw data along the analysis steps.

2.5.The term results, see Figure 1, pertains to a subset of processed data and other outcomes arising from the analysis of experimental data, excluding publications based on such analysis and intellectual property (IP) rights.

2.6. The term metadata, see Figure 1, describes information pertaining to data collected from instruments, including (but not limited to) the context of the experiment, the experimental team, experimental conditions and other logistical information.

Figure 2: Definitions of data classes as given in the Soleil Data Policy

A further additional data class of auxiliary data is identified in the PaNOSC data policy framework[30] for the wide variety of data resources which provide experimental context, given in Figure 3.

The term auxiliary data refers to data that provide contextual information regarding the experiment and its datasets but which are collected outside the context of the experiment conducted at the research facility, such as information about the sample images, provenance and preparation, data processing scripts, processing environment information such as software tools and versions used, etc.

Figure 3: Definition of auxiliary data from the PaNOSC Data Policy Framework

Different experiments may 'leave the facility' at different stages, so the scope of the policy may apply differently for different experiments. For the purpose of this framework, we shall use the term 'experimental data' to refer to all data related to an experiment using a facility's instrument, created and managed using a RI's resources, including auxiliary data items. We recognise that 'experimental data' has a number of subclasses, including raw, reduced, processed, and auxiliary, which are typically related to each other as in Figure 1 above. However, we shall not give at this

[30] Gotz, A., Perrin, J., Fanghor, H. et al. (2020). PaNOSC data policy framework. https://doi.org/10.5281/zenodo.3862701

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

stage a precise definition, but leave it to the facility setting the policy to use the most appropriate classifications of data.

4.4.2. Data within the scope of the policy

Facilities should specify the scope of the data policy. The facility experiment is typically within a wider scientific process that the user is undertaking, where other research actions are taken and data generated and analysed, which are outside the direct influence of the facility. These would have other intellectual property rights (IPR) conditions and are subject to different data policies. Thus, the RI should assert that its data policy applies to the experimental data within its purview.

7. The RI's data policy should seek to cover all classes of experimental data which are generated, stored, and analysed using the facility's resources (e.g. instruments, compute infrastructure, software, and staff) in the course of a user experiment.

Data gains meaning and context via accompanying information, known as metadata. Thus, the data policy should also apply to the metadata of the data in scope.

8. The RI's data policy should cover the metadata used to document experimental data identified in the policy's definitions.

Note that this does not constitute a commitment to handle all identified classes of data in the same way, have the same actors being responsible for them, or to keep them equally for the same time. However, the RI should avoid being responsible for experimental data that is not covered by the policy; if there is data associated with a user experiment that has uncertain status, it can lead to issues in terms of who is responsible for it and how it is handled. Thus the policy should provide statements on their treatment within the policy, even if they are subject to different regulation.

Facilities should consider, however, what policy actions can be taken to maximise the scientific value of the data resources. For example, facilities should consider how to maintain access to data that directly underpin or substantiate published research findings and are required for validation.

However, resource constraints (e.g. costs, ongoing storage capacity) are likely to limit the ability of facilities to commit to maintaining access to all data indefinitely, and so the data policy should indicate the criteria for the retention and deletion of experimental data.

9. The data policy should specify the retention policy for each class of experimental data, with a minimum retention period and criteria for deletion. As this includes auxiliary data, this also includes software and tools.

10. In the event that data are deleted, the facility should retain a record that the data existed. This could constitute a (metadata) record of their essential characteristics or a method to allow the reconstruction of the data. The facility should support as much as possible the provenance and validation of published research results in such circumstances.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Notes on elements 9 and 10:

- A non-exhaustive list of criteria for deletion might include: savings of resource in storage and funding; lack of evidence of use of the data; erroneous or otherwise nugatory data; request by a legal authority; data superseded by later results. The policy cannot pre-empt all reasons for deletion, but should have a process to select data for deletion and be able to provide a reason for the selection.
- Current data policies typically give a minimum retention period of 10 years for 'raw data', when it would have presumption of being retained and would not be deleted for reclaiming storage or because it was not being used. This of course does not mean that data may not be kept for a longer period.
- Data may also be deleted due to its funding source (e.g. commercially funded experiments); in this case, it may be deleted at the funder request (see section 4.6 on necessary restrictions below) and a public record may not be retained.
- Not all data may need a record after it is deleted; records of erroneous data or incorrect analyses may not be worthwhile.

Once the experiment is complete, then subsequent research actions are out of the facility's control. However, facilities can request that users continue to keep derived results data available and reusable as part of good research practice and within the community norms.

4.5 Enabling FAIR Data

Funder policy for publicly funded research within Europe now supports the aim of maximising scientific impact by releasing research results, in particular, research data, as openly, widely, and as early as possible. Furthermore, to encourage the use of data by third-parties, data should be interoperable with other data and software and reusable as widely as possible. Thus, this framework recommends that RIs should aim for the experimental data to be FAIR '*at the point of leaving the facility*'.

Thus, the data policy should commit the RI to manage the data in such a way as to make it as FAIR as is practical within funding, technological, and reasonable effort limitations, that is to:

- Support the ongoing **findability** of experimental data and their associated discovery metadata to uniquely identify experimental data to as wide a spectrum of users as possible.
- Support the ongoing availability of data and associated administrative metadata to allow users to **access** experimental data.
- Support the presentation of data and the provision of sufficient contextual metadata and supporting auxiliary data to maximise the opportunities for **interoperability** of experimental data with other data sources and with third-party software.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

- Support the presentation of data and the provision of sufficient contextual metadata and supporting auxiliary data to maximise the opportunities for **reuse** of experimental data in novel research contexts.

Applying FAIR adds value to the data for the experimental team as the prime users of the data as it makes the data better documented and accessible for their subsequent use. Further, by making data FAIR, there is an implicit commitment to making data as publicly available as possible, subject to the necessary restrictions as in section 4.6 below. By committing to Findability and Accessibility (using publicly available, globally unique persistent identifiers and providing access methods), the experimental data can be made open.

Additionally, we propose that data policies should include specific commitments in the way they will handle experimental data which would enable the production of FAIR data.

11. The RI's data policy should include commitments to enable the experimental data in scope to be FAIR. This may include the following commitments:

- ❖ The RI should provide the globally unique identification of experimental data via the association of an appropriate globally unique PID that conforms to the EOSC PID Policy.[31]
- ❖ The RI should annotate data with metadata in conformance to publicly available community and domain standards.
- ❖ The RI should support standard protocols for accessing data.
- ❖ The RI should provide data in formats conformant to publicly available standards.
- ❖ The RI should provide sufficient contextual metadata and auxiliary data.
- ❖ The RI should provide access to experimental data and associated metadata via human and machine-readable interfaces.

