D2.6 Self-evaluation Photon and Neutron RIs for FAIR data certification

Document Control Information

Settings	Value
Document Identifier:	D2.6
Project Title:	ExPaNDS
Work Package:	WP2
Work Package Lead	UKRI
Document Author(s):	Simon Lambert (UKRI), Abigail McBirnie (UKRI), Brian Matthews (UKRI), Stephen Collins (DLS), Maik Fiedler (HZDR), Brigitte Gagey (SOLEIL), Heike Görzig (HZB), Oliver Knodel (HZDR), George Kourousias (Elettra), Anders Markvardsen (ISIS), Carlo Minotti (PSI), Majid Ounsy (SOLEIL), Krisztian Pozsa (PSI), Silvan Schön (DESY), Nicolas Soler (ALBA), Darren Spruce (MAX IV), Andrei Vukolov (Elettra)

Document Contributor(s):	Christoph Deinert (HZDR), Patrick Fuhrmann (DESY), Alejandra Gonzalez Beltran (UKRI), Marta Gutierrez (EGI), Uwe Konrad (HZDR), Roberto Pugliese (Elettra)
Doc. Version:	1.0
Dissemination level:	Public
Date:	19/12/2022

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

Abstract

This ExPaNDS project deliverable describes a FAIR self-assessment undertaken by the ten ExPaNDS partner Photon and Neutron Research Infrastructures (PaN RIs) over the three-month period July – September 2022. After reviewing selected examples of existing FAIR evaluation frameworks designed to enable assessment at different levels (dataset, repository, and organisation), the report describes the evaluation approach adopted for the ExPaNDS FAIR self-assessment. As no existing framework met our specific need to focus on FAIR workflows and processes in PaN RIs, it was necessary to select, combine, and adapt existing frameworks. Supported by four underlying guiding principles, our approach drew heavily on the FAIR Principles, the RDA FAIR Data Maturity Model, and FAIRsFAIR's CoreTrustSeal+FAIRenabling framework. Post-evaluation feedback from ExPaNDS partners indicated that they found the FAIR self-assessment a useful and valuable exercise for understanding current levels of FAIRness at their facilities and for articulating what implementations they have in progress or planned to support FAIR in future. A key output of the ExPaNDS FAIR evaluation is the collected self-assessment reports from the ten partner facilities. These reports are published openly and in full as part of the deliverable. In addition, the self-assessments are supplemented with some high-level observations on the state of the FAIR journey across the ExPaNDS facilities.

Document Log

Version	Date	Comment	Author/Partner
Complete Draft	27/10/2022	Internal review version	Abigail McBirnie (UKRI), Simon Lambert (UKRI)
1.0	19/12/2022	Final version for submission	Simon Lambert (UKRI) Abigail McBirnie (UKRI) Brian Matthews (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

ACME-FAIR	Assessing Capability Maturity and Engagement with FAIR Enabling Practice
CMMI	Capability Maturity Model Integration
CTS	CoreTrustSeal
DCC	Digital Curation Centre
DESY	Deutsches Elektronen-Synchrotron
DLS	Diamond Light Source
DMP	Data Management Plan
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
HZB	Helmholtz-Zentrum Berlin
HZDR	Helmholtz-Zentrum Dresden-Rossendorf
KPI	Key Performance Indicator
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
PaN	Photon and Neutron
PEB	Project Executive Board
PID	Persistent Identifier
RDA	Research Data Alliance
RDI	Research Data Infrastructure

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

REST	Representational State Transfer
RFO	Research Funding Organisation
RI	Research Infrastructure
RISE	Research Infrastructure Self-Evaluation
RPO	Research Performing Organisation
STFC	Science and Technology Facilities Council
TFiR	Turning FAIR into Reality
UKRI	UK Research and Innovation
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 *D2.6: Self-evaluation Photon and Neutron RIs for FAIR data certification*, describes a recent FAIR self-assessment undertaken by the ten ExPaNDS partner PaN RIs. As well as providing background to and rationale for the evaluation approach used, deliverable D2.6 presents the resulting FAIR self-assessment reports from each ExPaNDS facility and draws out overall impressions of the exercise, including suggestions for improvements to the self-evaluation process and possibilities for next steps. Bearing in mind that the deliverable will be of interest to a range of readers, we also include some brief guidance (see end of Chapter One) on how readers can read the deliverable to meet differing needs and interests.

Background and context

A key role of the ExPaNDS project is to promote and embed the FAIR (Findable, Accessible, Interoperable, Reusable) Principles within the national Photon and Neutron Research Infrastructures (PaN RIs) across Europe. One aspect of this work focuses on developing policy and guidance; however, another important aspect centres on supporting practice, including the integration of the FAIR Principles into PaN RI workflows and processes. FAIR evaluation provides a means for facilities to examine these workflows and processes in a systematic way to better understand what they are doing well and where there may be potential for improvement.

Existing models suggest that FAIR evaluation serves two main purposes: to provide 'pass/fail' indicators in relation to the various components of FAIR (i.e. F, A, I, and R) and to measure progress along a 'journey' towards FAIR. Through the application of current FAIR assessment frameworks, it is possible to evaluate FAIRness at the level of the dataset, the repository, or the organisation; however, at present, no existing approach offers the means to evaluate FAIR specifically in relation to workflows and processes of the sort that feature in the context of PaN RIs.

Aims and purpose

Because PaN science requires PaN RIs, the processes and workflows of such facilities become absolutely crucial in enabling FAIR. If facility processes, workflows, and data management practices do not support FAIR across the experimental lifecycle, then PaN researchers have little hope of ending up with FAIR data from their experiments.

A task set out in the ExPaNDS description of work sought to support ExPaNDS partner RIs to undertake and report on an open FAIR self-assessment. For the reasons set out above, we chose to focus on facility workflows and processes. There was an inherent recognition that the outcomes of the exercise would differ for each ExPaNDS partner: what would be important would be what the facilities took away from the exercise for themselves, especially in terms of new insight and potential avenues for future development.

Evaluation approach

As outlined above, no existing FAIR evaluation approach was a perfect match for our needs. Thus, it was necessary to select, combine, and adapt existing frameworks to tailor an evaluation method especially suited to PaN RIs. To ensure our adaptation did not simply cherry-pick what seemed most relevant, we applied four underlying principles:

1. Our approach must link directly back to the FAIR Principles

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

2. Our approach must identify and take advantage of what existing FAIR evaluation frameworks have to offer

3. Our approach must take into account the relationships between existing FAIR evaluation approaches

4. Our approach must relate clearly to the processes and practices of PaN RIs.

Additionally, as set out in the ExPaNDs FAIR evaluation task description, our evaluation needed to be open in nature and take the form of a self-assessment.

In line with these four principles, we chose FAIRsFAIR's CoreTrust+FAIRenabling framework as the starting point for developing the ExPaNDS FAIR self-assessment questionnaire. Although it involves an element of external peer review, at its core, CoreTrustSeal+FAIRenabling features a strong self-evaluation component. As we explain in detail in our report, this framework could be most readily and sensibly adapted to focus on the workflows and processes that come into play across the experimental lifecycle within PaN RIs. Importantly, CoreTrustSeal+FAIRenabling also retains explicit links back to both the RDA FAIR Data Maturity Model and the FAIR Principles. These links were vital, both for ensuring that the questions we asked in the FAIR evaluation related explicitly to the FAIR Principles and in bringing necessary objectivity to our question development and selection process.

The ExPaNDS FAIR self-assessment exercise

During the three-month period July – September 2022, ExPaNDS partner PaN RIs each undertook a FAIR self-assessment, using a specially designed questionnaire and reporting template. The self-assessment was led in each facility by a volunteer 'facility coordinator', whose job it was to gather responses as necessary from multiple colleagues across their facility. Two workshops and bespoke one to one support offered additional opportunities for engagement with these facility coordinators and other colleagues in ExPaNDS over the course of the exercise. Each facility provided their responses to the questionnaire using the same reporting template. As the FAIR evaluation was intended to be open in nature, we publish these reports in the present deliverable (i.e. in Appendix B) as a key output of the FAIR evaluation task. Furthermore, though not initially envisaged, the self-evaluations are also supplemented with some high-level observations based on the questionnaires on the state of the FAIR journey across the ExPaNDS facilities.

Takeaways and future potential

Overall, the ExPaNDS partners were very positive about their experience of undertaking the FAIR self-evaluation exercise. They found it valuable and useful, both for understanding their facilities' current levels of FAIRness and in terms of articulating clearly what implementations are in progress or planned to support FAIR at their facilities in future. Indeed, several ExPaNDS partners could see real benefit in repeating the FAIR self-assessment on an annual basis for internal use at their facilities.

Specific feedback on the evaluation approach and questions asked served to highlight what worked well (e.g. such as achieving a good balance between usefulness and value of the exercise versus the time and resource needed to undertake the exercise) and also to suggest areas where improvements could be made (e.g. to particular questions and question format).

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

More broadly, it is the hope of ExPaNDS that our experience with this FAIR self-evaluation will make a significant contribution to the still developing area of FAIR evaluation and will help to address gaps that currently exist, especially around the detailed evaluation of workflows and experimental lifecycles where machines and automated processes play a major role. The concept of FAIR may seem simple, but its implementation in practice in such a context poses many challenges. There seems no point in undertaking a FAIR evaluation if it cannot be useful in some practical way. But, the end product of a FAIR evaluation is directly tied to that evaluation's usefulness. Certainly, when it comes to integrating FAIR into workflows and processes, details matter as does considered reflection about what might need to be changed. If this combination of actionable detail and useful insight are important goals (as they were for the ExPaNDS FAIR self-assessment) then the end product of the FAIR evaluation must reflect this.

At the time of writing the ExPaNDS proposal, it was envisioned that the goal of the FAIR evaluation task would be some kind of certification. However, in general, certification assesses the end result (i.e. presumably rewarding it with a 'certificate'); for example, in the context of FAIR, 'certification' might well mean achieving a pre-defined level of acceptable FAIRness. However, the overall purpose of the ExPaNDS FAIR self-assessment was not to judge or compare against some predetermined level, but rather, to provide a formal and systematic way for partner PaN RIs to gain insight about how they do and could implement the FAIR Principles within their facility workflows and processes. For some facilities, especially those still at an initial stage in their FAIR journey, the exercise provided a baseline from which to evaluate in future; for others, further along in their FAIR journey, the exercise served as a way to assess progress. In all cases, the self-evaluation offered an opportunity for systematic and critical self-examination. When approached, as by the ExPaNDS partners, with an attitude of openness and honest engagement, this reflective form of FAIR evaluation has real practical value, revealing useful, detailed information about current levels of FAIRness and highlighting possibilities for future development.

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.

1. Introduction and Context

The ExPaNDS project[1] is motivated by the desire to extend and deepen the application of FAIR principles within the Photon and Neutron (PaN) science community. The acronym 'FAIR' is sometimes used rather loosely as a standalone term (as in 'progress towards FAIR'), and this can be acceptable, though strictly it should be an adjective just like its constituent words. As initially conceived, the fundamental FAIR attributes (Findable, Accessible, Interoperable, Reusable) relate to datasets or collections of data. By extension they may be applied to services utilised in the research domain — thus the European Open Science Cloud (EOSC) aims to develop a 'Web of FAIR Data and services' for science in Europe.[2] In any case, FAIR qualities relating to outputs or tools of research are not achieved accidentally. The datasets acquire those attributes from the processes they undergo throughout their lifecycle: processes which will be more or less adapted to the generation of FAIR. In the case of large-scale scientific infrastructures such as PaN Research Infrastructures (RIs), data handling processes have developed over many years to suit the particular needs and behaviours of the facilities and their users, and there might be considerable variation between facilities in their practices.

Since the PaN RIs in ExPaNDS share a desire to open their data according to the FAIR Principles[3] (and specifically within the context of EOSC), it makes sense to ask them to examine their ways of working from the perspective of FAIR and to conduct a self-evaluation with a view to understanding more clearly where they are doing well, where there is potential for improvement, and to provide a baseline for comparison and learning from the practices of other facilities. This opening up of experience affords the opportunity to enhance the PaN RIs' capabilities to enable FAIR data through adaptation of their processes and workflows.

Such a self-evaluation need not start from scratch, as plenty of existing approaches to FAIR assessment have emerged in recent years. The ExPaNDS work plan envisaged a task to "Consider the use of emerging certification schemes for FAIR data within the Photon and Neutron community and assess the readiness of national RIs [with respect to] these schemes."[4] As explained in Chapter 2 of the present report, the existing FAIR evaluation frameworks apply at different levels: the individual dataset, the repository or the organisation. This breadth is a strength, but also necessitates work to adapt or profile the frameworks to the PaN context, with emphasis on the processes and workflows of the experimental lifecycle and the flow of data through them. In fact, this exercise is itself a valuable outcome of the work, leading to an understanding of how these generic frameworks of evaluation may be selected, combined, and adapted in the context of a particular scientific domain.

[1] ExPaNDS partners 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), UK Research and Innovation (UKRI), Science and Technology Facilities Council (STFC), ISIS Neutron and Muon Source, and European Grid Infrastructure Foundation (EGI). ExPaNDS (2020). Partners. https://expands.eu/partners/

[2] European Commision (n.d.). European Open Science Cloud (EOSC). https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc_en

[3] Wilkinson, M. D. et al. (2015). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data*, **3**:1. https://doi.org/10.1038/sdata.2016.18

[4] ExPaNDS (2018). ExPaNDS European Open Science Cloud Photon and Neutron Data Services [proposal].

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

It might be tempting to regard this self-evaluation as a kind of Key Performance Indicator (KPI) of the ExPaNDS project as a whole. After all, if the motivating goal is the application of FAIR Principles in the particular domain, and the work of the project is oriented towards that (albeit tempered by the emphasis on alignment with EOSC), surely the results of the evaluation should give some indication of whether the goal has been achieved? That, however, is too simple a viewpoint. As stated above, the value is expected to be for the PaN RIs themselves, giving hints for development of their policies, processes, and information infrastructure. These changes take time, and the work of ExPaNDS has been providing perspectives and tools of relevance, but it was never expected that any facility could become fully 'FAIR' within the lifetime of the project. The orientation towards EOSC, especially of the work on data catalogue services and data analysis services, though linked to FAIR data through, for example, enhancing the reuse of workflows, has a somewhat different orientation.

Nonetheless, there has been a strong cross-fertilisation between the FAIR self-evaluation and other work in the project. The development of the FAIR self-evaluation template was influenced by several other outcomes of the project, most notably the ExPaNDS FAIR metadata framework,[5] which is explicitly referenced in the template. Persistent Identifiers (PIDs) and related metadata are fundamental to FAIR, and the project's work in this area provides guidance on PID use and extension for PaN RIs.[6]

The present deliverable reports on the planning and conduct of the FAIR self-evaluation, and comprises six chapters:

- **Chapter One** (current section) introduces the aims and purpose of the ExPaNDS FAIR self-assessment exercise and explores links with previous work done in the ExPaNDS project.
- **Chapter Two** considers the foundations of FAIR evaluation and reviews selected examples of existing FAIR assessment frameworks, examining how these support the evaluation of FAIR at the level of the dataset, repository, and organisation.
- **Chapter Three** outlines the development and design of the ExPaNDS FAIR self-assessment questionnaire template and explains how we engaged with ExPaNDS partners.
- **Chapter Four** presents the ExPaNDS FAIR self-assessment questionnaire with its introduction.
- **Chapter Five** captures the overall impressions and outcomes of the FAIR self-assessment exercise, exploring what worked well and drawing together suggestions for changes and improvements.
- **Chapter Six** offers some general observations about the advancement of FAIR implementation across the ExPaNDS partners, based on the content of their self-assessment questionnaires.
- **Chapter Seven** concludes the deliverable, reflecting on the benefits of the FAIR self-assessment exercise for ExPaNDS partner RIs and what role this type of evaluation might play in future for PaN RIs.

[5] Soler, N., McBirnie, A., Gonzalez-Beltran, A. et al. (2022). ExPaNDS D2.7: Final recommendations for FAIR Photon and Neutron Data Management. https://doi.org/10.5281/zenodo.6799105
[6] Bunakov, V., Krahl, R., Matthews, B. et al. (2022). ExPaNDS D2.5: Advanced infrastructure for PIDs in Photon and Neutron RIs. https://doi.org/10.5281/zenodo.5905350

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

- **Appendix A** provides a detailed description of the process of development of the questionnaire.
- **Appendix B** reproduces the completed self-assessment reports from all participating facilities.

It should be emphasised that it is not necessary by any means to read this deliverable in its entirety. It is necessarily written with a range of audiences in mind, and, therefore, content in the different sections of the deliverable may be more or less relevant for certain audiences. As such, we offer the following guidance on 'how to read the deliverable', based on typical interests.

The ExPaNDS communications plan identifies a number of target groups for external communication and dissemination. Without implying that there are any for whom this report will be of no interest, there are clearly some for whom it will have specific relevance, and these can be grouped as follows:

1. *Community of PaN research infrastructures at the national and European level / PaN user community world-wide*

> **Guidance for reading:**
>
> - Chapter One for aims and purpose in PaN context
> - Chapter Four for the self-assessment questionnaire and reporting template
> - Chapters Five and Six for reflections on the exercise itself and general observations on the state of FAIR in the ExPaNDS RIs
> - Chapter Seven for final conclusions
> - As many of the facilities' questionnaires in Appendix B as of interest from a PaN point of view.

2. *European e-infrastructures / RIs managers, bodies, staff and IT professionals / Other EOSC cluster Projects / Other EOSC-related projects* to which can be added: *FAIR data advocates, implementers and policy maks (not necessarily in the PaN domain and so not included in the ExPaNDS target groups)*

> **Guidance for reading:**
>
> - Chapter One for aims and explanation of PaN context
> - Chapter Two for discussion of the landscape of FAIR evaluation and why the particular approach was chosen
> - Chapter Three for the process of development of the questionnaire and reporting template (and optionally Appendix A)
> - Chapter Five for reflections and conclusions based on the participants' experience of FAIR evaluation
> - Chapter Seven for final conclusions
> - Appendix A if desired for additional detail on the questionnaire development process.

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

2. Existing FAIR Evaluation Frameworks

This chapter considers the foundations of FAIR evaluation and reviews selected examples of existing tools and frameworks that support FAIR assessment at the level of the dataset, the repository, and the organisation respectively. The chapter concludes by drawing out the commonalities found across these frameworks and highlighting key gaps that remain in relation to undertaking meaningful and useful FAIR evaluation in the context of PaN RIs.

