ExPaNDS Guidance Note:

Key Policy Elements within a PaN RI Data Policy Framework

The ExPaNDS Data Policy Framework

Aim of the data policy framework

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.

Nine broad themes and 21 policy elements

Across nine themes, the ExPaNDS data policy framework discusses key principles that should be considered in relation to RIs' data policy making. The nine themes are:

1. General drivers and principles
2. Roles and responsibilities
3. Scope
4. Enabling FAIR data
5. Necessary restrictions to data sharing
6. Availability of infrastructure and responsibility for costs
7. Data management planning requirements
8. Recognition and reward for data usage
9. Reporting requirements, compliance monitoring, and any possible sanctions

Out of this discussion, 21 policy elements emerge to create the framework. These elements are presented in context in the sections below (numbered, in bold) and in summary in the Appendix to this Guidance Note. The elements are those in which RIs need to make choices on the level of commitments which they are prepared to make as well as the obligations that they require of users. In particular, in line with its two-fold purpose, the framework takes into account the needs of enabling FAIR data.[1]

[1] We have been guided on the structure of the principles by a CODATA note on best practice for research data management policies, which gives a useful categorisation of the topics that a data policy should consider. See: Hodson, S., and Molloy, L. (2015). Current best practice for research data management policies. http://doi.org/10.5281/zenodo.27872

Theme 1: 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 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:

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[2] and other requirements, such as on-boarding into the EOSC marketplace.[3]

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[4] 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.

---

[2] 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

[3] 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 .

[4] 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

Machine-readable policies are recommended within the Turning FAIR into Reality (TFiR)[5] 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.

Theme 2: 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.

[5] 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

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.
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.

Theme 3: Scope

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

3.1 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.

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.[6]

Figures 1 and 2 give an *illustrative example* the definitions of data as defined in the SOLEIL data policy[7] as representative of the definitions given in data policy.

[6] 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

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

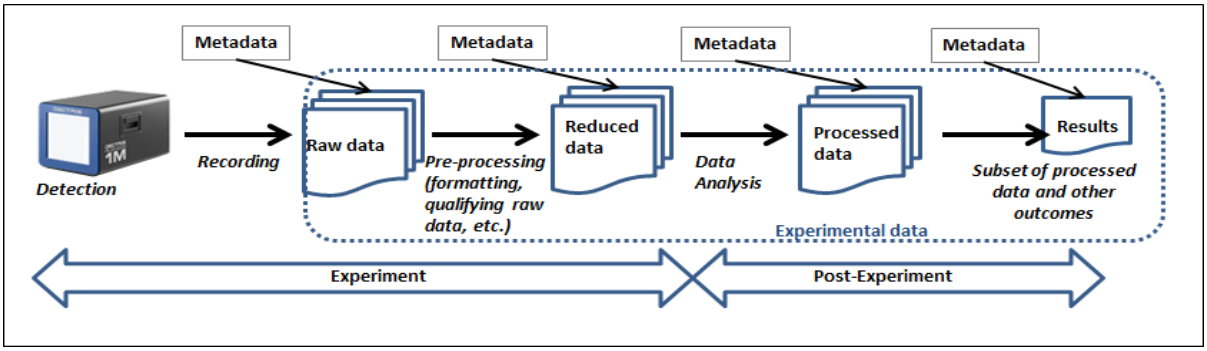

Figure 1: Simplified illustration of Classes of Experimental Data in the Science Life Cycle (from the Soleil Data Policy)

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[8] 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

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

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

3.2 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.

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.

Theme 4: 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.
- 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.[9]**
- **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.[10] 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.

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[11] 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.

Theme 5: 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.

[9] 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

[10] 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

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

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.

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)[12] 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)[13] additionally includes an exemption to the right of erasure in relation to processing for the

[12] 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

[13] 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,

purposes of scientific research. As Article 89 derogations (if any) may differ from country to country,[14] national RIs may wish to specify in their data policies any that apply in their jurisdiction.

Theme 6: 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.**

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.

Theme 7: 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.

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

[14] 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/

- 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.

Theme 8: 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.

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.

Theme 9: 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.

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.

The text of this Guidance Note is extracted from the ExPaNDS final deliverable on data policy: McBirnie, A., Matthews, B., Gagey, B. et al. (2021). D2.3: Final Data Policy Framework for Photon and Neutron RIs. Available at: https://doi.org/10.5281/zenodo.5205825 .

Date of Guidance Note: January 2022.

Appendix: Summary List of the 21 Elements that Comprise the ExPaNDS Data Policy Framework

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.
- 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.