These commitments are inherent in the commitment to enable FAIR data, which implies an adherence to the principles given in the commonly accepted definition of FAIR data.[32] Note that these are not implementation decisions brought into the policy, but rather, they form an explicit commitment that subsequent implementation decisions will be guided by the FAIR Principles, subject to practical, technological, and financial limitations. Some of these commitments require input from the experimental team, which may limit the facility's ability to annotate the data.

The policy should clearly refer to FAIR as opposed to open, which can have a broader and vaguer interpretation and access to data within a FAIR context are qualified. Rights and responsibilities with regard to the use of data are best specified within a data licence, in accordance with the FAIR Principles (Principle R1.1).

12. The RI's data policy should specify a licence under which the data are made available.

[31] EU (2020). A persistent identifier (PID) policy for the European Open Science Cloud (EOSC). https://op.europa.eu/en/publication-detail/-/publication/35c5ca10-1417-11eb-b57e-01aa75ed71a1

[32] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data*, **3**:1. https://doi.org/10.1038/sdata.2016.18

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Data licences are a fairly recent addition to most RIs' data policies and there is no general consensus on the most appropriate one to use. Creative Commons CC0[33] has been proposed by the wider research community as an appropriate case for reusable data without complications, and it is recommended that RIs should consider using it.

4.6 Necessary Restrictions to Data Sharing

FAIR does not mean open, and so restrictions on the access to data can be applied, and there are circumstances where it become desirable or necessary to restrict data. For the sake of transparency, facilities should be as clear as possible on the nature and extent of restrictions that are applied to data.

13. RIs should specify a decision making process for determining the grounds for restricting access to particular experimental data.

Typical grounds for limited access would include:

- data arising from experiments which are not publicly funded (typically, expressly excluded from data sharing);
- restrictions applied by reasons of national security or prevention of criminality;
- access to personal sensitive data.

However, it is not possible to enumerate all possible reasons for restrictions, so a process should be specified for determining exemptions. Note that some of these grounds for limiting access also extend to other aspects of the policy, such as differing retention policies.

Grounds for restriction should not necessarily change other aspects of making the data as FAIR as possible. While if the experiment is commercially sensitive, the facility may not want to assign a publicly accessible identifier, nevertheless, by applying other processes to the data for making it FAIR, such as rich annotations, it becomes a more valuable asset to the funder.

Further, facilities experiments are a joint enterprise between the facility and a user group. While most facilities would assert rights to experimental outputs, nevertheless, the subject and conduct of the experiment is the result of the contribution of the researcher, who has a wider research goal in conducting the experiment. The goals of the researcher to further their personal research agenda should be reflected in allowing them 'first use' of the experimental results to further their research objectives, by providing restricted access for a time-limited period.

14. RIs should specify the time limit (an 'embargo period') for which users are allowed exclusive access to experimental data. This should also specify who can access the data (e.g. facilities staff), who can determine who should be given access rights, including who can lift the embargo to allow early publication, and the appeals process established to alter the embargo period.

[33] Creative Commons. CC0 1.0 Universal (CC0 1.0) Public Domain Dedication. https://creativecommons.org/publicdomain/zero/1.0/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

In practice, an embargo period of 3 – 5 years is typical for facilities, and during the embargo period the PI is given the right to assign access rights.

Further, RIs have an obligation to comply with national legislation that requires restrictions on data, including GDPR:

15. The RIs must comply with the relevant national legislation, notably that under the GDPR framework, in the handling of personal and sensitive data.

Some personal data is typically included in published metadata (e.g. name and institution); this is analogous to a bibliographic reference for a published article, which is vital in terms of maintaining the scientific record.

Regarding GDPR, Article 89(2)[34] provides exemptions for research: member states, through legislative action, may derogate from rights to access, rectification, restriction, and to object to processing for research purposes, given appropriate conditions and safeguards are in place. Article 17(3)(d)[35] additionally includes an exemption to the right of erasure in relation to processing for the purposes of scientific research. As Article 89 derogations (if any) may differ from country to country,[36] national RIs may wish to specify in their data policies any that apply in their jurisdiction.

4.7 Availability of Infrastructure and Responsibility for Costs

Facilities should recognise that supporting a FAIR data policy comes with the provision of infrastructure to support the retention and distribution of FAIR data. Thus the data policy should commit that the facility should support the provision of infrastructure as far as the coverage of the data policy for the facility specifies. This would then commit the facility to identify how resources might be allocated to cover these costs, within the reasonable funding limitations available.

16. The RI's data policy may consider the extent to which it commits to providing infrastructure to support the retention and distribution of FAIR data, for example:

- ❖ a storage and curation service to keep experimental data for the specified retention periods;
- ❖ a data discovery service to keep experimental data or its record findable;
- ❖ a data and metadata access and movement service to allow users to interrogate the experimental context and access experimental data.

[34] EU (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Article 89. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504&qid=1532348683434#tocId115

[35] EU (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Article 17. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504&qid=1532348683434#tocId25

[36] Boardman, R. and Molnar-Gabor, F. (2019). GDPR Brief: How is Article 89 implemented across the EU/EEA? https://www.ga4gh.org/news/how-is-article-89-implemented-across-the-eu-eea/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

The policy should also specify which infrastructure and costs would reasonably be expected to be incurred by users. Note that this commitment does not bind the facility to a specific implementation strategy, and any financial statements are not within the scope of the policy.

4.8 Data Management Planning Requirements

DMPs specifically designed for facilities experiments are not standard within RIs. However, if data are to be well-managed and curated and also made FAIR, there is a need for users to cooperate with facilities staff to estimate the storage and computational needs of the experiment, and to assist in providing accurate metadata. This will guide the facility to provide computational resources that the experimental team might need as well as enable the data to be made FAIR. This may include preparing a DMP. If so, this should be specified in the policy.

17. The policy should specify the requirements on users to participate in the facilities data management planning activities, including whether the experimental team is responsible for preparing a DMP.

This might include:

- Providing accurate information on the experiment for inclusion in the experimental metadata.
- Providing estimates on the storage and computation requirements for data storage and data processing.
- Providing additional experimental metadata to enrich the contextual information of the experiment (e.g. via electronic laboratory notebooks).
- Specify software needed by the experimental team to process the data.

Notes:

- This section refers to a DMP for the purpose of planning the sound management of data and computation within the experimental process of the facility. It does not include assisting the users to prepare or conform to a DMP required by external agencies, such as funders, publishers, or other research performing institutions.
- Preparing the DMP should not be the responsibility of the PI or user scientists alone, but rather, be a responsibility of the whole experimental team, including assigned facilities staff. Facilities staff are frequently best placed to provide relevant information.