2.1 Foundations

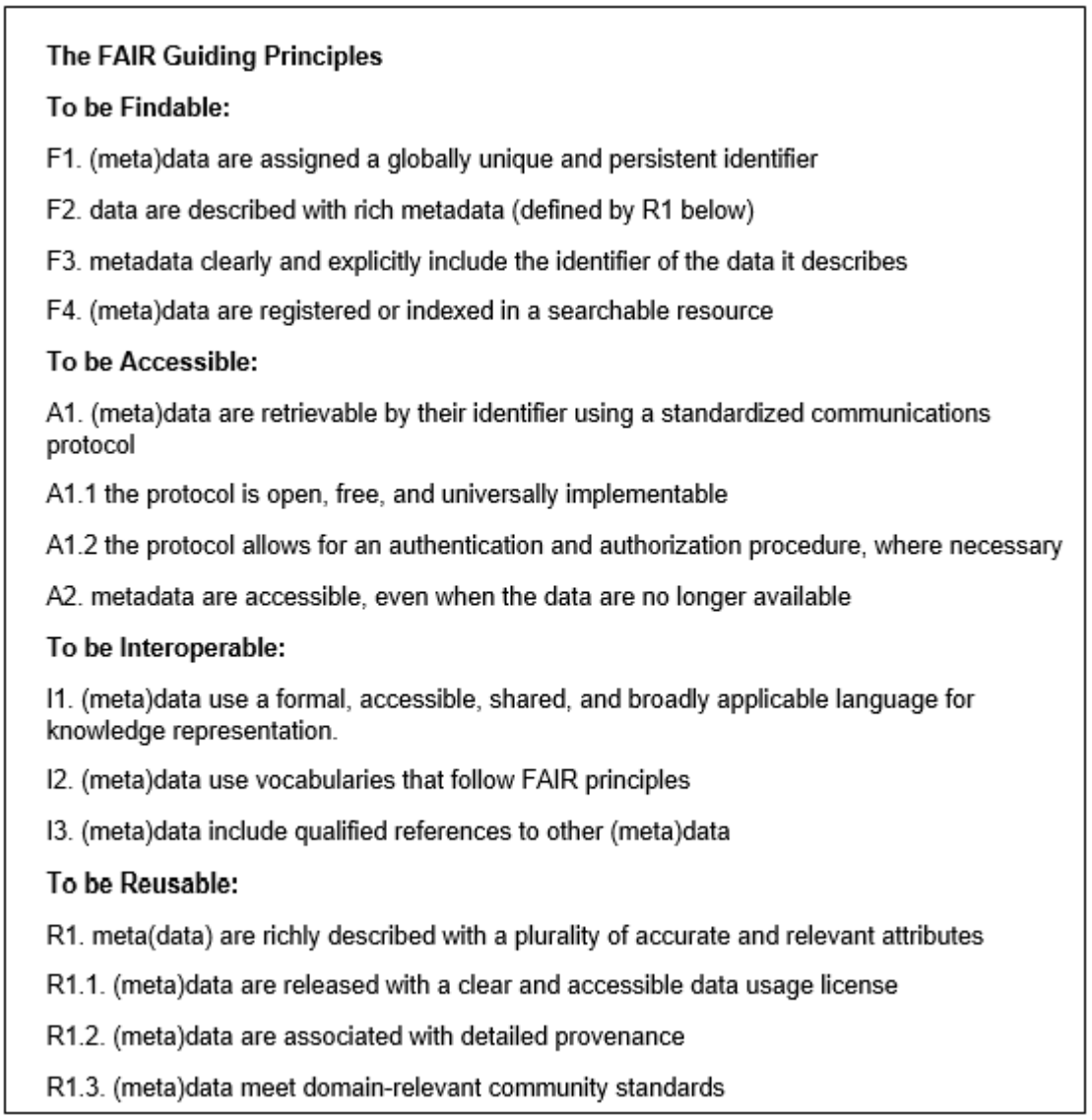
The ExPaNDS FAIR self-assessment for PaN RIs relies on two well-established models for its foundations: The FAIR Principles and the Research Data Alliance (RDA) FAIR Data Maturity Model.

2.1.1 FAIR Principles

Proposed in 2016,[7] the FAIR Guiding Principles for Scientific Data Management and Stewardship (aka the FAIR Principles) stand as fundamental to any FAIR assessment. The FAIR Principles (see Figure 1 below) outline fifteen related yet separate minimal requirements that data and/or their metadata must meet to be considered FAIR for both humans and machines. Since their inception, these fifteen requirements have come to define understanding of what is meant by FAIR data in the research context and have heavily underpinned policy and practice aimed at supporting and achieving FAIR data.

[7] Wilkinson, M. D. et al. (2015). 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.



The FAIR Guiding Principles

To be Findable:

F1. (meta)data are assigned a globally unique and persistent identifier

F2. data are described with rich metadata (defined by R1 below)

F3. metadata clearly and explicitly include the identifier of the data it describes

F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

A1. (meta)data are retrievable by their identifier using a standardized communications protocol

A1.1 the protocol is open, free, and universally implementable

A1.2 the protocol allows for an authentication and authorization procedure, where necessary

A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (meta)data use vocabularies that follow FAIR principles

I3. (meta)data include qualified references to other (meta)data

To be Reusable:

R1. meta(data) are richly described with a plurality of accurate and relevant attributes

R1.1. (meta)data are released with a clear and accessible data usage license

R1.2. (meta)data are associated with detailed provenance

R1.3. (meta)data meet domain-relevant community standards

Figure 1: The FAIR Guiding Principles for Scientific Data Management and Stewardship[8]

2.1.2 RDA FAIR Data Maturity Model

The FAIR Guiding Principles are just that — guiding principles. They inherently retain ambiguity and do not purport to set out rigid rules. As with any higher level guidance, this leaves interpretation of the FAIR Principles open to variety, depending on the context in which they are being applied and the interpreter. The RDA FAIR Data Maturity Model Specification and Guidelines (most recent version June 2020) seeks to reduce this diversity in interpretation, especially in the context of measurement and FAIR assessment, by providing "…a common set of core assessment criteria for FAIRness…".[9]

The RDA FAIR Data Maturity Model framework consists of three elements:

1. "Indicators, i.e. the individual aspects of FAIRness that are evaluated

[8] Ibid.

[9] RDA FAIR Data Maturity Model Working Group (2020). FAIR Data Maturity Model: specifications and guidelines. https://doi.org/10.15497/RDA00050

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

2. Priorities, i.e. the relative importance of the indicators
3. Evaluation method, i.e. the way that the results of the evaluation of the indicators can be given a value."[10]

The indicators derive directly from the FAIR Principles and set out measurable aspects of each. For example, the indicators RDA-A1.1-01M and RDA-A1.1-01D relate back to the principle of metadata and data (respectively) being accessible through a free access protocol, i.e. FAIR Principle A1.1.

In an assessment context, some indicators are likely to be deemed more important than others, and the priority level of an indicator attempts to capture this idea. In the case of the two example indicators above, RDA-A1.1-01M is assigned the top priority level, 'essential', because it relates to metadata (i.e. which is key to all FAIR data), whereas RDA-A1.1-01D is classed one priority level lower, as 'important'.

Drawing on both the indicator description and the priority level, the evaluation method component then aims to suggest ways for evaluating the indicator in practice. In particular, the evaluation methods aim to accommodate two different perspectives:

1. "**Measuring progress:** in this perspective, the emphasis lies on delivering a measure of the extent to which a resource under evaluation meets the requirements expressed in an indicator, giving an indication of which steps may be taken to achieve full satisfaction of an indicator.
2. **Measuring pass-or-fail:** in this perspective, the emphasis lies on determining whether a resource under evaluation meets the requirement of an indicator on a binary, pass-or-fail scale, providing a measure of how a resource under evaluation performs in reaching a particular target level of FAIRness."[11]

Despite seeking to reduce ambiguity and widely-divergent interpretation, the RDA FAIR Data Maturity Model still recognises the need to retain the flexibility inherent in the FAIR Principles. To this end the RDA framework, with its three components, should be considered descriptive rather than prescriptive. In other words, the indicators, priority levels, and evaluation methods serve to 'fill out' some of the descriptive details that can be seen as missing from the higher level guidance that makes up the FAIR Principles. Such descriptive detail can prove immensely valuable in an evaluation context because it allows for clearer articulation and specification of what exactly is being assessed.

2.2 A Variety of Approaches to FAIR Assessment

It can be seen from preceding discussion that the basic FAIR principles are not directly capable of being used for assessment of degree of FAIRness. Even the indicators of the RDA FAIR Data Maturity Model require interpretation and adaptation to particular domains. In consequence, and driven by the clamour of interest in FAIR, there has been a proliferation of approaches to FAIR assessment focussing on different aspects and in some cases particular types of resource. The approach taken in ExPaNDS has been to select a small number of frameworks (F-UJI, CoreTrustSeal, and ACME-FAIR) because of their prima facie relevance, to analyse their strengths and weaknesses, and to build on what they offer. However, acknowledging the existence of other approaches, the final subsection (see

[10] Ibid.
[11] Ibid.

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

Section 2.2.4) makes a brief survey of other approaches and explains why they are less relevant for our FAIR self-evaluation aims within ExPaNDS.

In any assessment context, it is important to establish the level at which the evaluation is taking place. Beyond the more generic FAIR Principles and RDA FAIR Data Maturity Model, existing FAIR assessment frameworks evaluate FAIRness specifically at the level of the dataset, repository, or organisation. Below, we consider selected examples that cover each of these levels in turn. As explained above, we devote the bulk of the discussion (see Sections 2.2.1 – 2.2.3) to those evaluation frameworks that had the greatest influence on our thinking around and development of the methodological approach that underpinned the ExPaNDS FAIR self-evaluation exercise.

2.2.1 F-UJI

The F-UJI tool,[12,13] developed in the FAIRsFAIR project,[14] is an online service that allows for the automated FAIR assessment of datasets. Each dataset is evaluated against a set of sixteen FAIR Object Assessment Metrics proposed by FAIRsFAIR.[15] These FAIRsFAIR metrics draw heavily on the FAIR Principles and the RDA FAIR Data Maturity Model.

To use the F-UJI tool, the user supplies an identifier (and, optionally, a metadata service, e.g. OAI-PMH) for the dataset to be evaluated. The tool then programmatically assesses the FAIRness of the dataset, producing an output that comprises three parts:

1. An **Evaluated Resource** section, which gives descriptive information about the dataset and states the overall assessed level of FAIRness (i.e. 0=incomplete, 1=initial, 2=moderate, 3=advanced)
2. A **Summary** section, which illustrates the FAIRness level as a percentage and also breaks down the scores achieved according to the four components of FAIR
3. A **Report** section, which includes details of the tests used and the dataset's results for each of the sixteen FAIRsFAIR assessment metrics.[16]

While relatively easy to use, the F-UJI tool does have known limitations. In particular, the automated assessment approach depends on clear, unambiguous interpretation being possible by a machine. Where the FAIR Principles are concerned, this is not always possible: even with the help of more descriptive models such as the RDA FAIR Data Maturity Indicators and the FAIRsFAIR Object Assessment Metrics, some criteria (e.g. 'rich', 'plurality', 'relevant') still require human judgement and interpretation.[17] On a more practical note, the ability to only evaluate one dataset at a time through the web-based interface is not especially efficient, although a REST service is available for testing against a larger number of datasets.[18]

---

[12] Devaraju, A. and Huber, R. (2020). F-UJI - An Automated FAIR Data Assessment Tool (v1.0.0). https://doi.org/10.5281/zenodo.4063720
[13] F-UJI: Automated FAIR Data Assessment Tool (n.d.). https://www.f-uji.net/
[14] FAIRsFAIR "Fostering FAIR Data Practices In Europe" project funded by the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 (grant agreement 831558). See https://www.fairsfair.eu/
[15] See https://www.fairsfair.eu/fairsfair-data-object-assessment-metrics-request-comments and https://docs.google.com/document/d/1ymkzVmF_BJmKTQZO0SRQ1YQJaPxeflJZ84AKUJUlGeM
[16] F-UJI: Automated FAIR Data Assessment Tool (n.d.). https://www.f-uji.net/
[17] F-UJI: Automated FAIR Data Assessment Tool (n.d.). About. https://www.f-uji.net/index.php?action=about
[18] See https://github.com/pangaea-data-publisher/fuji .

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

It is important to emphasise that F-UJI is still under development and continues to evolve. Its current strength lies in its capability to assess a single dataset in detail against defined criteria using specific tests, where the intention is to follow up the tool's automated assessment with a review by a human evaluator who wants to know where/how to make changes to improve the FAIRness of the dataset. In this sense, the tool adopts an evaluation approach resembling the 'measuring progress' evaluation method, i.e. as described by the RDA FAIR Data Maturity Model framework (see Section 2.1.2).

2.2.2 CoreTrustSeal and CoreTrustSeal+FAIRenabling

An assessment-related strand of work in the FAIRsFAIR project sought to align CoreTrustSeal,[19] an established certification approach for trustworthy repositories, with "repository characteristics that enable FAIR data".[20] The resulting CoreTrustSeal+FAIRenabling framework[21] provides an example of FAIR assessment aimed at the repository level.

CoreTrustSeal certification is awarded to repositories achieving an appropriate level of compliance with the sixteen CoreTrustSeal Trustworthy Data Repositories Requirements[22] as assessed through objective, independent review. The sixteen CoreTrustSeal requirements are collected under three themes:

1. **Organisational Infrastructure** (includes, for example, scope and mission, licenses, engagement with external guidance, governance, etc.)
2. **Digital Object Management** (incorporates aspects such as storage, data integrity, metadata, etc.)
3. **Technology** (covers technical infrastructure and security)

The evaluation process for CoreTrustSeal involves the applicant (i.e. a repository) providing evidence to show the reviewers how they meet each of the requirements. In addition, as an indicator of self-assessed progress, applicants must state what they consider to be their compliance level for each of the requirements. There are five compliance level options:

0 – Not applicable

1 – The repository has not considered this yet

2 – The repository has a theoretical concept

3 – The repository is in the implementation phase

4 – The guideline has been fully implemented in the repository

Reviewers, in turn, judge compliance on the basis of the evidence submitted by the applicant. Thus, the onus is on applicants to reflect realistically on their level of compliance and to ensure they can back up their self-assessed choice of compliance level with adequate and acceptable evidence. The judged compliance level also directly impacts the success (or not) of the CoreTrustSeal application. For example, level 1 and 2 compliance are not

[19] CoreTrustSeal Standards and Certification Board. (2019). CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022 (v02.00-2020-2022). https://doi.org/10.5281/zenodo.3638211
[20] L'Hours, H., von Stein, I., deVries, J. et al. (2021). M4.3 CoreTrustSeal+FAIRenabling, Capability and Maturity (1.0). https://doi.org/10.5281/zenodo.5346822
[21] Ibid.
[22] CoreTrustSeal Standards and Certification Board. (2019). CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022 (v02.00-2020-2022). https://doi.org/10.5281/zenodo.3638211

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

considered sufficient for certification, although reviewers may approve certification if some requirements are currently in the implementation phase (i.e. at level 3 compliance). Any requirements that the applicant deems not applicable (i.e. level 0) must be clearly justified as such.

The CoreTrustSeal+FAIRenabling framework adds FAIR assessment elements to seven of the sixteen CoreTrustSeal requirements. These seven CoreTrustSeal requirements are:

- R2: Licenses
- R7: Data integrity and authenticity
- R10: Preservation plan
- R13: Data discovery and identification
- R14: ReUse
- R15: Technical infrastructure
- R16: Security

CoreTrustSeal+FAIRenabling links these seven requirements back to the FAIR Principles, as illustrated in Figure 2 below.