4.9 Recognition and Reward for Data Usage

Credit should be given to the experimental team for the collection of data and its subsequent use. Facilities also need to be able to be recognised as contributing to science so that they can assess the impact and value of the use of their facilities and reward their staff appropriately. Thus, it is appropriate that the RI's data policy should encourage or specify how the use of experimental data should be recognised and cited.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

18. The RI's data policy should promote the recognition and citation of the use of facilities.

Specifically it should:

- Specify that use of experimental data should be acknowledged, including within citations.
- Encourage the citation of experimental data in publications by the experimental team and also re-users.
- Encourage re-users to contact the experimental team to express their interest in the experiment.
- Encourage the citation of software and instruments supplied by the facility.

The data policy cannot mandate this, as users are not within the control of the RI and can choose to take what they consider to be appropriate attribution steps. Further, because different publishers' editorial policies take different approaches to citing data and the use of experimental resources, this aspect also sits outside the control of RIs. However, by adding this element to the data policy, the user will be guided and encouraged to attribute the data suitably. Some facilities provide guidelines on the preferred form of data citation.

The data policy should be presented to users within a context that promotes FAIR data and that is supported with training material to ease the collection, exploitation, and citation of FAIR data. This is outside the scope of the policy itself, but should make its acceptance and use more straightforward to users.

4.10 Reporting Requirements, Compliance Monitoring, and Any Possible Sanctions

Data policies form part of the agreement of use of facilities; different RIs within different jurisdictions interpret the formality of this agreement differently, from a statement of intent to a contractual obligation. In all cases, the RI should be interested in the level of compliance to be able to assess the effectiveness of the data policy in achieving its goals and to monitor the acceptance of the data policy by the user community, and any resulting changes to their behaviour. Thus, data policies should indicate how compliance will be monitored, what reporting is required, and what sanctions may be imposed.

19. Users may be requested to report on compliance for previous experiments when applications for further access to the facility are received. RIs might consider that non-compliance may be a contributing factor in the refusal of further access.

The notion of FAIR data changes, and the appropriate and achievable level of FAIR-ness is likely to change over time. Achieving FAIR should not form a barrier to the introduction of emerging techniques and practices. If the data policy is specifying that experimental data should be FAIR, then again, the RI should assess the extent to which data is FAIR, judge its cost effectiveness, and consider any alternative approaches to raising the level of FAIR compliance.

20. RIs should have regular audits of their data management implementation and practices to evaluate compliance to the data policy and, in particular, the FAIR data principles.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Audit methods for FAIR data are emerging and their application within facilities is considered elsewhere in the ExPaNDS project.

The policy should also cover changes of circumstances or policy, for example arising from unforeseen restrictions on future budget or the continuity of service of the facility.

21. Changes or termination to the data policy will be given in sufficient time for PIs to take alternative action to provide alternative provision to comply with their funders' data policies.

This final element is especially important, given that PaN RIs have finite lifespans and resources. Thus, their ability to commit to manage data may be likewise time- and resource-limited.

5. Reflections on the Consultation and Next Steps

5.1 ExPaNDS Partners' Plans for Future Work

Based on the consultation feedback, this section overviews ExPaNDS partners plans for future work related to elements of the data policy framework. We consider what efforts are high priority and which remain as aspirations for further down the line. We also look at how priorities differ across ExPaNDS facilities.

5.1.1 High priority efforts

ExPaNDS partners plan to undertake future work in relation to a range of elements set out in the data policy framework. While not all ExPaNDS partners set out the same future work plans, a degree of overlap in intentions is evident.

Many high priority areas relate to implementation matters and FAIR data, including PIDs for data, metadata, data analysis, and software to support FAIR. Some facilities have specific projects related to these areas already underway or starting soon. Many of these future plans are influenced by the work going on in the ExPaNDS projects, e.g. around FAIR, data catalogues, and data analysis environments.

5.1.2 Actions for further in the future

Other framework elements, such as FAIR certification and PIDs beyond datasets, sit further away on the horizon for many ExPaNDS partners. While these activities are not priorities for the near future, they are areas that some facilities are watching with interest.

A big hurdle in terms of planning is that ExPaNDS partners see these 'horizon interests' as still very fuzzy in terms of implementation. They also feel that more foundational implementations (e.g. PIDs for data, metadata for FAIR) need to be put in place first. For example, referring to FAIR certification, one partner is very clear that this must wait for some time yet.

> *"…we are by far not yet at the stage to do this now."*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

5.1.3 Differing priorities

There are a couple of areas where it is notable that priorities differ across ExPaNDS partners: DMPs and data policy review.

Some facilities are very keen to progress work on DMPs; in particular, these include facilities where DMPs already feature, albeit not to full advantage. For other facilities, DMPs are not a priority or active area of planning at present, nor do these facilities have any concrete plans for future work around DMPs.

Statements on when ExPaNDS partners intend to review their current data policies are also very mixed. Even in light of the new ExPaNDS data policy framework, some facilities have no plans at present to review their data policies. Other partners express clear intentions to make changes in the near future, precisely because of the impact of the ExPaNDS framework.

5.2 Impact of PaN Data Policy Models and Frameworks

The consultation feedback references both the PaNdata and PaNOSC data policy models (i.e. as well as the ExPaNDS framework). While acknowledging the influence of the new PaNOSC model, most ExPaNDS partners do not plan to adopt the PaNOSC data policy outright. In this regard, partners see the ExPaNDS framework as offering greater flexibility. This said, many ExPaNDS partners feel that their data policies can benefit from ***all*** existing PaN data policy models (ExPaNDS, PaNOSC, and PaNdata), and they see no reason to restrict themselves to drawing on only one of these in their policy making.

5.2.1 Influence of the PaNdata and PaNOSC policy models

Comments from the consultation feedback reflect the strong influence that the PaNdata policy model[37] has to this day on data policy making in facilities.

> *"[Our] present [facility] data policy is based on the model formulated by the PaNdata Europe project."*

ExPaNDS partners are also very much aware of the PaNOSC data policy model[38] introduced in May 2020. However, the consultation feedback emphasises the point that ExPaNDS partners are not in a position to adopt the PaNOSC policy model outright.

One reason for this is that partners see that ExPaNDS facilities require more flexibility than the PaNOSC template provides.

> *"… PaNOSC did create a template – it was supposed to be like the work in ExPaNDS, i.e. a framework, but ended up being a template. So a template is a different thing, but actually, templates can be good. So, now we have the original PaNdata policy template, which for example [facility X] and [facility Y]*

[37] Dimper, R. (2011). D2.1: Common policy framework on scientific data. https://doi.org/10.5281/zenodo.3738498
[38] Götz, A., Perrin, J-F., Fanghor, H. et al. (2020). PaNOSC data policy framework. https://doi.org/10.5281/zenodo.3862701

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

have adopted, and now we have the PaNOSC template... It may be that for the PaNOSC facilities, the new PaNOSC template works. However, a lot of the ExPaNDS partners have indicated that they just couldn't adopt that template because it just would not fit."

"This data policy framework [i.e. the ExPaNDS framework] is more abstract in nature, i.e. then the PaNOSC framework. For example, there is no mention of specific technologies."