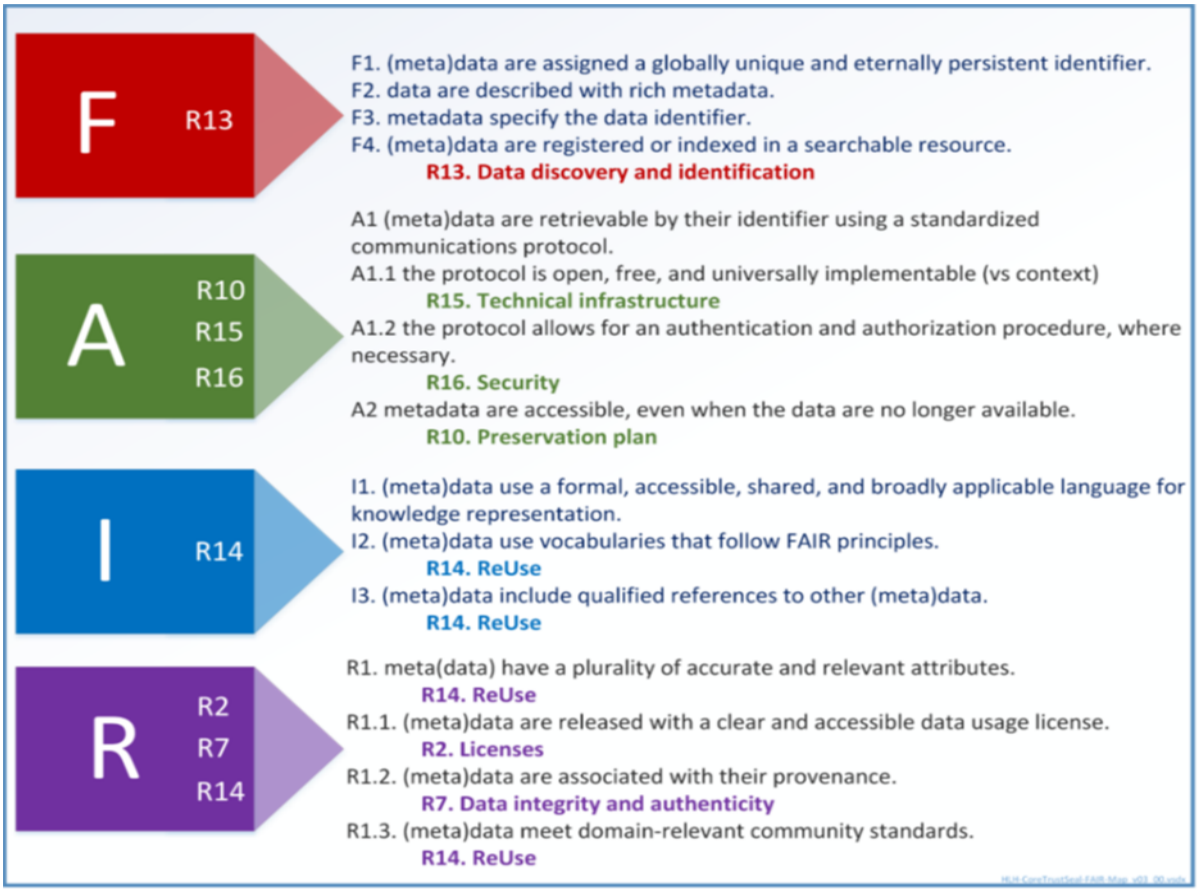
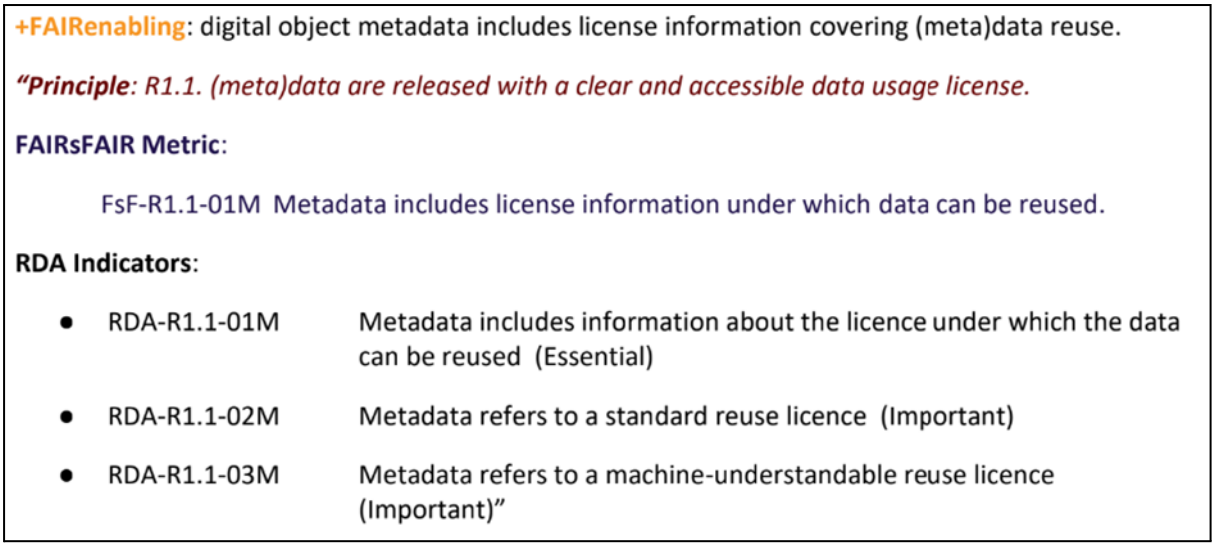


Figure 2: CoreTrustSeal to FAIR Alignment as illustrated in the CoreTrustSeal+FAIRenabling framework (v00.04)[23]

[23] L'Hours, H., von Stein, I., deVries, J. et al. (2021). M4.3 CoreTrustSeal+FAIRenabling, Capability and Maturity (1.0). https://doi.org/10.5281/zenodo.5346822

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

More specifically, under each of the relevant seven CoreTrustSeal requirements, CoreTrustSeal+FAIRenabling includes a +FAIRenabling element. This +FAIRenabling element is further clarified/specified by the inclusion of relevant FAIRsFAIR Object Assessment Metrics and RDA FAIR Maturity Indicators. Figure 3 below provides an example of this approach, taken directly from the CoreTrustSeal+FAIRenabling guidelines. The example relates to CoreTrustSeal requirement R2: Licenses.


+FAIRenabling: digital object metadata includes license information covering (meta)data reuse.

*"**Principle**: R1.1. (meta)data are released with a clear and accessible data usage license.*

FAIRsFAIR Metric:

FsF-R1.1-01M Metadata includes license information under which data can be reused.

RDA Indicators:

- RDA-R1.1-01M Metadata includes information about the licence under which the data can be reused (Essential)
- RDA-R1.1-02M Metadata refers to a standard reuse licence (Important)
- RDA-R1.1-03M Metadata refers to a machine-understandable reuse licence (Important)"

Figure 3: Example of +FAIRenabling element (in orange), including relevant FAIR Principle(s) (in red), FAIRsFAIR Object Assessment Metric(s) (in blue), and RDA Maturity Model Indicator(s) (in black), added to the CoreTrustSeal requirement R2: Licenses (not shown)[24]

As with the standard CoreTrustSeal, CoreTrustSeal+FAIRenabling also includes a self-assessment aspect in relation to compliance with the +FAIRenabling components. Formally, five 'Capability Maturity Model Indicators' are defined, based on the well-established Capability Maturity Model Integration (CMMI) framework.[25] As an appraisal tool, the CMMI, which stems from outside of the FAIR ecosystem, is designed to guide process improvement and/or assess process maturity. CoreTrustSeal+FAIRenabling identifies three Capability Maturity Model Indicators as particularly useful in the context of an initial FAIR self-assessment:

1. "**Initial:** Aware of the scope and issue within the area of focus. Lists of all items relevant to the area of focus exist.
2. **Managed**: Processes, procedures and other implementation measures are in place for all items on the lists
3. **Defined**: Managed areas of focus are further integrated into the wider organisational policy and practice."[26]

Work undertaken by FAIRsFAIR suggests that achieving level 2 'Managed' or level 3 'Defined' is generally sufficient for the +FAIRenabling elements, although, where possible, reaching higher levels of compliance (e.g. level 4 'Quantitatively managed' or level 5

[24] Ibid.
[25] See https://en.wikipedia.org/wiki/Capability_Maturity_Model_Integration .
[26] L'Hours, H., von Stein, I., deVries, J. et al. (2021). M4.3 CoreTrustSeal+FAIRenabling, Capability and Maturity (1.0). https://doi.org/10.5281/zenodo.5346822

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

'Optimising' — neither included in list above) would be desirable.[27] Thus, as with the F-UJI tool, CoreTrustSeal+FAIRenabling leans towards the 'measuring progress' evaluation approach, although it also incorporates a low bar, basic 'pass/fail' approach in that level 1 'Initial' is considered insufficient to show compliance with the +FAIRenabling elements, and therefore (one assumes), with the associated FAIR Principles (see Section 2.1.1).

Unlike CoreTrustSeal, which is an established certification (although the sixteen requirements are themselves formally reviewed every three years),[28] CoreTrustSeal+FAIRenabling has not been approved as a formal certification process, although it is possible that it may achieve such status in future, depending on the needs of the research and repository communities. Nonetheless, even in its present form, the CoreTrustSeal+FAIRenabling framework offers a useful and considered starting point for FAIR assessment at the repository level.

2.2.3 ACME-FAIR

Aimed at Research Performing Organisations (RPOs), Assessing Capability Maturity and Engagement with FAIR Enabling Practice (ACME-FAIR) is a guide developed within the FAIRsFAIR project.[29] Released in the first half of 2022, ACME-FAIR is currently in draft form.[30] The guide has already undergone some revision following a public consultation and review by the University of Helsinki; however, it remains open for further revisions, including a planned review by the Digital Curation Centre (DCC).[31] The main purpose of ACME-FAIR is "… to help managers of Research Data Management and related professional services to self-assess how they are enabling researchers, and the professional staff who support them, to put the FAIR data principles into practice…".[32]

The ACME-FAIR guide takes the form of seven documents, each covering an 'issue' deemed important for enabling FAIR data in RPOs:

- Issue #1: Defining the policy environment
- Issue #2: Developing sustainable business models
- Issue #3: Professionalising roles through training, mentoring, and recognition
- Issue #4: Supporting data management planning
- Issue #5: Defining data interoperability frameworks
- Issue #6: Selecting data, services, and repositories for FAIR
- Issue #7: Ensuring trustworthy curation[33]

[27] Ibid.
[28] CoreTrustSeal Standards and Certification Board. (2019). CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022 (v02.00-2020-2022). https://doi.org/10.5281/zenodo.3638211
[29] FAIRsFAIR (2022). ACME-FAIR: a guide for Research Performing Organisations (RPO). https://www.fairsfair.eu/acme-fair-guide-rpo
[30] FAIRsFAIR (2022). Assessing capability maturity and engagement with FAIR-enabling practice. https://zenodo.org/communities/acme-fair
[31] FAIRsFAIR (2022). ACME-FAIR: a guide for Research Performing Organisations (RPO). https://www.fairsfair.eu/acme-fair-guide-rpo
[32] FAIRsFAIR (2022). Assessing capability maturity and engagement with FAIR-enabling practice. https://zenodo.org/communities/acme-fair
[33] FAIRsFAIR (2022). ACME-FAIR: a guide for Research Performing Organisations (RPO). https://www.fairsfair.eu/acme-fair-guide-rpo

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

Each of these seven 'issue' guides incorporates "… a thematic introduction, an overview of the relevant capabilities [i.e. related to that issue], and a rubric for assessing the levels of maturity and community engagement for each capability."[34]

The ACME-FAIR assessment rubric, as with the rubric used by CoreTrustSeal+FAIRenabling, follows the CMMI framework (see Section 2.2.2). In this sense, ACME-FAIR, like CoreTrustSeal+FAIRenabling, places most of its emphasis on the 'measuring progress' evaluation approach, while still retaining some basic elements of a 'pass/fail' approach. Significantly, the CMMI framework can be applied at local levels (e.g. project, department) and/or at the level of the organisation, hence the reason that different FAIR assessment tools can use it at differing levels, i.e. CoreTrustSeal+FAIRenabling at the repository level and ACME-FAIR at the organisational level. It is important to note, however, that whereas in CoreTrustSeal+FAIRenabling, the compliance level (i.e. based on the CMMI framework) is self-assessed but then must be justified to reviewers, in the ACME-FAIR model, the indication of the level of compliance (i.e. also based on CMMI) stems entirely and only from self-assessment – there is no external review process.

More broadly, the ACME-FAIR guide is based on two institutional level assessment tools:[35]

1. the DCC's **Research Infrastructure Self-Evaluation (RISE) framework:** Aimed primarily at Higher Education institutions, the RISE framework "... is a benchmarking tool designed to facilitate RDM [Research Data Management] service planning and development at the institutional level";[36] and,

2. the Dutch National Coordination Point Research Data (LCRDM) task group's **Do I-PASS for FAIR self-assessment tool:** Intended for use by research institutes and universities, Do I-PASS for FAIR enables organisations, through their answers to a set of questions and their own evaluation of performance levels, to self-assess their level of FAIRness.[37]

ACME-FAIR also actively seeks to align with the recommendations presented in the Turning FAIR into Reality (TFiR) report and to complement the issues addressed in Science Europe's Guide to Sustainable Research Data.[38] However, unlike both the F-UJI tool for dataset assessment and the CoreTrustSeal+FAIRenabling approach for repository assessment, the ACME-FAIR organisational assessment framework incorporates neither the RDA FAIR Maturity Model Indicators nor the FAIRsFAIR Object Assessment Metrics. As well, although it makes general reference to the FAIR Principles — for example, through

[34] Ibid.
[35] Davidson, J., Whyte, A., Molloy, L. et al. (2022). Defining the Policy Environment: ACME-FAIR Issue #1 (2.0). https://doi.org/10.5281/zenodo.6345332
[36] Rans, J and Whyte, A. (2017). 'Using RISE, the Research Infrastructure Self-Evaluation Framework' v.1.1 Digital Curation Centre:www.dcc.ac.uk/guidance/how-guide
[37] de Bruin, T., Sarah Coombs, S., de Jong, J. et al. (2020). Do I-PASS for FAIR. A self assessment tool to measure the FAIR-ness of an organization (Version 1). https://doi.org/10.5281/zenodo.4080867
[38] Note that this Science Europe guidance has a remit that is wider than FAIR but there is, nonetheless, considerable implicit cross-over with FAIR. In particular, the "…guidance is designed to support RPOs [Research Performing Organisations], RFOs [Research Funding Organisations], and RDIs [Research Data Infrastructures] in developing their agenda for research data to achieve sustainable data sharing and interoperable systems. It takes the form of three complementary maturity matrices to allow collaboration with other organisations. The matrices present a framework and propose actions in six essential areas: Organisational engagement and commitment, Policy environment, Financial aspects, Training, Technical preparedness [and] Communication and awareness raising."[bullet points in original removed here] Science Europe (2021). Practical Guide to Sustainable Research Data. https://doi.10.5281/zenodo.4769703

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

mentions in capability descriptions such as 'Aligning policy on DMPs with FAIR Principles' (i.e. Maturity Capability #1 in the ACME Rubric: Supporting Data Management Planning)[39] — ACME-FAIR does not explicitly link assessment back to specific FAIR Principles (e.g. such as Principle R1.2, which has relevance for data management planning, i.e. the capability example provided above).

2.2.4 Other approaches to FAIR assessment

There are several services to perform automated assessment of online resources (such as datasets with a suitable identifier). These are instances of evaluation at the level of the dataset, and comparable with the F-UJI tool described above.

- The EOSC-synergy project[40] produced a FAIR Framework for validating EOSC FAIR data requirements, and in particular a tool called FAIR EVA[41] to check the FAIRness level of digital objects from different repositories or data portals. It is a web service that checks the metadata and data of a digital object against indicators that are those of the RDA FAIR Data Maturity Model.

- The FAIRsharing resources[42] include FAIR evaluation services to assess the FAIRness of digital resources.[43] Again, this is an online service that accepts an identifier of a resource to be evaluated, and assesses it against maturity indicators that may be defined according to community needs. The authors are emphatic that "... FAIR evaluations are not intended to be used as 'judgement', but rather as a means to objectively (AND TRANSPARENTLY!) [sic] test if a resource has successfully fulfilled the FAIRness requirements that that community has established."[44]

- The FAIR Enough data maturity indicators[45] implement a set of FAIR indicators to enable automated evaluation of a data resource. The indicators are related back to the basic FAIR Principles, interpreted so they are suitable for automated checking.

- FAIR-Checker,[46] developed by the interoperability working group of the French Institute for Bioinformatics, is another tool for assessing how FAIR are web resources. It uses semantic web technologies to check that metadata employ standards and recognised ontologies or controlled vocabularies.

These tools all have their place in FAIR evaluation, but because of their necessarily exclusive focus on FAIRness of individual datasets through automated assessment, they are not of great relevance to the ExPaNDS situation, where the wider context of workflows and processes is important.

---

[39] Whyte, A., Molloy, L, Grootveld, M, and Thorley, M. (2022). Supporting Data Management Planning: ACME-FAIR Issue #4. https://doi.org/10.5281/zenodo.6346747
[40] See https://www.eosc-synergy.eu . EOSC-synergy receives funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857647.
[41] EOSC-synergy (n.d.). FAIR Framework. https://www.eosc-synergy.eu/results/fair-framework/
[42] See https://fairsharing.org/ .
[43] FAIRsharing and FAIR metrics groups (n.d.). FAIR Evaluation Services. https://fairsharing.github.io/FAIR-Evaluator-FrontEnd
[44] Ibid.
[45] Emonet, V. (2022). FAIR Enough data maturity indicators [fair-enough-data]. https://fair-enough.semanticscience.org/collections/fair-enough-data
[46] Rosnet, T., Gaignard, A., Devignes, M. (n.d.) FAIR-Checker. https://fair-checker.france-bioinformatique.fr/

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

Finally, it is worth mentioning that there are some extensions to the range of FAIR Principles and evaluation frameworks going beyond data. An influential example is the FAIR Principles for Research Software (FAIR4RS Principles)[47] that has taken account of the special characteristics of software and their relationship to FAIR. Likewise the web service FOOPS![48] assesses the compliance of vocabularies or ontologies against the FAIR Principles. However, developments of this kind are of no relevance to the ExPaNDS situation.

2.3 Links, commonalities, and gaps

Several common aspects emerge from our review of selected example FAIR assessment frameworks. The review also highlights an important gap in the existing FAIR evaluation landscape.

2.3.1 Linking evaluation to the FAIR Principles

The frameworks reviewed in Section 2.2 illustrate that existing evaluation methods support FAIR assessment at different levels, including the level of the dataset, the repository, and the organisation. Importantly, given that the FAIR Principles act as the foundation for FAIR, the frameworks do all link back in some way to the FAIR Principles.

Despite their shared practice of linking to the FAIR Principles, how specifically (or not) the frameworks articulate these links differs. For example, both F-UJI (i.e. dataset level assessment) and CoreTrustSeal+FAIRenabling (i.e. repository level assessment) employ the RDA FAIR Data Maturity Indicators and/or the FAIRsFAIR Object Assessment Metrics to unambiguously link elements being assessed directly back to specific FAIR Principles. In contrast, the organisational level FAIR evaluation frameworks (e.g. such as ACME-FAIR and Do I-PASS for FAIR) acknowledge the FAIR Principles in the collective sense but make no reference to specific FAIR Principles.

This failure of existing organisational level FAIR evaluation frameworks to link explicitly to specific FAIR Principles results in a key gap in the current FAIR assessment landscape. The reason that this gap is particularly important is that it is precisely the organisational level frameworks that most strongly focus on evaluating practices, workflows, and processes through seeking to guide process improvement and/or appraise process maturity (see Section 2.2.3). Yet, these same frameworks cannot explicitly link their assessment approaches back to the details of the FAIR Principles. This is a crucial missing aspect, given that improvements to processes and workflows depend on information about what details should be included, added, left out, or changed. Certainly in the PaN RI context, where facility workflows and processes are crucial to data generation, storage, and (increasingly) analysis, having FAIR assessment results that include details that link back to specific FAIR Principles would help PaN RIs to be the most effective in terms of making improvements to support FAIR data.

2.3.2 Importance of self-assessment in FAIR evaluation

The review of existing frameworks also suggests a strong leaning towards self-assessment in the context of FAIR evaluation. F-UJI is the only example framework discussed above that

[47] Chue Hong, N., Katz, D., Barker, M. et al. (2022). FAIR Principles for Research Software (FAIR4RS Principles). https://doi.org/10.15497/RDA00068https://doi.org/10.15497/RDA00068
[48] Garijo, D., Corcho, O., and Poveda-Villalón, M. (2021). FOOPS! An Ontology Pitfall Scanner for the FAIR Principles. https://ceur-ws.org/Vol-2980/paper321.pdf

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

does not directly incorporate some method of self-assessment. Nonetheless, despite its programmatic, automated approach, F-UJI, with its tendency towards the 'measuring progress' type of assessment, lends itself well to follow on evaluation by a human. At the higher levels of assessment (e.g. repository, organisation), the role of self-evaluation is explicit: CoreTrustSeal+FAIRenabling incorporates self-assessment in combination with external review, while ACME-FAIR relies entirely on self-evaluation. Particularly at the organisational level, this echoes a wider trend that can be seen beyond FAIR-specific assessment; for example, both the RISE framework and the Science Europe research data sustainability guidance also rely on self-assessment as their evaluation method.

2.3.3 Describing the results of FAIR assessment

The results of assessment can take many forms, from purely quantitative measures to fully narrative reports. Indeed, it is the underlying purpose for the assessment that drives choices around the assessment approach and the results format that is deemed desirable and useful. For example, in educational assessment, an established and well-tested area, 'formative' assessment is undertaken throughout a student's course, whereas 'summative' assessment is used to produce the student's grade at the end of the course.[49] These two types of assessment are also employed as methods used for FAIR evaluation; for example, the RDA FAIR Data Maturity Model's two evaluation approaches — 'measuring progress' and 'pass/fail' — represent, respectively, formative and summative assessment methods.

The review of existing FAIR assessment frameworks indicates a notable preference for the 'measuring progress' approach, i.e. formative assessment. Even tools such as F-UJI, which do provide a type of summative assessment, do so with the overall purpose being to encourage improvement. In this sense, all of the assessment frameworks capture the idea of FAIR as 'a journey', and they seek to support improvement and progress along this journey. Where 'pass/fail' approaches do feature, e.g. such as in CoreTrustSeal+FAIRenabling, they are mostly low bar, meaning a 'fail' (level 1) generally indicates that the FAIR 'journey' has really yet to begin in earnest.

This emphasis on the 'measuring progress' approach is also evident in the measurements, indicators, and rubrics used by the various FAIR evaluation frameworks. Again, only the F-UJI tool employs a fully quantitative indicator — in this case, a percentage score. However, F-UJI also employs indicators that, while expressed as a 'score', are essentially qualitative in nature. Examples include the Capability Maturity Model Indicators such as 'initial' (level 1), 'defined' (level 2), and 'managed' (level 3). These same indicators are used by both CoreTrustSeal+FAIRenabling and ACME-FAIR. However, crucially, in these two frameworks, simply applying an indicator is not enough; both rely heavily on results in narrative format to support the assessment, i.e. the evidence supplied to reviewers in CoreTrustSeal+FAIRenabling and the rubric employed by ACME-FAIR. In other words, while formal indicators are included in results that derive from the use of existing FAIR evaluation frameworks, it is by far the qualitative explanations underpinning the self-assessed choice of compliance indicators that provide the most insight into the results of these FAIR evaluations.

[49] For a brief overview of educational assessment, see https://en.wikipedia.org/wiki/Educational_assessment .

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

3. Development and Design of the ExPaNDs PaN RI FAIR Self-Assessment

This chapter outlines the evaluation approach adopted for the ExPaNDS FAIR self-assessment, considers issues that impacted the design of the questionnaire and reporting template, and describes how we engaged with ExPaNDS partners over the course of the exercise.

3.1 Evaluation Method

This section sets out the key starting points for our FAIR evaluation approach, explores the four principles that underpinned our choice of method, and briefly overviews the question development process. Full details on the latter can be found in Appendix A.

3.1.1 Starting points

We saw in Chapter 2 that a variety of frameworks have already been developed for assessing degree of FAIRness, each with its own emphasis; and that because of these particular emphases, it was not possible to choose and apply one for the FAIR self-assessment exercise of ExPaNDS. Adaptation to the context of PaN RIs was absolutely necessary, where by context we mean the established processes and practices of the PaN facilities, accreted over years and motivated ultimately by the nature of the science and the techniques used. Self-assessment methods that focussed on the individual dataset (without considering the process by which it was created), on repositories (which are only one part of the operations of PaN RIs), or on the organisation as a whole (without taking account of what is and is not under the control of the facilities) would be confusing, difficult to apply and risk missing or distorting important aspects of FAIR in the facilities' operations.

3.1.2 Four underpinning principles

The decision was taken to select, combine, and adapt some of the existing frameworks to tailor an evaluation method especially suited to PaN RIs and capable of passing the test that it could be used for self-assessment by staff of those facilities with only minimal guidance. Indeed the process of developing such a method is regarded by the ExPaNDS team as a significant contribution in its own right to the understanding of FAIR, and a major outcome presented in this deliverable — quite apart from the applications of it by the ten participating facilities.

However, the adaptation of existing frameworks could not just be a fortuitous cherry-picking of what seemed most relevant. There had to be some underlying principles, and these four principles can be summarised as:

1. linking back to the FAIR Principles
2. identifying and taking advantage of what existing FAIR evaluation frameworks have to offer
3. taking account of the relationships between existing FAIR evaluation approaches
4. relating clearly to the processes and practices of PaN RIs

As mentioned in Section 2.3, existing frameworks do not really consider processes and workflows, but these are precisely the areas where PaN RIs can take action and make

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

changes. So, in the PaN context, it is essential to examine these. The existing framework that comes closest to allowing this is CoreTrust+FAIRenabling, which was selected as the starting point for developing the questionnaire, supported by the FAIR Principles and the RDA FAIR Data Maturity Model (i.e. addressing Principle 2 in the list above). Reference to these latter two frameworks ensures the completeness of coverage of the FAIR self-evaluation and its traceability back to the FAIR Principles (i.e. thereby addressing Principle 1 above). Furthermore, CoreTrustSeal+FAIRenabling is already linked to the FAIR Principles (i.e. this addresses Principle 3 in the list above), as can be seen in the extract from the CoreTrustSeal+FAIRenabling framework we presented earlier in Figure 3 (see Section 2.2.2).

Figure 4 below depicts in outline the procedure that was followed.

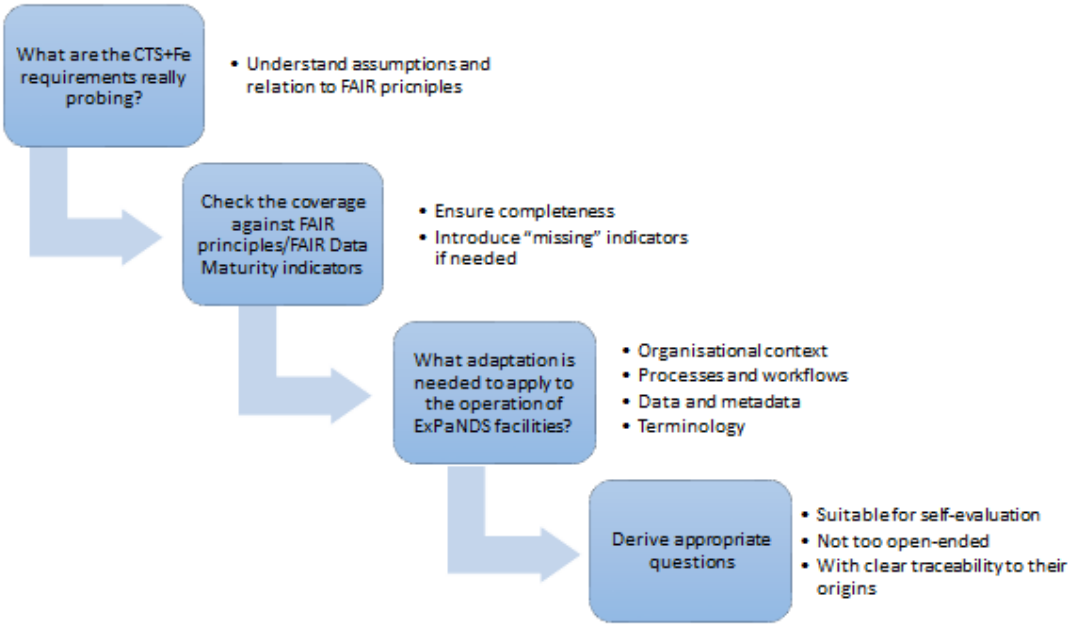


Figure 4: Schematic procedure by which the FAIR self-evaluation framework for ExPaNDS was developed

Some additional remarks can be made on the adaptations that were made in relating to the PaN context (i.e. to ensure that we addressed Principle 4 in the list above). CoreTrustSeal+FAIRenabling focuses on repositories, so our use of it required adaptation to the PaN experimental lifecycle and PaN RIs workflows. At one level this simply meant referring consistently to 'facilities', 'metadata catalogues', 'experimental lifecycle', etc. Going deeper, however, it was necessary to be aware of the distinction between the PaN RI as a whole and the individual instruments (or beamlines); for example, at the level of instruments, NeXus is the general metadata standard (and data format), whereas for the PaN RI as a whole (i.e. its metadata catalogue) it might be Dublin Core. Issues of this type arose constantly during the development of the questionnaire, and it was necessary to take care to retain the intent of the original criterion or indicator while adapting the assumptions and language to the PaN RI context.

3.1.3 Development of the questions

The whole exercise of refining the questions, underpinned by the four principles set out in Section 3.1.2, and especially, Principle 4, required much iteration as understanding

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

developed, but eventually resulted in a consistent, coherent set of questions that were ready for application. For reasons of brevity, we do not take the reader here through the many steps involved in the question development process; however, full details of the steps and how the question development traces back to the FAIR Principles, the RDA Maturity Model, and the CoreTrustSeal+FAIRenabling framework are available in the Appendix. For the 'final version' set of the 29 questions asked in the FAIR self-evaluation, see the copy of the questionnaire and reporting template in Chapter 4.

3.2 Questionnaire Design and Results Reporting

Alongside the foundational matter of the development of the evaluation method and questions (see Section 3.1), there were other factors, acknowledgedly, more practical in nature, but nonetheless, with the potential to contribute to a successful (or not) outcome for the ExPaNDS FAIR self-assessment exercise. These aspects included the design of the questionnaire (i.e. beyond the content of the questions themselves) and how we used and reported the results of the evaluation exercise.

3.2.1 Use of broader topics

During the question development process (see Section 3.1), it became evident that some questions were more related than others. This led to the natural emergence of a useful set of broader, high-level 'topics', under which questions could be grouped. In the end, we arrived at a set of topics that covered seven broad areas of relevance to PAN facility workflows and processes in the context of FAIR:

1. the existence, completeness, and richness of metadata related to experiments at the facility
2. the flexibility and capability of any search functionality/service provided by the facility
3. standardisation as used in research data management processes at the facility
4. the indexing and harvesting of the facility's metadata by machines
5. the use of PIDs by the facility
6. the access to data that the facility provides to human users and machines
7. the facility's long term curation of data

We used these topics to section the questionnaire. Each topic was introduced with a brief explanation of its relevance to FAIR. After each of these introductions, we listed the specific FAIR Principles that the questions grouped under that topic addressed. Thus, it was possible for users of the questionnaire to not only relate the questions to a general aspect of FAIR but also to specific elements of the FAIR Principles. This approach also allowed the mixing of different components of the FAIR Principles (e.g. the 'F' and the 'R') when this seemed the natural thing to do; for example, 'existence, richness, and completeness of metadata' makes sense as a broad topic but it also combines elements of both findability ('F') and reusability ('R'). If, as was our initial inclination, we had instead grouped the questions under the four components of FAIR (i.e. without the use of the broader topics), not only would we have encountered the undesirable problem of repetition of questions, but also we would not have been able to tie the questions being asked to more general understandings of FAIR as well as we could through the use of the broader topics.

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

3.2.2 Other design considerations

The survey tool used and the format of questions were other key design considerations.

The use of Google Docs for the template, i.e. rather than a survey tool such as Google Forms or similar, ensured that the template could be as flexible as possible (e.g. respondents could even add additional options to tick boxes or add additional free text if they wished) and also that multiple colleagues in a facility could work on responses at the same time (i.e. because our expectation was that input from multiple areas of expertise would be needed, given the range of questions). Although, in the end, the option to custom shape aspects of the template was not actually taken up by participants, we felt that, given the self-assessment and reflective nature of the FAIR evaluation, it was important to keep open the option to customise responses.

Another important aspect was what type of question to employ — for example, binary 'yes/no', multiple choice, free text, etc. A careful balance was needed. While on the one hand, we sought to take care not to over burden respondents, i.e. as could happen, for example, if we required too many free text answers,[50] on the other hand, we wanted to ensure that we encouraged sufficiently reflective and thoughtful responses, something that the overuse of 'yes/no' or multiple choice questions could work against. In the end, we took the middle path, making use of 'yes/no' and multiple choice question formats, but including free text boxes after many of these, asking participants to expand on responses and/or add optional detail if they wished. We also added two questions (Questions 28 and 29) to the end of the template that aimed to capture overall impressions and feedback.

3.2.3 Management of results reporting

As set out in the project's description of work, the ExPaNDS PaN RI FAIR evaluation exercise was intended to be an "open self-assessment".[51] Two important points can be drawn from this description:

1. The results of the exercise should be made **open**, presumably for each facility separately; and,
2. The focus is on **self-assessment**, implying that the main beneficiaries of the exercise should be the facilities themselves, and furthermore, that the primary usefulness of the results for each facility will be for that facility.

The first of these points was implemented by ensuring that the ExPaNDS partner RIs understood from the start that their finalised reporting templates would be published as part of the ExPaNDS FAIR assessment D2.6 deliverable (i.e. the current document) and that we had their permission to do so. We included notices to this effect in several places on the questionnaire and reporting template (see Chapter 4). We also raised this issue in the FAIR assessment workshops (see Section 3.3) and addressed any concerns (of which, there were only one or two, related to purely practical matters, e.g. what file format in which to submit the finalised responses).

Emphasising the self-assessment aspect required a multi-pronged approach. As set out in Section 3.2.2, one action was to ensure we included sufficient opportunity for thoughtful and reflective responses through our design of the questionnaire. In addition to this, though, we

[50] See Section 5.2.1 for further discussion on this point.
[51] ExPaNDS (2018). ExPaNDS European Open Science Cloud Photon and Neutron Data Services [proposal].

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

also went to great lengths to drive the point home about the self-reflective nature of the assessment in our advocacy and at the workshops (see Section 3.3) when we discussed the aim and purpose of the FAIR evaluation exercise with ExPaNDS partners. The final key component of our emphasis on self-assessment was to set out the bottom line right at the start of the task that we would not be seeking to compare responses across facilities, beyond summarising overall impressions around the usefulness of the evaluation for the facilities and drawing out specific points of feedback related to the evaluation method, questions, and design of the reporting template.[52] This was deemed an appropriate stance given the stated focus on self-assessment.

It is important to note that existing evaluation frameworks that involve self-assessment such as CoreTrustSeal (and, therefore, one assumes, also CoreTrustSeal+FAIRenabling) and ACME-FAIR do not invoke comparisons across repositories or organisations; indeed, in the case of CoreTrustSeal, a repository's 'supporting evidence' is only made public if/once they achieve certification, and ACME-FAIR has no requirement at all for the publication of results, meaning it can be used purely internally if organisations wish. It was our strong belief that if we wanted to support ExPaNDS partners to feel free to be open with 'where they are' on the FAIR journey, then it was vital that facilities did not feel that they would be unfairly compared with others, especially given that, despite all being national PaN RIs, the ExPaNDS partners work within different resource constraints, are at different stages of development and implementation, and may vary in their immediate and mid-term strategic priorities and planning goals.

3.3 Engagement with ExPaNDS Partners

Our engagement with ExPaND partners around the FAIR self-assessment exercise involved advance advocacy, a workshop series, and ongoing, individualised support during the evaluation process.

3.3.1 Scene setting advocacy

In advance of the self-assessment exercise being run over summer 2022 and into the early autumn, advocacy related to the ExPaNDS FAIR self-assessment task began in January 2022 with a formal presentation, given by the leader of WP2, about the assessment task at the ExPaNDS extended Project Executive Board (PEB) meeting. The intention behind this early advocacy work was that, because colleagues from all ExPaNDS PaN RIs were invited to attend this extended PEB, the meeting provided an excellent opportunity to formally brief all partners on the upcoming self-assessment task. Providing this advance briefing and notice was deemed especially important, given that ExPaNDS partners would need to set aside sufficient time and resources over a period of several months mid-year to complete the FAIR self-assessment.

3.3.2 Workshops

Once the FAIR self-assessment template was finalised, WP2 ran two workshops to support the self-assessment exercise. The first of these, held in July 2022, presented the aims of the ExPaNDS PaN RI FAIR self-assessment, introduced the questionnaire template to the facility coordinators for the exercise, and offered an opportunity for discussion and questions. The follow on workshop, which ran at the end of September 2022, focused on presentations from those colleagues who had coordinated the FAIR self-assessment at their respective facilities. Each ExPaNDS partner who presented shared high-level outcomes of

[52] These are presented in Chapter 5.

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

the exercise for their facilities and provided feedback on the questions that made up the FAIR self-assessment questionnaire.

3.3.3 Individualised support

In addition to these two workshops, WP2 colleagues involved in the design of the FAIR self-assessment template were available by email and through meetings to respond to any questions or concerns raised by ExPaNDS partners throughout the duration of the FAIR evaluation exercise. In particular, we helped to clarify the meaning of questions on the questionnaire template, where partners were unsure. Ad hoc discussions focused on the purpose and goals of the FAIR self-assessment also took place in the regular WP2 monthly catch up meetings, as well as in meetings with individual ExPaNDS partners.

4. ExPaNDS PaN RI FAIR Self-Assessment Questionnaire and Reporting Template

This chapter presents the **blank** questionnaire and reporting template we designed and used in the ExPaNDS PaN RI FAIR self-evaluation exercise. We reproduce the template in full, including all introductory text. The development and design of this template is overviewed in Chapter 3 and described in more detail in Appendix A.