Nonetheless, ExPaNDS partners do see that the PaNOSC model offers a complementary approach.

"… it is absolutely clear that [our facility] isn't in a position to adopt all of the things in the [PaNOSC] policy, but we could still mention them in our policy."

"If we take the 30 element framework and the PaNOSC one and then combine these with what we already have at [our facility] and tweak it, then would should have a very good data policy covering what we want it to do in terms of enabling FAIR more than we have been."

Indeed, ExPaNDS partners have clearly given thought to the differences between the policy models of PaNdata and PaNOSC, and they see definite value in certain aspects of the PaNOSC policy. For example, partners highlight auxiliary data as an important concept missing in the PaNdata policy but introduced by the PaNOSC policy.

"…auxiliary data is not covered in [our] current [facility] data policy, i.e. because it was not covered in the original PaNdata policy framework."

"… for auxiliary data, we agree that this is a good idea that PaNOSC brought up and that is something important to be added."

5.2.2 Impact of ExPaNDS data policy framework

It is clear from the consultation feedback that the ExPaNDS data policy framework (even in its draft form) is already having an impact on facilities' data policy planning.

"It is generally a good framework, and the overall outcome of the consultation is of overall consensus and that the RI's policy can greatly benefit from the framework."

"[Our current data policy] is likely to change in the next future … and the framework will be valuable during that review."

"Thanks to the 30-elements framework, some additions / precisions that could be made in the next version of [our facility] data policy have been identified."

It is also notable that many ExPaNDS partners mention the consultation process itself as influential in spurring facilities on to reconsider their data policies.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

> *"The framework consultation already highlighted some weak points of the existing data policy."*
>
> *"The present consultation may help us to identify potential areas for a future review..."*
>
> *"There are some things based on the [consultation] discussion that [our facility] would seek now to change in its current version of the data policy."*
>
> *"After the chat today and looking at notes, it could be useful to chat to some colleagues within [our facility] about looking to review the data policy."*

One comment from the feedback even suggests a potential reach for the ExPaNDS data policy framework beyond the context of PaN RIs.

> *"It is a useful framework, and also one that could be useful for our national discussions … There is a lot of good stuff in the framework, so we could definitely take this and say, 'look, this is how some people are working on this and can we use this as a starting point'."*

5.3 Reflections on the Consultation Feedback

This section takes the opportunity to reflect on the consultation feedback, as well as the consultation process itself. Was the feedback on the framework as expected? Were there any surprises? Was the consultation process a useful exercise?

5.3.1 Deep engagement with the consultation process

Due to our 2019 Landscaping Survey[39] and because ExPaNDS WP2 has been engaged in work around data policy ever since, we already had a foundational understanding of ExPaNDS partners' data policies prior to undertaking the data policy framework consultations. This said, the consultation meetings provided an unprecedented opportunity to have detailed conversations with partners both about their data policies and their thoughts around policy making itself.

These two way exchanges were much more nuanced and rich in detail than the Landscaping Survey. As such, we emerged with a deeper understanding of ExPaNDS partners' approaches to data policy. Based on the consultation feedback, we also know that ExPaNDS partners' thinking was influenced by participation in the consultation.

While we did expect the consultation process itself to have an impact in terms of raising awareness across ExPaNDS partners about topics related to data policy, we were not necessarily expecting the high level of interest and engagement shown. The active participation of ExPaNDS partners was notable, not only in terms of the length and detail of the consultation feedback, but also in relation to partners' desire to engage deeply with the topics discussed.

[39] Ashton, A., Da Graca Ramos, S., Matthews, B. et al. (2019). ExPaNDS data landscaping survey. https://doi.org/10.5281/zenodo.3673811

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

5.3.2 Two-way dialogue and exchange of ideas

The consultation conversations were truly two-way in nature. Both ExPaNDS WP2 and ExPaNDS partner facilities came away with a deeper awareness of issues related to data policy making. While we had hoped that this would indeed be the case, we were also not sure at the outset how the consultation process would be received, especially as it made considerable demands on ExPaNDS partners' time.

One result of the thoughtful dialogues that took place was the emergence of themes that we had not considered (sufficiently or at all) in relation to the data policy framework, including, for example, matters such as role of the policy, policy audience, and policy structure.

5.3.3 Feedback that was as expected

In terms of the consultation feedback itself, a fair amount was along the lines of what we expected, i.e. due to both the Landscaping Survey results and our ongoing work in the area of data policy. For example feedback on specific framework elements such as PIDs, classes of data, and embargo reflected what we already knew from other avenues of discussion in the context of the ExPaNDS project. Indeed, neither were we surprised by the different views on and current state of these areas across ExPaNDS partners.

The desire to keep implementation out of the data policy has been raised previously by ExPaNDS partners, so again, this was not a surprise. Many of the themes that partners highlighted in relation to implementing the data policy have already been discussed in other contexts, for example, during the course of ExPaNDS' work with PaNOSC on developing their data policy model.

5.3.4 Unexpected aspects

Some aspects of the consultation feedback, however, were unexpected. Certainly, the feedback served to highlight just how lengthy and involved the processes of both the creation of a data policy and its review are in the PaN RI context. It was clear from comments in the feedback that, in particular, facilities do not see the review of data policy as something to be undertaken often.

This view drives home the point that the ExPaNDS project, while clearly having an influence on thinking around data policy, is realistic in its position that it does not necessarily expect the adoption of new data policies at partner facilities during the relatively short (i.e. 3 year) course of the project. Data policy review and adoption is a longer process.

Another area that emerged more strongly than we expected was the extent to which ExPaNDS partner facilities are aware of and prioritise the wider ecosystem of national policy and legislation. While we already had some sense of the significance of this ecosystem for national PaN RIs,[40] its dominance and impact is even more evident from the consultation feedback. For example, while ExPaNDS partners value harmonisation of data policy across PaN RIs, they are also clear that, as national facilities, the need for alignment with the national ecosystem takes priority.

What is not so clear from the feedback is whether the next level of alignment prioritisation is with local policy (i.e. other policy at the facility) or across the national PaN RIs. What is also notable is

[40] See Chapter 2 in: Matthews, B., McBirnie, A., Vukolov, A. et al. (2020). D2.1: Draft extended data policy framework for Photon and Neutron RIs. https://doi.org/10.5281/zenodo.4014811

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

that references to EOSC policy and other policy at the pan-European level rarely feature in the consultation feedback. Both of these situations suggest that additional conversations may be needed around the alignment needs of policy and how this plays out in national PaN RIs.

5.4 Shifts in Thinking around Data Policy

The consultation feedback illustrates clearly how the diversity of ExPaNDS partner RIs is reflected in their data policies and suggests that this trend will continue in future. The feedback also emphasises how deeply facilities are engaging with data policy issues, not only in policy making but also in relation to implementation decisions and activities.