Copies of the **completed** self-evaluation reports for each of the ten ExPaNDS partner RIs are available in Appendix B. Although these reports stand as a key output of the FAIR evaluation task, i.e. they are not simply ‘supplementary material’, contrary to standard writing practice, we include them in an appendix (rather in the body of the text) so as not to disrupt the flow of the argument made in the chapters that comprise the body of the deliverable.

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

Questionnaire and Reporting Template for ExPaNDS PaN RI FAIR Self-assessment (with Introduction)

Introduction

Aim and purpose of the ExPaNDS FAIR self-assessment exercise

The description of work for ExPaNDS WP2 includes a task focused on FAIR evaluation. Specifically, the task seeks to support ExPaNDS partner facilities to undertake and report on a FAIR self-assessment. While it is our sincere hope that all the facilities will benefit from this exercise, we fully expect the outcomes to be different for every facility: as with all self-assessments, what is important is what the facilities take away from the exercise for themselves, especially in terms of new insight and potential avenues for future development.

Existing models that seek to evaluate the FAIRness of data, i.e. its Findability, Accessibility, Interoperability, and Reusability, assess either the end product (i.e. the FAIR dataset - see, for example, the F-UJI tool) or the entry point to the FAIR dataset (i.e. the repository - see, for example, CoreTrustSeal+FAIRenabling). Given, however, that PaN science requires PaN facilities, the processes and workflows of those facilities become absolutely crucial in enabling FAIR. If facility processes, workflows, and data management practices do not support FAIR across the experimental lifecycle, then PaN researchers have little hope of ending up with FAIR data from their experiments. Thus, the questions that make up the ExPaNDS FAIR self-assessment methodically address the FAIR principles (see Box 2; 'The FAIR Guiding Principles), seeking to make explicit how and to what extent these principles are incorporated into facility workflows, processes, and data management across the experimental lifecycle.

Publication of self-assessment reports

IMPORTANT ***PLEASE READ***

The information you provide below in response to the questions in the FAIR self-assessment template will be published openly 'as is' and unedited as part of the ExPaNDS project's deliverable *D2.6: Self Evaluation Photon and Neutron RIs for FAIR data certification*. As such, we recommend that you do not include any sensitive or confidential information that you would not want to be made public. By submitting your completed responses below to ExPaNDS WP2, you confirm that you have the appropriate sign off from your facility (if/as required) and that you agree to the publication of your responses in full in the D2.6 deliverable as set out above.

Acknowledgement of coordinators and contributors

Unless requested otherwise by the individuals involved, the coordinators for the facility self-assessment exercises will be included in the list of authors of the D2.6 deliverable.

We will also include a 'List of Contributors' as part of the text of D2.6. If you contribute to the self-assessment responses and would like your name to be included on this list, please let

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

your coordinator know, so that they can pass your request on to ExPaNDS WP2. Contributors may also remain anonymous, if they prefer.

Questions that make up the FAIR self-assessment

In this template, the questions to be addressed during the self-assessment exercise are grouped under seven broad areas of relevance to FAIR:

1. the existence, completeness, and richness of metadata related to experiments at your facility
2. the flexibility and capability of any search functionality/service you provide
3. standardisation as used in research data management processes at your facility
4. the indexing and harvesting of your facility's metadata by machines
5. the use of persistent identifiers (PIDs) at your facility
6. the access to data that you provide to human users and machines
7. your facility's long term curation of data

We also list the FAIR principle(s) specifically addressed by the questions that fall under each topic. There are 29 questions in total, some of which are in multiple parts.

The questions in this template derive from a systematic analysis of the FAIR principles and existing FAIR assessment models, including CoreTrustSeal+FAIRenabling and the RDA FAIR data maturity indicators. We are happy to provide the details of this analysis on request. ExPaNDS deliverable D2.6 will also set out the analysis approach in full.

Some of the questions are intentionally vague with respect to the subject of the question, for example when referring to "data" (which could include raw data, analysed/processed data, published data etc.), or when there are multiple beamlines or instruments that might handle things differently. The intent is that the interpretation of such questions is up to the individual facility, choosing whatever seems most useful or natural. An introductory question (numbered 0) allows you to explain this context, so that the interpretation is clear to other readers.

Please note that the questions are all intended to capture a snapshot of the present state. Any foreseen future developments can be mentioned under question 28 as part of the summary reflections.

FAIR self-assessment questionnaire and reporting template

IMPORTANT ***PLEASE READ***

The information you provide below in response to the questions in the FAIR self-assessment template will be published openly 'as is' and unedited as part of the ExPaNDS project's deliverable *D2.6: Self Evaluation Photon and Neutron RIs for FAIR data certification*. As such, we recommend that you do not include any sensitive or confidential information that you would not want to be made public. By submitting your completed responses below to ExPaNDS WP2, you confirm that you have the appropriate sign off from your facility (if/as

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

required) and that you agree to the publication of your responses in full in the D2.6 deliverable as set out above.

Comments on context

0. If you wish, please explain how you have chosen to handle references in the questionnaire to "data" (for example, what type[s] of data are covered by your responses), and/or the situation of multiple beamlines/instruments with different practices (for example, by focusing on only one), and any other context that will help to interpret your responses.

Existence, completeness, and richness of metadata

FAIR – and especially Findability and Reusability – relies heavily on data having associated metadata. Metadata is structured information about data. There are many types of metadata, each providing different information about a data resource. For example, descriptive metadata, with its focus on identifying characteristics and what the data is about, is key to discovery and citation, while rights metadata sets out conditions on reuse of the data and provenance metadata details the origin and history of the data, including the processes and workflows that produced it. To best support the Findability and Reusability of data, its metadata should be complete and as rich as possible, documenting a range of different attributes about the data.

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

1. Is metadata associated with the data that are created/collected across the experimental lifecycle at your facility?

☐ Yes ☐ No

2. Looking at the metadata listed as essential in the ExPaNDS metadata framework across all stages of the experimental lifecycle (see Section 6.3, pages 54-56, Essential metadata to capture in each stage of the experimental lifecycle), is all of this metadata captured (i.e. either automatically or manually) at your facility?

☐ Yes ☐ No

If there are any essential metadata that you do not capture, what are they?

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

3. When data from your facility is accessed, do you provide any metadata that is specifically intended to aid the reuse of that data, i.e. in distinction from metadata for discovery of the data (e.g. do you provide metadata that gives contextual information about how the data was generated)?

☐ Yes ☐ No

If so, what is the metadata that you provide to aid reuse?

Search (flexibility and capability)

Findability is the first component of FAIR. Search, underpinned by metadata, enables Findability. Search should be flexible and capable of meeting a range of needs, from browsing and basic discovery to highly-specified, focused queries. In practice, this means that metadata needs to be searchable in a variety of ways, for different purposes, and by general users, domain experts, and machines.

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- F4. (Meta)data are registered or indexed in a searchable resource

4. Is it possible to search metadata related to data from the experimental lifecycle at your facility?

☐ Yes ☐ No

5. Does the metadata enable basic discovery (e.g. does the metadata include bibliographic information such as author, title, date, etc.)?

☐ Yes ☐ No

Can you make multi-faceted, PaN-specific queries (e.g. technique, experimental parameters, instrument, sample)?

☐ Yes ☐ No

Any additional comments on how your metadata enables discovery of data via search?

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

6. How is the metadata searchable? (select all that apply)

☐	Free text search
☐	Controlled vocabulary
☐	Filters (e.g. date, topic, instrument, etc.)

Are there any other ways of searching your metadata?

7. Who can search the metadata (e.g. searching restricted in some way, searchable by anyone)?

Standardisation

Standardisation can take many forms in data management. For example, metadata standards, controlled vocabularies, formally-expressed linked relationships (e.g. triples), and file formats are all types of standardisation. In relation to FAIR, standardisation is especially important for Findability, Interoperability, and Reusability in that it supports richness and enables machine interoperability.

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- F4. (Meta)data are registered or indexed in a searchable resource.
- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data.
- R1.2 (Meta)data are associated with their provenance

8. Which metadata standard(s) does your facility use for data? (select all that apply)

☐	Dublin Core
☐	DataCite
☐	DCAT
☐	NeXus

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

☐	B2FIND
☐	OpenAIRE (i.e. DataCite with minor adjustments)

Do you use any other metadata standards for data?

9. Does your facility use controlled vocabularies for metadata description?

☐ Yes ☐ No

If 'Yes', which one(s) do you use and for which types of metadata (e.g. PaNet ontology for identifying the technique used, list of keywords covering topics or research areas, fixed list of instruments, etc.)?

10. Does your facility link data/metadata with other relevant data/metadata (e.g. do you link data/metadata from your facility to related experiments, resulting publications, calibration data, etc.)?

☐ Yes ☐ No

If 'Yes', do you do this in a way that standardises and captures these relationships formally (e.g. through the use of a 'related resource' metadata field or similar)?

☐ Yes ☐ No

Any additional comments on how your metadata supports formalised links between data/metadata?

11. Does your facility record and make available information about the provenance of the data offered for reuse (e.g. the experiment or research project with which it originated, the processing it has undergone, any curation actions that have been applied)?

☐ Yes ☐ No

If it does, is the provenance information made available in any specific standard form, either for a particular community (e.g. in NeXus for the PaN community) or cross-domain (e.g. the PROV-O language)?

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

☐ Yes ☐ No

If 'Yes', then please provide further details on the specific standard(s) you use for provenance information:

12. What file formats are used to store and process data at your facility?

13. If your facility employs multiple file formats, how do you make these interoperable within and across your facility?

More generally, how do you ensure that the file formats you use are interoperable with standard file formats widely used outside of PaN facilities (e.g. .txt, .csv, .pdf, .png, etc.)?

Indexing and harvesting of metadata by machines

A fundamental aspect of the FAIR principles is that they seek to make data Findable, Accessible, Interoperable, and Reusable for both humans *and* machines. Indeed, the prediction of an increasing role for machines in the research data sphere was a key driver behind the creation of the FAIR principles. Thus, when it comes to enabling Findability, it is very important that machines can index and harvest metadata related to datasets.

The questions asked in this section relate specifically to the following FAIR principle:

- F4. (Meta)data are registered or indexed in a searchable resource.

14. If your facility maintains a metadata catalogue, can this metadata be queried and retrieved using an API?

☐ Yes ☐ No

15. If your facility maintains a metadata catalogue, are the metadata standards you use compatible with OAI-PMH?

☐ Yes ☐ No

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

If 'No', do you maintain a mapping from your metadata to the metadata standards (e.g. Dublin Core, DataCite) commonly used by OAI-PMH?

☐ Yes ☐ No

PIDs

Persistent identifiers (PIDs) can represent a wide variety of digital and real-world entities in the research process. Not only do they enable instant resolution to the entity in question, but they can also support the linking together of digital objects, thus providing a context for their interpretation and reuse. The assignment and use of PIDs is fundamental to the FAIR principles, especially for the attributes Findable and Accessible.

The questions asked in this section relate specifically to the following FAIR principles:

- F1. (Meta)data are assigned a globally unique and eternally persistent identifier.
- F3. Metadata specifies the data identifier
- A1. (Meta)data are retrievable by their identifier using a standardised communication protocol

16. What specific persistent identifier service(s) does your facility use? (select all that apply)

☐	Digital Object Identifier (DOI)
☐	Handle
☐	Open Researcher and Contributor ID (ORCID)
☐	Research Organization Registry (ROR)
☐	International Generic Sample Number (IGSN)

Does your facility use any other persistent identifier services?

17. If your facility has a metadata catalogue, do you include hyperlinked PIDs as part of the metadata provided on landing pages in that catalogue?

☐ Yes ☐ No

18. Do you use any types of PIDs within your facility's data management processes that you don't make available on your metadata catalogue landing pages?

☐ Yes ☐ No

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

If 'Yes', please provide further details (e.g. we store ORCIDs in our proposal system, but we do not include ORCIDs alongside author names on our catalogue landing pages).

Access to data by users and possibly by machines

The FAIR principle of Accessibility requires that the user should know how the data of interest, once found, may be accessed. This has two aspects: a method for being able to retrieve it, and possibly authentication and authorisation if there are conditions under which the data is available. The second aspect is also connected with Reusability: being clear about the conditions for reuse by means of licenses. Licenses should preferably be standard and familiar to the expected users. Automated access to data through defined interfaces may be important to allow not only automated discovery but the possibility of subsequent access and processing without human intervention.

The questions asked in this section relate specifically to the following FAIR principles:

- A1. (Meta)data are retrievable by their identifier using a standardised communication protocol
- A1.1 The protocol is open, free, and universally implementable
- A1.2 The protocol allows for an authentication and authorization procedure, where necessary
- R1.1 (Meta)data are released with a clear and accessible data usage license

19. How do potential (re)users of your facility's data know how they can access it? (e.g. this information is described on a web page, provided as part of the metadata, etc.)

20. Are there any authentication and authorisation measures in place to control or monitor who has access to your facility's data?

☐ Yes ☐ No

21. What protocols are employed for accessing the data? Please comment on whether the protocols are open and free (no cost) (e.g. HTTP, FTP), whether they depend on the size or other properties of the dataset, and whether you allow for the possibility of automated access to data through the protocols, as well as human access.

22. Does your facility have license agreements on data that it stores and makes available that apply to the users of that data and make clear how they may reuse it?

☐ Yes ☐ No

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

23. Are the licences you use standard ones (e.g. Creative Commons 4.0, public domain dedication CC0, etc.)?

☐ Yes ☐ No

Please list the types of licenses you apply to data at your facility.

24. Are the licences you use machine-understandable as well as readable by humans?

☐ Yes ☐ No

Curation of data

Long-term curation of data is not an explicit requirement of FAIR, but it is implicit in both Accessibility and Reusability. Maintaining integrity and authenticity of data over time are obviously essential for the data to be a reliable resource. Even if a dataset might be removed for some reason, it is desirable for FAIR that the metadata should be retained, even if it only points to a “tombstone” record. Monitoring of possible changes in the environment, such as community practices, is desirable to ensure continued reusability of data in the future.

The questions asked in this section relate specifically to the following FAIR principles:

- A2. Metadata should be accessible even when the data is no longer available
- R1.2 (Meta)data are associated with their provenance
- R1.3 (Meta)data meet domain-relevant community standards

25. Do you have measures in place at your facility to ensure the integrity of data files against unintentional or unauthorised alteration?

☐ Yes ☐ No

If ‘Yes’, please provide further details of the measures you have in place.

Reminder: please ensure that you do not include any sensitive or confidential information in your response.

26. Is it ever possible that a dataset generated with its metadata during the experimental lifecycle could be removed from the facility’s record?

☐ Yes ☐ No

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

If 'Yes', is the associated metadata retained when the dataset is removed?

☐ Yes ☐ No

Please provide any additional comments on the deletion or retention of data/metadata at your facility:

27. Do you take steps against the possibility of changes over time (e.g. evolution of community standards) that might affect the reusability of the data your facility holds?

☐ Yes ☐ No

If 'Yes', please provide further details of the steps you take.

Summary reflections

Looking back across all seven topics addressed in the FAIR self-assessment as well as your facility's overall experience of the exercise, please provide summary comments and reflections.

28. What does your facility take away from the FAIR self-assessment exercise, especially in terms of new insight into the FAIRness of your workflows and data management processes and potential avenues for future development?

29. Please provide feedback and suggestions for improvement of the approach used for the FAIR self-assessment (e.g. questions asked, facility coordinator role, workshops, explanatory notes provided, links to existing FAIR evaluation models, focus on workflows and processes, etc.).

NOTE: please ensure you do not include any personal identifying information (e.g. names) in your response.

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

5. Reflections on the Self-Evaluation Exercise

This chapter considers ExPaNDS partners' overall impressions of the FAIR self-assessment exercise and summarises key feedback, including participants' views on what worked well and their suggestions for changes and improvements.

5.1 Was it Useful?

A key aspect used to judge the successfulness of the FAIR self-assessment exercise was its usefulness (or not) to ExPaNDS partner facilities. Overall, partners found undertaking the evaluation useful, and indeed, several remarked that they saw benefit in repeating the exercise in future. Participants also found it helpful that the approach could be used for two different purposes when evaluating FAIRness: to establish a baseline and to measure progress.

5.1.1 Overall impressions

Overall feedback from the ten ExpaNDS partner PaN RIs strongly indicated that they found the FAIR self-assessment a useful undertaking. Many participants commented that the results provided a good baseline for understanding their facility's current level of FAIRness. The exercise also helped ExPaNDS partners to gain useful insight and encouraged them to articulate clearly what implementations are in progress or planned to support FAIR at their facilities in future.

As further evidence of the usefulness of the FAIR evaluation exercise, several ExPaNDS partners remarked that they could see real benefit in repeating the FAIR self-assessment on an annual basis for internal use at their facilities.

5.1.2 Dual purpose

Partners emphasised that they felt that the FAIR self-assessment could be employed with different purposes in mind. In particular, the evaluation exercise could serve a dual-purpose:

1. To establish a baseline or measure current levels of FAIRness; and/or
2. To evaluate progress towards FAIRness.

In this regard, the ExPaNDS partners' observation strongly echoes the two different evaluation perspectives (i.e. ‘'pass/fail' and 'measuring progress') supported by existing FAIR assessment tools such as the RDA Maturity Model (see Section 2.1.2). On the one hand, partners felt the self-assessment provided good information on their current state of FAIRness, highlighting what was in place and what was not (i.e. primarily through the 'yes/no' questions on the template — but see further comments on this point in Section 5.3 below); on the other hand, partners felt able to use the self-assessment to better understand their planned journey towards increasing FAIRness and to articulate what was in progress and how far along that work was.

5.2 What Worked Well

ExPaNDS partners felt that several aspects of the FAIR self-assessment exercise worked especially well. These aspects included the level of burden, the clarity of the template questions and the support provided when things were not clear, and the reflective nature of the exercise.

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

5.2.1 Balancing the usefulness of the exercise and the effort required

A key concern in the design of the FAIR self-assessment exercise was the need to balance usefulness and thoroughness in capturing detailed information related to FAIR with realistic resource expectations in a time-limited exercise. To this end, we were keen to develop an approach that would not be seen as over-burdensome. Feedback from ExPaNDS partners strongly suggested that we achieved our aim in this regard: partners felt that the level of commitment required and the time and resource needed to complete the FAIR assessment was appropriate, but that at the same time, the exercise was useful and successful.

5.2.2 Support offered and clarity of questions

Participants in the FAIR self-assessment exercise noted that, in general, the questions included in the template were clear and easy to understand. In the few cases where additional clarity was needed, this was dealt with over email or through other discussion one to one with the partners. Related to this, participants also indicated that the support provided to the facility FAIR assessment coordinators by WP2 leads for the FAIR self-assessment task was helpful. The workshops (see Section 3.3) offered additional opportunities for help with any outstanding issues.

5.2.3 Reflective nature of the exercise

As explained in Chapter 3, there was a reason to avoid the use of formal indicators in the FAIR self-assessment. In particular, we sought to promote the exercise as an opportunity for ExPaNDS partners to reflect on issues related to FAIR and their progress on addressing these. Feedback from ExPaNDS partners suggests that they, too, felt not including formal indicators in the assessment was the right course of action, especially as this was the first FAIR self-assessment that any of the partners had undertaken. There was some discussion that, perhaps in future, some use of indicators might have a role, especially in terms of allowing a facility to compare itself year on year. However, in general — and, certainly at present — ExPaNDS partners welcomed the free text options and flexibility of the template and felt that these encouraged reflection on FAIRness levels, whereas, the use of formal indicators could have been discouraging at this stage.

5.3 Suggestions for Changes and Improvements

As part of the feedback from the exercise, the WP2 FAIR assessment task leads were keen to receive suggestions for changes and improvements to the overall approach used and/or the questions asked on the template. In general, participants seemed happy with the methodology and approach used, and there were no major suggestions for changes. However, we did receive some consistent feedback relating to aspects of the reporting template and the questions themselves.

5.3.1 Use of binary ‘yes/no’ questions

Several ExPaNDS partners remarked on the ‘yes/no’ approach used for many of the questions. As highlighted above in Section 5.1.2, this approach echoes aspects of a ‘pass/fail’ evaluation method. While partners did acknowledge this point and see its potential usefulness, for example, in establishing a baseline, they also felt it did not let them show the bigger picture adequately at times. This was particularly the case when work on an implementation aspect was underway but not yet complete; for such cases, some partners

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

would have welcomed an option such as 'work in progress' or similar. It was agreed that even more use of free text boxes might offer another way to address the issue.

This feedback was quite interesting for the WP2 task leads, given that we had sought to reduce the burden on participants by not requiring too many free text answers. There also remains the open question as to whether or not partners would have been able to see (i.e. acknowledge) the areas of FAIR that needed work as clearly if they had not had to choose from the binary 'yes/no' option.

5.3.2 Feedback on the question about 'essential' metadata (i.e. Question 2 in the self-assessment questionnaire)

While the clarity of the questions included in the FAIR self-assessment questionnaire and reporting template was not generally raised as an issue, there was one question — Question 2 — that partners felt should be reworked, especially in light of the 'yes/no' discussions outlined in section 5.3.1. Question 2 comprises two parts, a 'yes/no' answer followed by further details supplied as free text:

> "Looking at the metadata listed as essential in the ExPaNDS metadata framework across all stages of the experimental lifecycle (see Section 6.3, pages 54-56, Essential metadata to capture in each stage of the experimental lifecycle), is all of this metadata captured (i.e. either automatically or manually) at your facility? ['yes/no' tick boxes]
>
> If there are any essential metadata that you do not capture, what are they? [free text box]"[53]

There were two main issues that ExPaNDS partners had with Question 2. The first was the 'yes/no' option. Because the answer applied to all 31 'essential' metadata listed in the ExPaNDS metadata framework and because PaN RIs are still at an early stage of the FAIR journey, participants argued that no facility would realistically be able to answer 'yes' to this question in this initial FAIR self-assessment. And, indeed, this would seem to be the case: only one facility answered 'yes' but even this answer was followed by a clarifying statement in the free text box, indicating that some essential metadata could be missing if they were deemed not relevant to the experiment.

The second concern with Question 2 related to the definitions of the 'essential' metadata types as given in the ExPaNDS metadata framework documentation. As shown in the quote above, Question 2 included a link to ExPaNDS deliverable D2.2: Draft recommendations for FAIR Photon and Neutron Data Management,[54] and, in particular, to the section of that document that lists concisely the 'essential' metadata elements of the framework. The problem encountered by partners undertaking the FAIR self-assessment was that the list in Section 6.3 of D2.2 does not include the description of the metadata element, but rather, only its name. The definitions and descriptions of the metadata elements are actually found scattered under the relevant sections throughout D2.2, which means that considerable extra effort is needed to find them. As well, in some cases, the definitions themselves can be unclear, further confusing matters. While this is not a problem with the FAIR self-assessment template per se, if Question 2 (or similar) is asked again in a future FAIR self-assessment, it would likely be helpful for participants to have easy access, not only to the name of the metadata element, but also its detailed definition/description.

[53] See Chapter 4.

[54] Salvat, D., Gonzalez-Beltran, A., Görzig, H. et al (2020). ExPaNDS D2.2: Draft recommendations for FAIR Photon and Neutron Data Management. https://doi.org/10.5281/zenodo.4312825

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

5.3.3 For PaN RIs, the context is not always PaN

The WP2 FAIR self-assessment task intended and was designed to focus on PaN RIs. However, some ExPaNDS partners operate in a wider context, for example, where a PaN facility is a part of a wider, coherent science institution. In such cases, issues such as metadata may well need to be considered in a context that is wider than simply PaN. Feedback about the reporting template suggested a need to consider this more in the future design of PaN-related FAIR self-assessment exercises, especially in relation to the wording of questions. This said, partners also recognised a need not to stray too far from the detail of the PaN context, especially where workflows and processes are a desired focus of the FAIR self-assessment.

6. Further Observations on FAIR in ExPaNDS PaN RIs

6.1 Introduction

The work presented in this deliverable, involved the development of a FAIR self-evaluation template (the questionnaire), based on and adapting existing approaches to FAIR assessment, and which would be used by the RIs in the ExPaNDS project to examine their own progress towards FAIR. The method was also reviewed with recommendations for change for future assessments.

Feedback after the exercise was complete recommended that a commentary should be included on the content of the questionnaire responses themselves. Thus this section highlights some general findings, while remaining consistent as possible with the initial aims and methodology.

- During the assessment, facilities were assured that there was to be no view taken of individual progress or comparison between facilities to encourage openness and honesty. To preserve this commitment, the observations have been kept at a general level, to give an overall sense of where the ExPaNDS partnership currently stands on the FAIR journey.
- Feedback was received from participants on the binary yes/no nature of some of the questions, where participants felt this approach was too rigid to reflect the status of their facility. Consequently summarising results of these questions was not felt to be informative in its current form, though general observations are possible.
- We sought consent from participating partners and circulated the additional observations.

With these points in mind, we make some observations on the general state of FAIR implementation across the partnership.

6.2 Observations on FAIR Readiness

The first point to emphasise is that PaN facilities in ExPaNDS find themselves at different stages of their FAIR journey, which they recognise clearly. This means that there are in some cases wide divergences between facility responses in parts of the self-evaluation. We divide further observations structured around the four FAIR characteristics.

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

6.2.1 Findability

Assigning metadata and PIDs are key to findability. The facilities have different processes for populating metadata, involving both automated and manual sources, which have implications for consistency and uniformity of the metadata. Having said that, all the facilities associate some metadata with data that is collected as part of experiments, and most provide metadata for basic discovery. There are differences between beamlines depending on their age, established practices, and their use of data catalogues, with implementations in place or under-development, using ICAT, SciCAT or bespoke systems.

DOIs are commonly accepted, becoming implemented and used as PIDs, although others are also in use, and ORCIDs are recognised as valuable to identify researchers.

6.2.2 Accessibility

A complicating factor across the facilities is the approach taken to openness of data: time-limited embargos on access to data are common, and there may be restrictions on access and search, typically to members of the experimental team. Most facilities provide an API to query the metadata catalogue. Access to the data is possible through a variety of protocols, typically HTTPS, FTP and Globus. Again there are varying degrees of openness and automation according to policies and technical constraints.

6.2.3 Interoperability

Linking of data/metadata is widespread, with use of appropriate standards. NeXus/HDF5 is put forward as fundamental as a source of detailed experimental metadata, with DataCite metadata also important for citation and discovery. Controlled vocabularies are not widespread; some facilities use their own vocabularies, and adoption of the PaNet ontology is envisaged in future. Filters are offered as an alternative to free text search.

Very few facilities capture all the metadata listed as “essential” in the ExPaNDS metadata framework across all stages of the experimental lifecycle. Sometimes this is because metadata catalogues are still under development, or else it is not available. Often sample information is not captured; this evidences efforts in the community that seek to address this issue, such as the LEAPS-STARS[55] initiative. Further, some specialised fields that might not always be relevant in context, such as proposed experiment conditions, are not always supplied.

6.2.4 Reusability

Most facilities provide metadata intended to aid reuse as well as metadata for basic discovery. The exact metadata fields vary but may cover sample environment or beam intensity and other technical characteristics. In some cases, the metadata is provided elsewhere e.g. electronic logbooks, or the metadata within NeXus files is considered sufficient.

Licensing is typically some variant of Creative Commons.

There is some availability of provenance of data, though specialised formats like PROV-O are not used. Links in NeXus and DataCite metadata help to establish provenance of datasets and publications.

[55] https://leaps-initiative.eu/wp-content/uploads/2021/10/DIGITAL-LEAPS-August-2021.pdf

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

Regarding long-term preservation, all facilities have measures to ensure integrity of data files against alteration. The priorities and perceptions differ: some ExPaNDS partners are concerned mainly with integrity of files through backups, while others emphasise the situation of deletion (whether intentional or accidental) and its consequences. A small majority allow the possibility of the removal of a dataset from record. A minority actively monitor for changes affecting future reuse, though the NeXus standard itself is seen as offering long-term security.

7. Concluding Remarks

The FAIR Principles have captured the imagination of many who work with and manage research data. The acronym is appealing; the four top-level desiderata are memorable and easy to grasp; the breakdown to the next level of criteria is both logical and intuitive. What is more, this second level of FAIR criteria is temptingly expressed as statements that invite 'yes/no' evaluation (as, for example, "F1. (Meta)data are assigned a globally unique and persistent identifier"[56]). It might be thought, therefore, that conducting an evaluation against the FAIR Principles would be straightforward, especially for the ExPaNDS partner RIs, a group of well-established scientific facilities working in a common domain of PaN science. Moreover, these PaN RIs already have a commitment to and acquaintance with FAIR — they would hardly have chosen to participate in the ExPaNDS project if that were not the case.

Of course, the reality is very different: the concept of FAIR may be simple, but its implementation raises many difficulties that must be resolved, as the length of this deliverable attests. Considerable intellectual input was required to make the ideal of FAIR evaluation into something achievable and useful. The first difficulty lies in understanding the nature of the evaluation itself: what is expected to be its end product. Clearly, this has implications for the process and the role of the participants. It was initially thought, at the time of writing the ExPaNDS proposal, that the goal would be some kind of certification: a relatively formal process with a degree of rigour, sometimes conducted by an external party for impartiality and objectivity. However, certification seems to assess primarily the end result (presumably rewarding it with a 'certificate'), whereas the real aim of the exercise was to help PaN RIs support and enhance FAIR at the current stage in their FAIR journey. It was in any case always envisaged that self-evaluation would be conducted, implying a critical self-examination which, if approached with an attitude of openness and honest engagement, would reveal the strengths and weaknesses and signpost future opportunities.[57] In the final version of this deliverable, the self-evaluations are supplemented with some high-level observations on the state of the FAIR journey across the ExPaNDS facilities, highlighting some general points without comparison or judgement.

With this view of the purpose of the evaluation in mind, it would not have been reasonable simply to ask the facilities to gauge themselves against the raw FAIR Principles. Although the FAIR Principles are expressed as seemingly factual statements, they were not intended to be used directly in evaluation. It was for precisely this reason that the RDA Working Group on the FAIR Data Maturity Model produced its comprehensive indicators, recognising that the FAIR Principles themselves are not "strict rules" but rather "may lead to diverse interpretations and ambiguity".[58] But these RDA indicators are domain-agnostic and require

[56] Wilkinson, M. D. et al. (2015). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data*, **3**:1. https://doi.org/10.1038/sdata.2016.18

[57] This is the distinction between 'formative' and 'summative' assessment described in Section 2.3.3.

[58] RDA FAIR Data Maturity Model Working Group (2020). FAIR Data Maturity Model: specifications and guidelines. https://doi.org/10.15497/RDA00050

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

interpretation in particular disciplinary contexts. Other frameworks have also emerged for FAIR assessment (for selected examples, see Section 2.2); hence the need for the four principles underlying the design of the ExPaNDS self-evaluation, set out in Section 3.1.2. A further consideration underlying the whole effort was that the ExPaNDS PaN RI FAIR self-evaluation should be achievable without excessive effort required of the participating facilities.

It can be seen that a considerable effort was needed to produce the questionnaires that were distributed to nominated staff at the facilities for the self-evaluation exercise. These staff in turn took their responsibility seriously, and the WP2 FAIR evaluation task leads would like to thank them for their engagement and effort in completing the questionnaires in a timely fashion: without this commitment from the ExPaNDS partners, the task could not have produced worthwhile results. It is gratifying that, as reported in Chapter 5, the burden on partners was considered worthwhile.

It should be reemphasised that the whole exercise is a contribution to the methodology of FAIR assessment, a systematic and principled application of the FAIR Principles, through the adaptation of existing frameworks, to produce tailored indicators for a particular domain. Although within the ExPaNDS project this has been a once-only effort, there is no need for it to be left there. Opportunities present themselves for:

- repetition over time (an opportunity that was raised by several of the participants themselves) so as to measure progress
- comparison between facilities, not in the sense of competition but rather in uncovering what can be learnt from the practices of others
- extension of the approach to other disciplines

So although what was envisaged as a “certification scheme”[59] in the ExPaNDS work programme may still be a distant prospect, what has been achieved is a profiled adaptation of existing FAIR evaluation frameworks for PaN RIs leading to “an open self-assessment exercise of the national research infrastructures”,[60] just as was planned.

[59] ExPaNDS (2018). ExPaNDS European Open Science Cloud Photon and Neutron Data Services [proposal].
[60] Ibid.

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

References

Bunakov, V., Krahl, R., Matthews, B. et al. (2022). ExPaNDS D2.5: Advanced infrastructure for PIDs in Photon and Neutron RIs. https://doi.org/10.5281/zenodo.5905350

Chue Hong, N., Katz, D., Barker, M. et al. (2022). FAIR Principles for Research Software (FAIR4RS Principles). https://doi.org/10.15497/RDA00068https://doi.org/10.15497/RDA00068

CoreTrustSeal Standards and Certification Board. (2019). CoreTrustSeal Trustworthy Data Repositories Requirements 2020–2022 (v02.00-2020-2022). https://doi.org/10.5281/zenodo.3638211

Davidson, J., Whyte, A., Molloy, L. et al. (2022). Defining the Policy Environment: ACME-FAIR Issue #1 (2.0). https://doi.org/10.5281/zenodo.6345332

de Bruin, T., Sarah Coombs, S., de Jong, J. et al. (2020). Do I-PASS for FAIR. A self assessment tool to measure the FAIR-ness of an organization (Version 1). https://doi.org/10.5281/zenodo.4080867

Devaraju, A. and Huber, R. (2020). F-UJI - An Automated FAIR Data Assessment Tool (v1.0.0). https://doi.org/10.5281/zenodo.4063720

Emonet, V. (2022). FAIR Enough data maturity indicators [fair-enough-data]. https://fair-enough.semanticscience.org/collections/fair-enough-data

EOSC-synergy (n.d.). FAIR Framework. https://www.eosc-synergy.eu/results/fair-framework/

European Commision (n.d.). European Open Science Cloud (EOSC). https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science/european-open-science-cloud-eosc_en

ExPaNDS (2018). ExPaNDS European Open Science Cloud Photon and Neutron Data Services [proposal].

FAIRsFAIR (2022). ACME-FAIR: a guide for Research Performing Organisations (RPO). https://www.fairsfair.eu/acme-fair-guide-rpo

FAIRsFAIR (2022). Assessing capability maturity and engagement with FAIR-enabling practice. https://zenodo.org/communities/acme-fair

FAIRsharing and FAIR metrics groups (n.d.). FAIR Evaluation Services. https://fairsharing.github.io/FAIR-Evaluator-FrontEnd

F-UJI: Automated FAIR Data Assessment Tool (n.d.). About. https://www.f-uji.net/index.php?action=about

F-UJI: Automated FAIR Data Assessment Tool (n.d.). https://www.f-uji.net/

Garijo, D., Corcho, O., and Poveda-Villalón, M. (2021). FOOPS! An Ontology Pitfall Scanner for the FAIR Principles. https://ceur-ws.org/Vol-2980/paper321.pdf

L'Hours, H., von Stein, I., deVries, J. et al. (2021). M4.3 CoreTrustSeal+FAIRenabling, Capability and Maturity (1.0). https://doi.org/10.5281/zenodo.5346822

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

Rans, J and Whyte, A. (2017). 'Using RISE, the Research Infrastructure Self-Evaluation Framework' v.1.1 Digital Curation Centre:www.dcc.ac.uk/guidance/how-guide

RDA FAIR Data Maturity Model Working Group (2020). FAIR Data Maturity Model: specifications and guidelines. https://doi.org/10.15497/RDA00050

Rosnet, T., Gaignard, A., Devignes, M. (n.d.) FAIR-Checker. https://fair-checker.france-bioinformatique.fr/

Salvat, D., Gonzalez-Beltran, A., Görzig, H. et al. (2020). ExPaNDS D2.2: Draft recommendations for FAIR Photon and Neutron Data Management. https://doi.org/10.5281/zenodo.4312825

Science Europe (2021). Practical Guide to Sustainable Research Data. https://doi.10.5281/zenodo.4769703

Soler, N., McBirnie, A., Gonzalez-Beltran, A. et al. (2022). ExPaNDS D2.7: Final recommendations for FAIR Photon and Neutron Data Management. https://doi.org/10.5281/zenodo.6799105

Whyte, A., Molloy, L, Grootveld, M, and Thorley, M. (2022). Supporting Data Management Planning: ACME-FAIR Issue #4. https://doi.org/10.5281/zenodo.6346747

Wilkinson, M. D. et al. (2015). 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 A: Detailed Explanation of the Question Development Process

This appendix explains in detail the steps we followed to develop the final set of questions asked in the ExPaNDS FAIR self-assessment.