5.4.1 Data policy diversity will remain

The consultation feedback highlights that the diversity found across ExPaNDS partners' data policies is highly likely to continue to remain a feature in future. A key reason for this is that the different ExPaNDS facilities have different views on and approaches to research data and research data management, especially when it comes to implementation matters. Some of these differences result from historical implementation decisions or differences in conception. For example, questions such as whether the policy applies to internal staff only or to external users or to raw data only or to all types of data are fundamental to both the design and implementation of a data policy.

5.4.2 Data policy is a powerful and important tool

The consultation feedback does provide some insight into how thinking about data policy may be shifting. ExPaNDS partners see data policy as a powerful and important tool, perhaps more than they have done in the past. Indeed, the data policy framework consultation process flagged up just how deeply many partners are thinking about data policy issues.

There is a definite recognition of the significance of FAIR and also an active awareness of facilities' aims to support FAIR data. As well, ExPaNDS partners are very aware of the national policy and legislative ecosystem in which they sit. Indeed, they are most likely to review their data policies in response to changes in that ecosystem.

5.4.3 Focus of future activity

In terms of activity, ExPaNDS partners' plans for future work in relation to data policy focus not (in most cases) on the review of the policy itself but rather on ensuring that key implementations are in place to support the data policy – for example, PIDs for data and metadata for FAIR. Work on data catalogues and data analysis environments is also a priority for the near future for many facilities.

ExPaNDS partners also express an awareness of other areas that they see as potentially important in the longer term future. These include DMPs (with some partners, interested in these now), additional uses of PIDs (again, with some partners interested in these now), and FAIR compliance. Of these areas, FAIR compliance is seen as the most fuzzy and distant on the horizon.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

5.5 Next Steps

This report represents the final output of the ExPaNDS project on data policy. Data policy is constantly evolving as it is subject to a changing legal, cultural, and technological environment. RIs' cycles of data policy review and revision are time consuming as they require wide user consultation and senior managerial approval, so changes are infrequent. The future plans on revisions to data policy reported in the consultation show that RIs vary between reviewing policy immediately to several years in the future. Nevertheless, the ExPaNDS data policy framework has been effective on influencing the direction of travel for the RIs, and both the ExPaNDS data policy framework and the PaNOSC model data policy will be reflected in the changes we would expect to see in data policy and implementation over the coming years.

ExPaNDS will continue to advocate and promote the definition and implementation of data policy as part of its ongoing aim of providing guidance on the realisation of FAIR data in practice, with a particular focus on areas where the implementation of data policy is unclear (as in section 5.1 above). These include:

- Priorities and standards for metadata to annotate an experiment in detail
- Best practices around the use of PIDs
- Exploration of how improved data management planning can complement the processes of facilities in collecting and managing FAIR data

Assessment and metrics allow RIs to evaluate their progress on the implementation of data policy, and also assist in determining the value to the facilities science of the policy. In the next stage of the project, we turn our attention to providing a suitable assessment framework for data management in facilities, including, in particular, with respect to FAIR data. Building on common standard approaches to assessment, this will provide a tiered approach, so that RIs can prioritise actions that most effectively move them towards the provision of FAIR data.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

References

Ashton, A., Da Graca Ramos, S., Matthews, B. et al. (2019). ExPaNDS data landscaping survey. https://doi.org/10.5281/zenodo.3673811

Bicarregui, J., Matthews, B. and Schluenzen, F. (2015). PaNdata: Open Data Infrastructure for Photon and Neutron Sources, *Synchrotron Radiation News*, 28:2, 30-35, 10.1080/08940886.2015.1013418

Boardman, R. and Molnar-Gabor, F. (2019). GDPR Brief: How is Article 89 implemented across the EU/EEA? https://www.ga4gh.org/news/how-is-article-89-implemented-across-the-eu-eea/

Cambridge Dictionary (2020). Framework. https://dictionary.cambridge.org/dictionary/english/framework

Creative Commons. Attribution 4.0 International (CC BY 4.0). https://creativecommons.org/licenses/by/4.0/

Creative Commons. CC0 1.0 Universal (CC0 1.0) Public Domain Dedication. https://creativecommons.org/publicdomain/zero/1.0/

Davidson, J., Grootveld, M., Whyte, A. et al. (2020). D3.3 policy enhancement recommendations. https://doi.org/10.5281/zenodo.3686901

Dimper, R. (2011). D2.1: Common policy framework on scientific data. https://doi.org/10.5281/zenodo.3738498

EC (2018). Prompting an EOSC in practice. https://ec.europa.eu/info/sites/info/files/prompting_an_eosc_in_practice.pdf

EC (2020). European Open Science Cloud (EOSC). https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/european-open-science-cloud-eosc_en

EC (2020). Progress on Open Science: Towards a shared research knowledge system. Final report of the Open Science Policy Platform. https://ec.europa.eu/research/openscience/pdf/ec_rtd_ospp-final-report.pdf#view=fit&pagemode=none

EC Expert Group on FAIR data (2018). Turning FAIR into reality. https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf

EOSC Executive Board Rule of Participation Working Group (2021). EOSC rules of participation. https://op.europa.eu/en/publication-detail/-/publication/a96d6233-554e-11eb-b59f-01aa75ed71a1/language-en/format-PDF/source-184432576

EOSC Glossary Interest Group (2020). EOSC glossary December 2020. https://docs.google.com/document/d/1zcF95LChshSCv1bigS-AWG12VRyRyzZ5SKyMF5cyk3k/edit#heading=h.rzgwuch68sm

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

EU (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:02016R0679-20160504

EU (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Article 17. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504&qid=1532348683434#tocId25

EU (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Article 89. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02016R0679-20160504&qid=1532348683434#tocId115

EU (2020). A persistent identifier (PID) policy for the European Open Science Cloud (EOSC). https://op.europa.eu/en/publication-detail/-/publication/35c5ca10-1417-11eb-b57e-01aa75ed71a1

Gagey, B. (ed.) (2018). SOLEIL data management policy. https://www.synchrotron-soleil.fr/en/file/11308/

Gotz, A., Perrin, J., Fanghor, H. et al. (2020). PaNOSC data policy framework. https://doi.org/10.5281/zenodo.3862701

Hodson, S., and Molloy, L. (2015). Current best practice for research data management policies. http://doi.org/10.5281/zenodo.27872

Matthews, B., McBirnie, A., Vukolov, A. et al. (2020). D2.1: Draft extended data policy framework for Photon and Neutron RIs. https://doi.org/10.5281/zenodo.4014811

Wikipedia (2020). Policy framework. https://en.wikipedia.org/wiki/Policy_framework

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data*, **3**:1. https://doi.org/10.1038/sdata.2016.18

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Appendix 1: Thirty Elements to Consider within a PaN Data Policy Framework (draft version)

Key Message

This guidance note presents 30 elements that should inform a data policy framework for photon and neutron research infrastructures. In essence, each PaN RI **needs to make choices** in relation to these 30 elements — both on the level of commitments that they themselves are prepared to make as well as with regards to the obligations that they will place on their users.