A1.1 Underpinning Models and Frameworks

As highlighted in the 'four principles' described in Section 3.1.2, the starting point for both our evaluation approach and the question development was the FAIR Principles. It was fundamental that each question asked in the assessment could be linked explicitly to one or more of the FAIR Principles, a point emphasised in the introductory text to the ExPaNDS FAIR Self-assessment Questionnaire and Reporting Template (see Chapter 4):

> "…the questions that make up the ExPaNDS FAIR self-assessment methodically address the FAIR principles …, seeking to make explicit how and to what extent these principles are incorporated into facility workflows, processes, and data management across the experimental lifecycle." (template introduction, see Chapter 4)

As the template's introductory text goes on to note, we then drew on two existing evaluation frameworks to facilitate the question development process:

> "The questions in this template derive from a systematic analysis of the FAIR principles and existing FAIR assessment models, including CoreTrustSeal+FAIRenabling and the RDA FAIR data maturity indicators." (template introduction, see Chapter 4)

A1.2 Steps in the Process

The development of the questions for the ExPaNDS FAIR self-assessment involved four main steps. The diagram illustrated in Figure A1 presents an overview of these steps, each of which is described in more detail in the sections below.

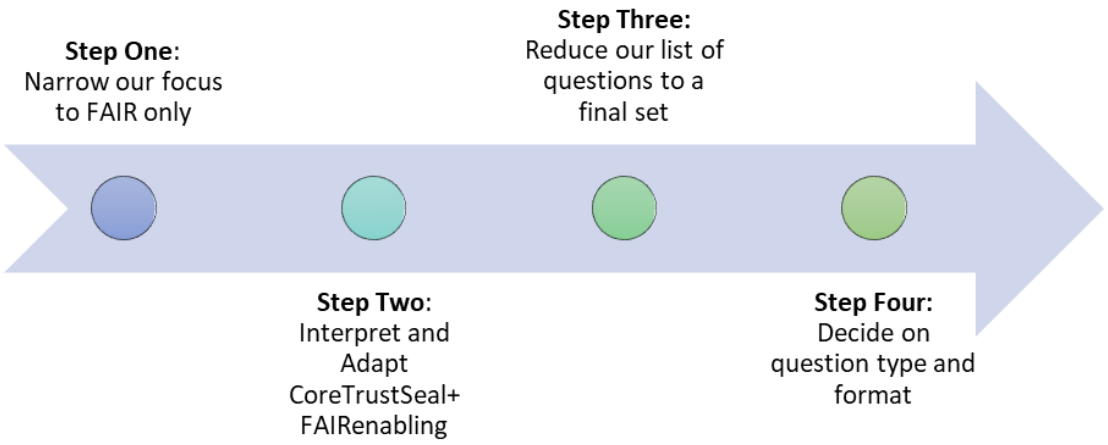


Figure A1: Steps in the ExPaNDS FAIR Self-assessment Question Development Process

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

A1.2.1 Step one: Narrow our focus to FAIR

Having decided to employ the CoreTrustSeal+FAIRenabling framework (see Section 3.1.2), our first step was to narrow our focus to only the +FAIRenabling components of that framework, i.e. because we were undertaking a FAIR assessment specifically rather than a general assessment of processes and workflows.

This meant that we concentrated on the seven (out of sixteen) CoreTrustSeal requirements that included a +FAIRenabling element (see Section 2.2.2):

- R2: Licenses
- R7: Data integrity and authenticity
- R10: Preservation plan
- R13: Data discovery and identification
- R14: ReUse
- R15: Technical infrastructure
- R16: Security

In the CoreTrustSeal+FAIRenabling framework, each of these CoreTrustSeal (CTS) elements is associated with one or more FAIR principles (FAIR):

- R2: Licenses (CTS)
  - ➢ R1.1 (Meta)data are released with a clear and accessible data usage license (FAIR)
- R7: Data integrity and authenticity (CTS)
  - ➢ R1.2 (Meta)data are associated with their provenance (FAIR)
- R10: Preservation plan (CTS)
  - ➢ A2. Metadata are accessible even when the data is no longer available (FAIR)
- R13: Data discovery and identification (CTS)
  - ➢ F1. (Meta)data are assigned a globally unique and eternally persistent identifier. (FAIR)
  - ➢ F2. Data are described with rich metadata (FAIR)
  - ➢ F3. Metadata specifies the data identifier (FAIR)
  - ➢ F4. (Meta)data are registered or indexed in a searchable resource (FAIR)
- R14: ReUse (CTS)
  - ➢ I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation (FAIR)
  - ➢ I2. (Meta)data use vocabularies that follow FAIR principles (FAIR)
  - ➢ I3. (Meta)data include qualified references to other (meta)data. (FAIR)
  - ➢ R1. (Meta)data have a plurality of accurate and relevant attributes

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

- ➢ R1.3 (Meta)data meet domain-relevant community standards (FAIR)

- R15: Technical infrastructure (CTS)
  - ➢ A1. (Meta)data are retrievable by their identifier using a standardised communication protocol (FAIR)
  - ➢ A1.1 The protocol is open, free, and universally implementable (FAIR)
- R16: Security (CTS)
  - ➢ A1.2 The protocol allows for an authentication and authorization procedure, where necessary (FAIR)

Thus, taken together, the seven +FAIRenabling CoreTrustSeal elements address all fifteen FAIR Principles. For readers who find a visual representation helpful, Figure A2 illustrates this mapping (see also Section 2.2.2).

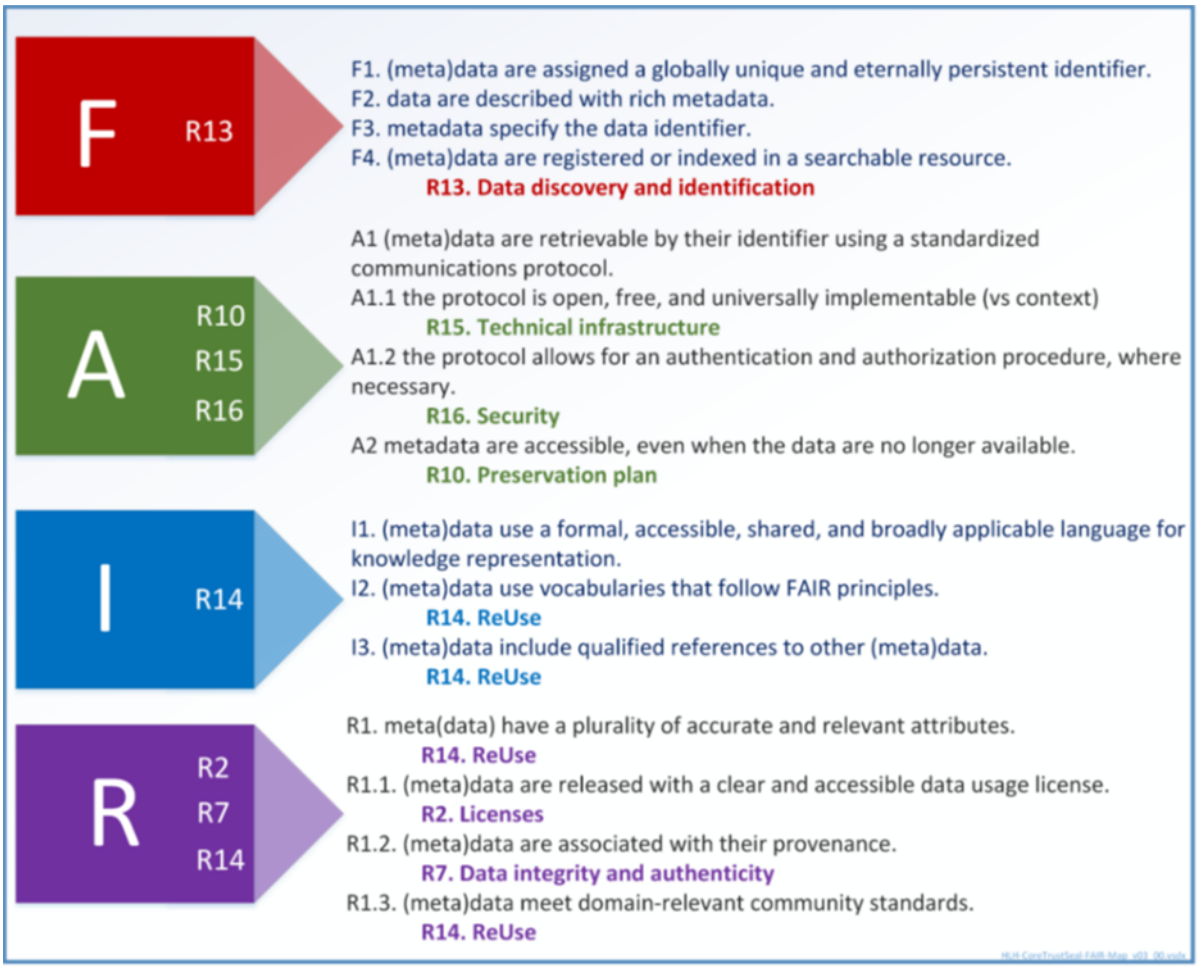


Figure A2: CoreTrustSeal to FAIR Alignment as illustrated in the CoreTrustSeal+FAIRenabling framework (v00.04[61]

As well as maintaining the relationship described above between CoreTrustSeal requirements and the FAIR Principles, each +FAIRenabling statement in the

[61] L'Hours, H., von Stein, I., deVries, J. et al. (2021). M4.3 CoreTrustSeal+FAIRenabling, Capability and Maturity (1.0). https://doi.org/10.5281/zenodo.5346822

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

CoreTrustSeal+FAIRenabling framework is associated not only with its related FAIR Principle(s) but also with a set of relevant indicators drawn both from the FAIRsFAIR FAIR Object Assessment Metrics and the RDA FAIR Data Maturity Model as illustrated in the following example for CoreTrustSeal requirement R2: Licenses:

> '**+FAIRenabling**: digital object metadata includes license information covering (meta)data reuse.
>
> ***"Principle**: R1.1. (meta)data are released with a clear and accessible data usage license.*
>
> **FAIRsFAIR Metric**:
>
> FsF-R1.1-01M Metadata includes license information under which data can be reused.
>
> **RDA Indicators**:
>
> - RDA-R1.1-01M Metadata includes information about the licence under which the data can be reused (Essential)
> - RDA-R1.1-02M Metadata refers to a standard reuse licence (Important)
> - RDA-R1.1-03M Metadata refers to a machine-understandable reuse licence (Important)" '[62]

Thus, using the CoreTrustSeal+FAIRenabling framework, it is possible to explicitly map selected CoreTrustSeal requirements (i.e. requirements 2, 7, 10, 13, 14, 15, 16), +FAIRenabling statements, FAIR Principles, FAIRsFAIR FAIR Object Assessment Metrics, and RDA Maturity Model Indicators to each other. This mapping can be undertaken in total or in part; for example, if one wants to link all of the elements above, then that is entirely possible to do, and indeed, this is a real strength of the CoreTrustSeal+FAIRenabling framework. That said, using the CoreTrustSeal+FAIRenabling framework, in theory, one could also choose to simply make use of the FAIR Principles to RDA indicator mapping; however, in practice, of course, it would probably make more sense simply to employ the RDA FAIR Data Maturity Model framework for this purpose.

In our case, we made use of the entire mapping except for the FAIRsFAIR FAIR Object Assessment Metrics. As explained in Section 2.2.1, the FAIRsFAIR FAIR Object Assessment Metrics draw heavily on the RDA FAIR Data Maturity Model Indicators, and thus, we felt it was unnecessary to use both models in our approach. Our decision was to make use of the more established model of the two, i.e. the RDA FAIR Data Maturity Model.

A1.2.2 Step two: Interpret and adapt CoreTrustSeal+FAIRenabling

Our next step was to interpret — and adapt, where this seemed sensible — the various +FAIRenabling components of the CoreTrustSeal+FAIRenabling framework for the needs of the PaN RI FAIR self-assessment.

As explained in Section 3.1.2, one aspect of this interpretation/adaption was essentially a translation exercise: instead of thinking in the context of 'repositories', translating the context to elements such as 'the experimental lifecycle' and 'workflows and processes' that have specific relevance to PaN RIs. This translation was not always straight-forward, however, so another aspect that became very important was exploring the questions provided under each

[62] Ibid.

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

of the CoreTrustSeal requirements. These questions are meant to help guide the submission of evidence for CoreTrustSeal certification, but they also proved very useful in helping us to articulate relevant questions to ask in our FAIR self-assessment.

As an example of our interpretation process involving both the translation element and the use of the CoreTrustSeal questions, let us consider in detail how we developed the ExPaNDS FAIR self-assessment questions related to the Findability (F) component of the FAIR Principles.

We can see in Figure A2 in Section A1.2.1 above that all elements of the Findability (F) component map to the R13: Discovery & Identification requirement of CoreTrustSeal. For clarity, it is important to note that some of the FAIR components do not have such a straight-forward one to one mapping; for example, both Accessibility (A) and Reusability (R) map to three CoreTrustSeal requirements each.

In the CoreTrustSeal+FAIRenabling framework, there are four corresponding +FAIRenabling statements for R13: Discovery & Identification — two related to 'discovery' and two related to 'identification':

> "*Discovery* …
>
> **+FAIRenabling:** The repository provides evidence that resource discovery metadata is sufficient for their designated community of users. A disciplinary repository may be expected to provide information for both general purpose resource discovery systems (exposure for indexing by search engines, high level metadata such as Dublin Core or DataCite), and metadata to support their more specialist designated community of users."
>
> **+FAIRenabling:** This mapping focuses on the Principle [F4.[63]]
>
> *Identification* …
>
> "**+FAIRenabling:** All objects in the repository are persistently identified. Any exceptions are documented and explained, including a timetable for complete coverage of persistent identifiers.
>
> ***+FAIRenabling:*** *The repository provides evidence that digital object metadata includes its persistent identifier."*[64]

For requirement R13, CoreTrustSeal provides the following set of suggested questions intended to guide evidence submission:

- Does the repository offer search facilities?
- Does the repository maintain a searchable metadata catalogue to appropriate (internationally agreed) standards?
- What persistent identifier systems does the repository use?
- Does the repository facilitate machine harvesting of the metadata?

[63] Note that this +FAIRenabling statement seems to have been left incomplete in the CoreTrustSeal+FAIRenabling framework, so we took it as relating to the matter of (Meta)data being registered or indexed in a searchable resource.

[64] L'Hours, H., von Stein, I., deVries, J. et al. (2021). M4.3 CoreTrustSeal+FAIRenabling, Capability and Maturity (1.0). https://doi.org/10.5281/zenodo.5346822

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

- Is the repository included in one or more disciplinary or generic registries of resources?
- Does the repository offer recommended data citations?

In relation to these questions, considering them in the context of PaN RI workflows and processes and the PaN experimental lifecycle as well as the four +FAIRenabling statements reproduced above, we came up with twenty associated PaN RI-related questions:

- Does the repository offer search facilities? (CTS)

1. Which data from which stage(s) across the experimental lifecycle does your facility make searchable? (e.g. info from the proposal, raw data, analysed datasets)
2. How is the data searchable? (e.g. free text, controlled vocabulary, filtering)
3. Who can search the data? (e.g. restricted in some way, searchable by anyone)

- Does the repository maintain a searchable metadata catalogue to appropriate (internationally agreed) standards? (CTS)

4. Is metadata associated with the data that are created/collected across the experimental lifecycle at your facility?
5. What metadata standard(s) does your facility use?
6. Can metadata related to data from the experimental lifecycle at your facility be searched?
7. Is the metadata such that it enables basic discovery (e.g. bibliographic information)? Multi-faceted, PaN-specific queries (e.g. technique, experimental parameters, instrument, sample)?

- What persistent identifier systems does the repository use? (CTS)

8. At what stage(s) of the experimental lifecycle are persistent identifiers assigned at your facility?
9. To what (people, places, things) does your facility assign persistent identifiers?
10. What specific persistent identifier service(s) does your facility use?
11. If your facility has a metadata catalogue and makes use of persistent identifiers, are these persistent identifiers included in the metadata provided on landing pages in the metadata catalogue? If you do not include certain persistent identifiers on your landing pages, which type are they and to what do they apply?

- Does the repository facilitate machine harvesting of the metadata? (CTS)

12. If your facility maintains a metadata catalogue, can this metadata be queried and retrieved using an API?
13. If your facility maintains a metadata catalogue, are the metadata standards you use compatible with OAI-PMH? If not, do you provide a crosswalk?

Broader questions:

14. At what stage(s) of the experimental lifecycle at your facility is metadata automatically created and recorded by machines?

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

15. At what stage(s) of the experimental lifecycle at your facility is metadata manually created and recorded by humans?

16. Looking at the metadata listed as essential in the ExPaNDS metadata framework, is this metadata captured (i.e. either automatically or manually) during the experimental lifecycle at your facility for each of the stages of the experimental lifecycle? If essential metadata are not captured at any stage(s), which metadata types are these?

- Is the repository included in one or more disciplinary or generic registries of resources? (CTS)

17. Do you make data/metadata related to the experimental lifecycle at your facility available to others beyond staff and users of your facility? If so, how? (Note: not really on topic – but maybe still a useful question?)

18. If you have a metadata catalogue, what external PaN-specific and/or generic resources formally link to or list your catalogue?

- Does the repository offer recommended data citations? (CTS)

19. Does your facility provide examples of how to cite data from the facility?

20. Where/how are these examples provided? (e.g. on landing pages in the metadata catalogue, in the facility data policy, through training activities)

Working in parallel, we undertook this process of associating relevant PaN-related questions to CoreTrustSeal questions for each of the four elements of FAIR (i.e. F, A, I, R). It is important to note that despite the fact that we had the benefit of the +FAIRenabling statements (as well as their corresponding FAIR Principles and RDA Maturity Model Indicators) and the CoreTrustSeal set of suggested questions to guide us, devising the PaN-specific questions could never be a purely objective exercise. There was a necessary subjective element as well, that required us to draw on our knowledge and experience of FAIR, the PaN experimental lifecycle, and the processes and workflows of PaN RIs and their instruments.

A1.2.3 Step three: Reduce questions to final set

Once we had developed our initial set of possible questions for each of the four components of FAIR, our next step was to narrow down this rather large resulting set of 48 questions to a more focused set to use in our FAIR assessment. Again, this step combined objective and subjective aspects.