Rationale for Change

Developed within the ExPaNDS project,[41] the 30 element data policy framework set out below has a particular focus on enabling FAIR data. Currently, most PaN data policies are based on the original 2011 PaNdata Policy Framework,[42] which was drafted prior to the formal emergence of the FAIR principles.[43] Since then, the FAIR principles have come to play an increasingly significant role in both national and European science policy. Given this, it is both important and timely for PaN RIs to refresh their research data policies to reflect these developments.

Themes

The 30 elements of the framework sit within 9 broad themes identified as relevant to data policy:

1. General drivers and principles
2. Scope
3. Data sharing and necessary restrictions
4. Enabling FAIR data
5. Rights and responsibilities
6. Infrastructure and costs
7. Data management planning
8. Recognition and reward
9. Reporting and compliance

1. General Drivers and Principles

Maximise scientific value and public investment return

RIs wish to maximise their scientific value and their broader societal impact, while at the same time serving the requirements of their user communities. The construction and operation of user facilities are significant long-term investments for public sector research, and therefore, those facilities are obliged to seek the maximum return for the public expenditure. This has been

[41] Matthews, B., McBirnie, A., Vukolov, A. et al. (2020). D2.1: Draft extended data policy framework for Photon and Neutron RIs. https://doi.org/10.5281/zenodo.4014811
[42] Dimper, R. (2011). D2.1: Common policy framework on scientific data. https://doi.org/10.5281/zenodo.3738498
[43] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data*, **3**:1. https://doi.org/10.1038/sdata.2016.18

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

traditionally measured in terms of publications and other research outputs (e.g. patents and products). However, as facilities science becomes more data intensive, the value of the experiment is increasingly encapsulated in those data.

Data volume and complexity

The volume and complexity of PaN data has meant that user communities have found it increasingly challenging to store and process those data. Consequently, facilities themselves have taken a role to help manage the data directly, with data storage capacity, specialised analysis software, and often, significant compute power.

Need for data policies

This additional responsibility of user facilities has meant that RIs have needed to clarify the rights and responsibilities of their facilities and their user communities, and thus the need has emerged for data policies.

In relation to the need for data policies, framework elements 1 – 9 apply:

1. RIs should **openly publish a data policy**, including the period in which the policy is in force and when it is planned to be reviewed. A **Persistent Identifier (PID)** should be used to refer to the published version of the policy.

The RIs' data policies should seek to **address the following objectives**:

2. Clarify the **ownership and access** to data collected at a facility.

3. Specify the extent to which the facility will supply the user with **support to access and analyse** the data, within the experiment and beyond for subsequent analysis, to maximise the opportunity for the user to develop their science.

4. Specify the obligations on users in the **subsequent use** of the data.

5. Specify the actions the RI should undertake to ensure that experimental data is **made available and reusable** to maximise the scientific impact of the experiment

The RIs' data policies should **take the wider landscape into account**. This includes aspects such as funders, the shared PaN user community, and the need to integrate with initiatives such as the EOSC:[44]

6. The data policy should be **sufficiently flexible** to accommodate the data policies of RIs' national funders and users' funders.

7. A RI's data policy should **comply** with their **national research funder's** policy.

8. RIs should **seek to align** their data policies, within the constraints of divergent national funder policies and legal frameworks.

[44] EC (2020). European Open Science Cloud (EOSC). https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/european-open-science-cloud-eosc_en

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

9. RIs should consider making their policies **machine-readable**.

2. Scope

The data policy should define the scope of its coverage and what it excludes.

Definitions and common elements

As is standard in most policy documents, PaN RIs' data policies should include definitions of fundamental concepts to ensure clarity on the coverage and scope of the data policy. These would normally be placed in a 'definitions' section of the policy document. Where possible, RIs should seek to harmonise the definitions they use.

More generally, in the design of their data policies, RIs' should aim for similar broad coverage and scope, even allowing for the fact that specific details of RIs' data policies will differ.

In relation to this need for harmonised definitions and common coverage, framework element 10 applies:

10. RIs should seek to **harmonise the terminology** used, if possible, and **use common data policy elements**.

Data within scope

Fundamental to defining the scope of data policy is the classification of the data in scope. However, agreement on how to classify and define the various sorts of data related to experiments (e.g. raw, reduced, processed, analysed, results, etc.) has not been reached within the PaN community. Indeed, it is recognised that local traditions around defining data types are likely to remain in place for some time to come.

In relation to RIs' data policies, therefore, the most important thing is that the local definitions for data types are clearly set out in the data policy (see previous section on definitions).

RIs must bear in mind that the facilities experiment typically sits within a wider scientific process that the user is undertaking. This process may involve other data and data-related actions that are outside of the RIs' direct influence and that are subject to different data policies. The RIs can only make assertions related to facility data.

In relation to data within the scope of the data policy, framework elements 11 – 12 apply:

11. RIs' data policies should **apply to data which are generated, stored, and analysed using facility resources** (e.g. instruments, compute infrastructure, software, staff).

12. RIs' data policies should **cover the general classes of experimental data** (although specific definitions of these classes may differ, according to local tradition):

- **Raw data** generated directly from the use of instruments and stored on facilities' storage resources.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

- **Reduced data** generated from raw data produced using facilities' compute and software resources.
- **Processed data** generated from the use of instruments and produced using facilities' compute and software resources.
- **Auxiliary data** to provide contextual information, in so far as it is owned by the facility, processed using facility resources, or submitted by users to provide supplementary information.
- **Metadata** used to describe, contextualise, and constrain the data resources above.

Note that this does not constitute a commitment to handle all these classes of data in the same way, or to keep them equally for the same time. The policy should provide statements on their treatment within the policy, if they are subject to different regulation.

Maximising data value while recognising constraints

RIs should consider what policy actions can be taken to maximise the scientific value of data resources. For example, facilities should consider how to maintain access to data that directly underpin or substantiate published research findings and are required for validation.

However, resource constraints (e.g. costs, ongoing storage capacity) are likely to limit the ability of facilities to commit to maintaining access to all data indefinitely, and so the data policy should indicate the criteria for the selection of experimental data.

In relation to the maximisation of data value while recognising constraints, policy framework elements 13 – 15 apply:

13. RIs' data policies should **seek to maximise the scientific impact** of experiments through **enabling** the **validation** of research results and **maximising** opportunities for **reuse**.

14. RIs' data policies should **specify the retention policy** for each class of experimental data, with a **minimum retention period** and **criteria for deletion**. As this includes auxiliary data, this also includes software and tools.

15. **In the event that data are deleted**, the facility should **retain a "digital footprint"** of the data. This could constitute a (metadata) record of their essential characteristics or a method to allow the reconstruction of the data. The facility should **support** as much as possible the **provenance and validation of published research results**.

Once the experiment is complete, then subsequent research actions are out of the facilities' control. However, facilities can request that users continue to keep derived results data available and reusable.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

3. Data Sharing and Necessary Restrictions

FAIR data sharing

Funder policy for publicly funded research within Europe now largely supports the aim of maximising scientific impact by releasing research results, in particular, research data, as openly, widely, and as early as possible. Furthermore, to encourage the use of data by third-parties, data should be interoperable with other data and software and reusable as widely as possible. [45,46,47]

These aims draw heavily on the FAIR principles. To reflect these principles appropriately while also taking into account the reality that, once the experiment is complete, subsequent research actions are out of the facilities' control, we recommend that a realistic aim for PaN RIs should be for data that are FAIR "*at the point of leaving the facility*".

In relation to enabling FAIR data sharing, framework element 16 applies:

16. RIs' data policies should **enable** the experimental data in scope to be **FAIR**:

- **Support** the ongoing **findability** of experimental data and their associated discovery metadata to uniquely identify experimental data to as wide a spectrum of users as possible.
- Support the ongoing availability of data and associated administrative metadata to **allow** users, including new users, to **access** experimental data.
- Support the presentation of data and provide sufficient contextual metadata and supporting auxiliary data to **maximise** the opportunities for **interoperability** of experimental data with other data sources and with third-party software.
- Support the presentation of data and provide sufficient contextual metadata and supporting auxiliary data to **maximise** the opportunities for **reuse** of experimental data in novel research contexts

Where possible, RIs should seek to present the data policy to users within a context which promotes FAIR data and is supported with training material to ease the collection and exploitation of FAIR data. While is outside of the scope of the policy itself, this should make acceptance and use of the data policy more straightforward to users.

Data licence and access

RIs' data policies should clearly refer to FAIR as opposed to open, which can have a broader and vaguer interpretation. FAIR does not mean open, and facilities should be clear on the nature and extent of restrictions that are applied to data, which should be specified in a data licence.

[45] EC Expert Group on FAIR data (2018). Turning FAIR into reality. https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf
[46] EC (2018). Prompting an EOSC in practice. https://ec.europa.eu/info/sites/info/files/prompting_an_eosc_in_practice.pdf
[47] EC (2020). Progress on Open Science: Towards a shared research knowledge system. Final report of the Open Science Policy Platform. https://ec.europa.eu/research/openscience/pdf/ec_rtd_ospp-final-report.pdf#view=fit&pagemode=none

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

RIs data policies should also apply any relevant legislation around the handling of personal or sensitive data and should specify the grounds for restricting access to data.

In relation to access and data licences, framework elements 17 – 19 apply:

17. RIs' data policies should **specify the licence** under which the data are made available.

18. RIs should **acknowledge** the application of the **relevant national legislation** under the **GDPR framework** in the handling of personal and sensitive data.

19. RIs should specify the grounds for **restricting access to data**. Typical grounds would be:

- data arising from **experiments which are not publicly funded** (typically, expressly excluded from data sharing)
- restrictions applied by reasons of **national security** or **prevention of criminality**
- access to **personal sensitive data**

Embargo period

Facility experiments are a joint enterprise between the facility and users. While most facilities would assert rights to experimental outputs, nevertheless the subject and conduct of the experiment is the result of the contribution of the researcher, who has a wider research goal in conducting the experiment. The goals of the researcher to further their personal research agenda and objectives should be reflected in allowing the researcher "first use" of the experimental results, by providing restricted access for a time-limited period.

In relation to supporting the concept of "first use", framework element 20 applies:

20. RIs should **specify** the time limit (an "**embargo period**") for which users are allowed exclusive control on the use of experimental outputs. This should also **specify who can access** the data (e.g. facilities staff), who can determine **who should be given access rights**, and the **appeals process** established **to alter the embargo period**.

4. Enabling FAIR data

RIs data policies should include specific commitments on the way they will handle experimental data to enable the production of FAIR data. Note that these commitments are not implementation decisions brought into the policy, but rather form an explicit commitment that subsequent implementation decisions will be guided by the FAIR principles.

In relation to the commitment to enable FAIR data, framework element 21 applies:

21. RIs' data policies should include **commitments to enabling FAIR data**, which may include:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

- Provision for the **globally unique identification of experimental data** via the association of an appropriate globally unique PID that conforms to the EOSC PID Policy[48]
- Annotation of data with **sufficient rich metadata** in conformance to publicly available **community and domain standards**
- Support for **standard protocols for accessing** data
- Provision of data in **formats** conformant to **publicly available standards**
- Provision of sufficient **contextual metadata** and **auxiliary data**
- Provision of access to experimental data and associated metadata via **human and machine-readable interfaces**

5. Rights and Responsibilities

Actors

The data policy should outline the rights and responsibilities, with respect to the data policy, of the actors involved. An initial proposed set of actors includes RIs, facility support staff, principal investigators, experimental team, and data re-users; however, there may well be other actors to consider.

Rights

Rights often centre on issues such as the right to determine, develop, control, or access. For example, while the PI may have the right to determine who has access to experimental data during the embargo period, the RI has the right to develop a data policy that grants facility staff access to embargoed data, for example, for the purposes of adding metadata for data curation.

Responsibilities

Responsibilities and their focus will differ depending on the actors involved. The RI may well have its eye on big-picture responsibilities, such as maximising scientific impact, long-term data stewardship, or respecting national funder policy, whereas the PI and experimental team will tend to focus much more on responsibilities related to their particular experiment.

In relation to actors and their rights and responsibilities, policy framework element 22 applies:

22. RIs should **specify** the **rights and responsibilities** of particular **classes of actors** involved in the experimental process.

6. Infrastructure and Costs

RIs should recognise that supporting a FAIR data policy comes with the provision of infrastructure to support the retention and distribution of FAIR data. Thus the data policy should commit the facility to support the provision of infrastructure as far as the coverage of the facility's data policy

[48] EU (2020). A persistent identifier (PID) policy for the European Open Science Cloud (EOSC). https://op.europa.eu/en/publication-detail/-/publication/35c5ca10-1417-11eb-b57e-01aa75ed71a1

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

specifies. In turn, this commits the facility to identifying how resources might be allocated to cover these costs, within the reasonable funding limitations available.

The policy should also specify which infrastructure and costs would reasonably be expected to be incurred by users.

Note that these commitments do not bind RIs to specific implementation strategies.

In relation to infrastructure and related costs, policy framework element 23 applies:

23. RIs' data policies should consider the **extent** to which the **RI commits** to providing **infrastructure** to **support** the **retention and distribution of FAIR data**. Such infrastructure might include:

- a **storage and curation service** to keep experimental data for the specified retention period
- a **data discovery service** to keep experimental data or its record findable
- a **data and metadata access and movement service** to allow users (including new users) to interrogate the experimental context and access experimental data

7. Data Management Planning

Data Management Plans (DMPs) are not necessarily standard within facilities. However, if data are to be well-managed and curated and also made FAIR, there is a need for users to cooperate with facility staff to estimate the storage and computational needs of the experiment and to assist in providing accurate metadata. This may include preparing a DMP – if so, this should be specified in the data policy.

In relation to data management planning, framework elements 24 – 25 apply:

24. RIs data policies should specify the **requirements on users** to participate in **data management planning activities**. These activities might include:

- Providing **accurate information** on the experiment **for inclusion in the experimental metadata**
- Providing **estimates on the storage and computation requirements** for data storage and data processing
- Providing **additional experimental metadata**, using tools provided by the facilities (e.g. via electronic laboratory notebooks)
- Specifying **software needed** by the experimental team to process the data

25. RIs' data policies should **specify whether** the PI is responsible for preparing a **DMP**.

8. Recognition and Reward

RIs data policies should encourage or specify how the use of experimental data and facility resources should be recognised and cited.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

In relation to citation and acknowledgement, policy framework element 26 applies:

26. RIs' data policies should **promote the recognition and citation of the use of facilities**. Specifically, a data policy should:

- Specify that **use of experimental data should be acknowledged**, including within citations
- Encourage the **citation of experimental data in publications** by the experimental team and also re-users
- Encourage **re-users to contact the experimental team** to express their interest in the experiment
- Encourage the **citation of software and instruments** supplied by the facility.

9. Reporting and Compliance

Monitoring, including FAIR

RIs' data policies should indicate how compliance will be monitored, what reporting is required, and what sanctions may be imposed. For example, reporting needs around enabling FAIR data may include FAIRness audits.

Notification of changes

Data policies should also cover changes of circumstances or policy. These may be significant changes, for example arising from the termination of the data policy itself or from unforeseen restrictions on future budget or the continuity of service of the facility. Changes may also be minor, for example resulting from regular review of the data policy. RIs data policies should address the matter of timely notification of data policy changes to relevant actors, for example, PIs.

In relation to reporting, compliance, and change notification, policy framework elements 27 – 30 apply:

27. RIs should have **regular audits** on **compliance to FAIR data**.

28. Compliance to policy by users may be **monitored and checked**.

29. Users may be requested to **report on compliance for previous experiments** when applications for further access to the facility are received. **Non-compliance** may be a **contributing factor** in the **refusal of further access**.

30. Changes or termination to the data policy will be **given in sufficient time** for PIs to take alternative action to provide alternative provision to comply with their funders' data policies.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Appendix 2: Set of Framing Questions

This set of framing questions was used to provide an initial structure to guide the consultation meeting discussions. Not all of the consultations covered all of the questions, and additional topics (not included below) were often discussed.

Q1: The 'thirty elements' framework ➜ What are your thoughts on the broad, thematic areas covered by the 'thirty elements that should inform a data policy'? These broad areas include:

1. General drivers and principles?
2. Scope?
3. Data sharing and necessary restrictions?
4. Enabling FAIR data?
5. Rights and responsibilities?
6. Infrastructure and costs?
7. Recognition and reward?
8. Reporting and compliance?

Q2: Specific elements ➜ We would like to draw your attention to particular elements of the 'thirty elements' framework:

- Use of Persistent Identifiers
- Differentiation between raw and processed data
- Embargos and their length

Are these elements part of your current data policy/practice? Do you foresee that these elements will pose any particular challenges for your facility?

Q3: Impact of FAIR on the policy ➜ Do you feel that the 'thirty elements' engage with all aspects of FAIR adequately? If you compare the ExPaNDS 'thirty elements to inform a data policy' with your current data policy, can you distinguish where changes have been made (i.e. from the original PaNdata policy framework)[49] to accommodate FAIR principles?

Q4: Consequences of legal frameworks ➜ Do you feel that the 'thirty elements' framework can accommodate existing legal frameworks (e.g. GDPR, licensing) sufficiently and appropriately?

Q5: Implementing the 'thirty elements' framework in practice (bear in mind that the framework is 'best practice' guidance, i.e. some aspects may yet be aspirational for facilities):

- Are there any aspects of the 'thirty elements' framework with which you would disagree or that you consider unhelpful?
- Are there any aspects that would form a barrier to users or that you would find difficult/impossible to implement?

[49] Dimper, R. (2011). D2.1: Common policy framework on scientific data. https://doi.org/10.5281/zenodo.3738498

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

- Are there any tools/roles that you would need in place to implement the 'thirty elements' framework?
- Based on your understanding of the 'thirty elements' framework, are there elements that you consider mandatory/optional? Are there elements that, in the case of your facility, you would choose not to address (i.e. bear in mind that the 'thirty elements' framework sets out policy-related elements about which the facility needs to make a decision or choice. Thus, a valid choice would also include the choice not to act on an element.)

Q6: Missing elements ➜ Is anything missing from the 'thirty elements' framework that you would like to see included?

Q7: Plans for your own data policy ➜ Can you update on your plans for your own data policy? Do you have plans to make any specific modifications in light of the 'thirty elements' framework?

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.

Appendix 3: List of Major Changes Made to the Final Version ExPaNDS Data Policy Framework

1. In overview, clarified the purpose of ExPaNDS providing a DP recommendation.
2. Removed the notion of Ownership as not helpful. Instead talk of rights and responsibilities over data collected and managed on RI's resources as part of User Experiments.
3. Made Element 2/3/4 sub-parts of Element 1 and added notes on the element.
4. Element 2 (was 5): changed "action" to "commitment" and clarified that it includes sharing with third parties as appropriate.
5. Merged Element 7 and Element 8 into a new Element 3.
6. Element 9 on machine readable policies deleted.
7. Section 6 on responsibilities moved forward to section 3.
8. Element 6 (was 10) restricted to cover only the terminology used within the Facility and not to cover the harmonisation across facilities.
9. Rewrote section on data terminology, including dropping section on ExPaNDS Glossary.
10. Clarified element 7 (was 11).
11. Changed element 8 (was 12) to make it apply to Metadata, and took out the data subclasses.
12. Deleted Element 13 – as repetition of Element 5 (now 2)
13. Added notes on element 9 and 10.
14. Renamed Section 5 Enabling FAIR Data and moved element 21 into it as a new element 11
15. Note on CC0 added after Element 12
16. Renamed Section 6 Necessary restrictions to data sharing. Moved elements 13, 14, and 15 into this section.
17. Section 6.7 – made some changes to clarify that this does not require an explicit financial statement in the DP
18. Section 6.8 – merged the two recommendations into one and added clarifications on the scope and responsibilities of the DMP.
19. Section 6.9 – added some clarifications
20. Section 6.10 – merged elements 20 and 21.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641.