Looking at both the FAIR Principles and the RDA Maturity Model Indicators, it became evident that some of the questions we had devised could be left off the ‘final’ list because they did not actually address a FAIR Principle specifically. For example, the FAIR Principles do not include reference to data citation, so questions related to data citation (such as questions 19 and 20 in the list directly above in section A1.2.2) could be removed:

- Does the repository offer recommended data citations? (CTS)

19. Does your facility provide examples of how to cite data from the facility?

20. Where/how are these examples provided? (e.g. on landing pages in the metadata catalogue, in the facility data policy, through training activities)

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

There were also many cases where questions were asking the same things in slightly different ways. In such cases, it was often possible to choose just one of the questions, or perhaps, to combine a few questions into one. It is important to emphasise that similar questions could well be created, especially since the work on F-, A-, I-, and R-related questions was done in parallel.

For example, consider the three similar questions below. The first question came out of work done on the 'F' component, while the second two questions resulted from work on the 'I' component. In this example, the two I-related questions can be subsumed into the F-related question:

- What metadata standard(s) does your facility use? **(F-related)**
- What metadata standard(s) are used by your facility to document data/metadata across the experimental lifecycle? **(I-related)**
- If your facility offers a metadata catalogue, what metadata standard(s) does it employ? **(I-related)**

To ensure that objectivity also came into these decisions, we double checked our subjective impressions about the similarity of questions against the RDA Indicators that were linked to the questions. For example, where questions that we considered were similar also addressed the same RDA indicator(s), then we could be more confident that the questions did indeed address a similar aspect of FAIR.

Again, we can see this in action using the example three questions discussed above. Table A1 below presents the three questions, along with their associated RDA FAIR Data Maturity Indicators. Here, we can clearly see that the first question does indeed subsume the next two questions: all address indicator RDA-I1-01M, but only the first question addresses additional elements. Therefore, given that our aim was to reduce the number of necessary questions while still ensuring complete coverage of all FAIR Principles, out of these three example questions, it made the most sense for us to select the first one and leave the other two out.

Relevant RDA Maturity Indicator	**PaN FAIR Assessment Question (i.e. Examples 1 – 3)**
F2 RDA-F2-01M Rich metadata is provided to allow discovery (Essential)	What metadata standard(s) does your facility use?
F4 RDA-F4-01M Metadata is offered in such a way that it can be harvested and indexed	

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

I1 RDA-I1-01M Metadata uses knowledge representation expressed in standardized format (Important)	
I1 RDA-I1-01M Metadata uses knowledge representation expressed in standardized format (Important)	What metadata standard(s) are used by your facility to document data/metadata across the experimental lifecycle?
I1 RDA-I1-01M Metadata uses knowledge representation expressed in standardized format (Important)	If your facility offers a metadata catalogue, what metadata standard(s) does it employ?

Table A1: The three example PaN-related questions, along with their associated RDA FAIR Data Maturity Indicators. Note that the RDA Maturity Indicators are reproduced here as they are presented in the relevant sections of the CoreTrustSeal+FAIRenabling framework (see discussion in Section A1.2.2).[65]

The end result of this reduction process was a final set of 27 questions.[66] As described in Section 3.2.1, these questions were then grouped under seven broad topics. Up until this point, as the steps discussed in this appendix outline, a clear link had been retained from a particular question to its related FAIR Principle(s). However, the grouping of questions into broad topics effectively hid these links. It was for this reason that we included the introductory text to each topic in the template as well as the list of FAIR Principles that the questions under the topics address (see the full version of the template in Chapter 4).

While we did consider listing the relevant FAIR Principle(s) after each question, our feeling was that it would be perhaps easier for users of the questionnaire to gain insight about FAIRness at their facility if they could consider the point of a question in the context of the broader topics rather than within the limited confines of a single FAIR Principle. This was especially the case where a broad topic covered multiple FAIR Principles and it was necessary to understand how those principles worked together under that topic. Both the initial feedback we received (e.g. at the first workshop [see Section 3.3.2] and from one to one discussions) and the final feedback provided post-assessment indicated that our hunch was correct and that participants felt we had provided links back to the FAIR Principles at the correct level of detail.

[65] L'Hours, H., von Stein, I., deVries, J. et al. (2021). M4.3 CoreTrustSeal+FAIRenabling, Capability and Maturity (1.0). https://doi.org/10.5281/zenodo.5346822

[66] Note that the final template actually includes 29 questions in total (see Chapter 4). Questions 28 and 29 were added to capture summary information around insight gained and about feedback on the self-assessment process (see Section 3.2.2). As such, these two final questions did not relate to any specific FAIR Principle(s), nor was it the intention of these two questions to do so.

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

A1.2.4 Step four: Decide on question format

Once we had settled on our final list of questions, the next and last step was to choose a format for each of the questions. This aspect has already been covered in the main body of this report (see Section 3.2.2) so we do not discuss it further here.

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

Appendix B: ExPaNDS FAIR Self-Assessment Reports

This appendix presents the responses to the self-assessment questionnaire as submitted by each of the ExPaNDS partner RIs. The reports appear in alphabetical order by facility name. For brevity, we do not repeat the introductory text or the explanatory text that was supplied alongside each section of the template; for full details on these, see the blank version template supplied in Chapter 4 of the body of this deliverable. For ease of reference, however, in the facility reports, we do retain the questions as asked and also the notes about which FAIR Principles are relevant for each section of the questionnaire.

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

B1.1 ALBA

Comments on context

0. If you wish, please explain how you have chosen to handle references in the questionnaire to "data" (for example, what type[s] of data are covered by your responses), and/or the situation of multiple beamlines/instruments with different practices (for example, by focusing on only one), and any other context that will help to interpret your responses.

The first beamline at ALBA in which our data catalogue (ICAT) has been installed, is NCD-SWEET, dedicated to small and wide angle X-ray scattering. The answers in this questionnaire related to findability/accessibility will therefore largely refer to datasets produced in this beamline or in the other beamline for which ICAT has entered in production (MIRAS, an infrared spectroscopy beamline). In both cases, ICAT is not yet open to the outside world. Please also note that at the moment, only raw data is stored and not processed data. Since the installation of ICAT is recent, we don't have yet open data (having passed the embargo period) available for reuse.

Finally, new beamlines in ALBA such as LOREA tend to adopt NeXus/HDF5 from the beginning so that we refer to it in some of the answers.

Existence, completeness, and richness of metadata

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

1. Is metadata associated with the data that are created/collected across the experimental lifecycle at your facility?

☑ Yes ☐ No

2. Looking at the metadata listed as essential in the ExPaNDS metadata framework across all stages of the experimental lifecycle (see Section 6.3, pages 54-56, Essential metadata to capture in each stage of the experimental lifecycle), is all of this metadata captured (i.e. either automatically or manually) at your facility?

☐ Yes ☑ No

If there are any essential metadata that you do not capture, what are they?

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

Experiment::Sample information (appears somewhat in the title but not in a systematic way)
Storage::Dataset information
Storage::Preservation description information
Data publication::Ressource identity (since we are putting in place the DOI system)

3. When data from your facility is accessed, do you provide any metadata that is specifically intended to aid the reuse of that data, i.e. in distinction from metadata for discovery of the data (e.g. do you provide metadata that gives contextual information about how the data was generated)?

☑ Yes ☐ No

If so, what is the metadata that you provide to aid reuse?

Energy (wavelength), detector type and xy pixel size, elapsed time.

Search (flexibility and capability)

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- F4. (Meta)data are registered or indexed in a searchable resource

4. Is it possible to search metadata related to data from the experimental lifecycle at your facility?

☑ Yes ☐ No

5. Does the metadata enable basic discovery (e.g. does the metadata include bibliographic information such as author, title, date, etc.)?

☑ Yes ☐ No

Can you make multi-faceted, PaN-specific queries (e.g. technique, experimental parameters, instrument, sample)?

☐ Yes ☑ No

Any additional comments on how your metadata enables discovery of data via search?

Basic search functionalities of iCAT in icatprod are available, meaning only filters on lists displayed. No mechanism to search e.g metadata details among all datasets is available (except if they are provided in the title).

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

6. How is the metadata searchable? (select all that apply)

☑	Free text search
☐	Controlled vocabulary
☑	Filters (e.g. date, topic, instrument, etc.)

Are there any other ways of searching your metadata?

Not yet

7. Who can search the metadata (e.g. searching restricted in some way, searchable by anyone)?

Users having performed the experiment, authorised ALBA staff for data under embargo. We also have an “Open data” tab where data will be viewable by anyone logged in once they pass the embargo period.

Standardisation

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- F4. (Meta)data are registered or indexed in a searchable resource.
- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data.
- R1.2 (Meta)data are associated with their provenance

8. Which metadata standard(s) does your facility use for data? (select all that apply)

☐	Dublin Core
☐	DataCite
☐	DCAT
☑	NeXus
☐	B2FIND
☐	OpenAIRE (i.e. DataCite with minor adjustments)

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

Do you use any other metadata standards for data?

Not yet (under study).

9. Does your facility use controlled vocabularies for metadata description?

☑ Yes ☐ No

If 'Yes', which one(s) do you use and for which types of metadata (e.g. PaNet ontology for identifying the technique used, list of keywords covering topics or research areas, fixed list of instruments, etc.)?

For the instruments using NeXus, the vocabulary used comes from NeXus application definitions.

10. Does your facility link data/metadata with other relevant data/metadata (e.g. do you link data/metadata from your facility to related experiments, resulting publications, calibration data, etc.)?

☐ Yes ☑ No

If 'Yes', do you do this in a way that standardises and captures these relationships formally (e.g. through the use of a 'related resource' metadata field or similar)?

☐ Yes ☐ No

Any additional comments on how your metadata supports formalised links between data/metadata?

11. Does your facility record and make available information about the provenance of the data offered for reuse (e.g. the experiment or research project with which it originated, the processing it has undergone, any curation actions that have been applied)?

☑ Yes ☐ No

If it does, is the provenance information made available in any specific standard form, either for a particular community (e.g. in NeXus for the PaN community) or cross-domain (e.g. the PROV-O language)?

☐ Yes ☑ No

ALBA

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

If 'Yes', then please provide further details on the specific standard(s) you use for provenance information:

12. What file formats are used to store and process data at your facility?

NeXus / HDF5 in new beamlines (e.g LOREA), various specific formats in others

13. If your facility employs multiple file formats, how do you make these interoperable within and across your facility?

The output file formats used for raw data are known (standard) in the different relevant communities so that users know which software to use in order to explore and process the data. In the case that users are unsure, beamline scientists provide guidance.

More generally, how do you ensure that the file formats you use are interoperable with standard file formats widely used outside of PaN facilities (e.g. .txt, .csv, .pdf, .png, etc.)?

Conversion scripts are provided by the Computing division if the users or beamline scientists need them.

ALBA

Indexing and harvesting of metadata by machines

The questions asked in this section relate specifically to the following FAIR principle:

- F4. (Meta)data are registered or indexed in a searchable resource.

14. If your facility maintains a metadata catalogue, can this metadata be queried and retrieved using an API?

☑ Yes ☐ No

15. If your facility maintains a metadata catalogue, are the metadata standards you use compatible with OAI-PMH?

☐ Yes ☑ No

If 'No', do you maintain a mapping from your metadata to the metadata standards (e.g. Dublin Core, DataCite) commonly used by OAI-PMH?

☐ Yes ☑ No

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

PIDs

The questions asked in this section relate specifically to the following FAIR principles:

- F1. (Meta)data are assigned a globally unique and eternally persistent identifier.
- F3. Metadata specifies the data identifier
- A1. (Meta)data are retrievable by their identifier using a standardised communication protocol

16. What specific persistent identifier service(s) does your facility use? (select all that apply)

☑	Digital Object Identifier (DOI)
☐	Handle
☐	Open Researcher and Contributor ID (ORCID)
☐	Research Organization Registry (ROR)
☐	International Generic Sample Number (IGSN)

Does your facility use any other persistent identifier services?

No

ALBA

17. If your facility has a metadata catalogue, do you include hyperlinked PIDs as part of the metadata provided on landing pages in that catalogue?

☐ Yes ☑ No

18. Do you use any types of PIDs within your facility's data management processes that you don't make available on your metadata catalogue landing pages?

☐ Yes ☑ No

If 'Yes', please provide further details (e.g. we store ORCIDs in our proposal system, but we do not include ORCIDs alongside author names on our catalogue landing pages).

Access to data by users and possibly by machines

The questions asked in this section relate specifically to the following FAIR principles:

- A1. (Meta)data are retrievable by their identifier using a standardised communication protocol
- A1.1 The protocol is open, free, and universally implementable

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

- A1.2 The protocol allows for an authentication and authorization procedure, where necessary
- R1.1 (Meta)data are released with a clear and accessible data usage license

19. How do potential (re)users of your facility's data know how they can access it? (e.g. this information is described on a web page, provided as part of the metadata, etc.)

At the moment, accessibility instructions for users only are given via a web page accessible from within ALBA premises (https://intranet.cells.es/Intranet/Help/BeamlinesHelp/sftp/). No open data is available yet.

20. Are there any authentication and authorisation measures in place to control or monitor who has access to your facility's data?

☑ Yes ☐ No

21. What protocols are employed for accessing the data? Please comment on whether the protocols are open and free (no cost) (e.g. HTTP, FTP), whether they depend on the size or other properties of the dataset, and whether you allow for the possibility of automated access to data through the protocols, as well as human access.

SFTP, HTTPS, (future: Globus or Aspera-> proprietary IBM)

ALBA

22. Does your facility have license agreements on data that it stores and makes available that apply to the users of that data and make clear how they may reuse it?

☑ Yes ☐ No

23. Are the licences you use standard ones (e.g. Creative Commons 4.0, public domain dedication CC0, etc.)?

☑ Yes ☐ No

Please list the types of licenses you apply to data at your facility.

Creative Commons 4.0

24. Are the licences you use machine-understandable as well as readable by humans?

☑ Yes ☐ No

Curation of data

The questions asked in this section relate specifically to the following FAIR principles:

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

- A2. Metadata should be accessible even when the data is no longer available
- R1.2 (Meta)data are associated with their provenance
- R1.3 (Meta)data meet domain-relevant community standards

25. Do you have measures in place at your facility to ensure the integrity of data files against unintentional or unauthorised alteration?

☑ Yes ☐ No

If 'Yes', please provide further details of the measures you have in place.

Reminder: please ensure that you do not include any sensitive or confidential information in your response.

Access to data is regulated by user group permissions. Access to the data catalogue is regulated by LDAP authentication. Backups are in place in case some data is altered by mistake.

26. Is it ever possible that a dataset generated with its metadata during the experimental lifecycle could be removed from the facility's record?

☐ Yes ☑ No

ALBA

If 'Yes', is the associated metadata retained when the dataset is removed?

☐ Yes ☐ No

Please provide any additional comments on the deletion or retention of data/metadata at your facility:

From the beamline, a user would be able to delete a dataset accidentally. The metadata present in the metadata catalogue would however remain.

27. Do you take steps against the possibility of changes over time (e.g. evolution of community standards) that might affect the reusability of the data your facility holds?

☑ Yes ☐ No

If 'Yes', please provide further details of the steps you take.

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

We aim at keeping standard file formats that are Adopting NeXus as a data format, maintaining other formats.

Summary reflections

28. What does your facility take away from the FAIR self-assessment exercise, especially in terms of new insight into the FAIRness of your workflows and data management processes and potential avenues for future development?

This questionnaire is useful for us in the sense that it allows concretely pinpointing which are the FAIR-related aspects to improve and provides clear directions to follow in order to achieve better FAIR data practices at our facility.

This synthesis also allows us to reflect on the concepts and tools developed in ExPaNDS WP2. In particular, certain aspects of the metadata framework might need development in the future like for instance, reusability-wise, agreeing on what is the minimal information to collect regarding samples and instruments.

29. Please provide feedback and suggestions for improvement of the approach used for the FAIR self-assessment (e.g. questions asked, facility coordinator role, workshops, explanatory notes provided, links to existing FAIR evaluation models, focus on workflows and processes, etc.).

NOTE: please ensure you do not include any personal identifying information (e.g. names) in your response.

While we understand that the questions aimed at capturing an “instant snapshot” of the facility’s state regarding FAIR, some of them might have deserved a text-box allowing to explain why “yes” or “no” was ticked, so as to know whether a particular FAIR-related process was undergoing or not envisaged at all by the facility.

FAIR-related aspects regarding raw and processed data could have been treated in separate questions since they relate to different time-points in the PaN-data ODI Data Continuum model (that underpins the WP2 FAIR metadata framework).

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

B1.2 DESY

Comments on context

0. If you wish, please explain how you have chosen to handle references in the questionnaire to "data" (for example, what type[s] of data are covered by your responses), and/or the situation of multiple beamlines/instruments with different practices (for example, by focusing on only one), and any other context that will help to interpret your responses.

For the following questionnaire, we took the beamlines and experimental setups of DESY photon science into consideration, in particular PETRA III.
As “data”, we consider all data created during experiments or processed post-beamtime.
Currently there is no metadata catalogue at DESY (work in progress). The metadata that is currently available is part of DESY’s DOOR (DESY Online Office for Research with Photons) and Gamma portal, where the data of all experiments performed at DESY can be overviewed.

Existence, completeness, and richness of metadata

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

1. Is metadata associated with the data that are created/collected across the experimental lifecycle at your facility?

☑ Yes ☐ No

2. Looking at the metadata listed as essential in the ExPaNDS metadata framework across all stages of the experimental lifecycle (see Section 6.3, pages 54-56, Essential metadata to capture in each stage of the experimental lifecycle), is all of this metadata captured (i.e. either automatically or manually) at your facility?

☐ Yes ☑ No

If there are any essential metadata that you do not capture, what are they?

Most of the ones in the Proposal Section (1-7) are captured, as are some in the Experiment section (9,12).
All others are not systematically captured.

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

3. When data from your facility is accessed, do you provide any metadata that is specifically intended to aid the reuse of that data, i.e. in distinction from metadata for discovery of the data (e.g. do you provide metadata that gives contextual information about how the data was generated)?

☐ Yes ☑ No

If so, what is the metadata that you provide to aid reuse?

Search (flexibility and capability)

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- F4. (Meta)data are registered or indexed in a searchable resource

4. Is it possible to search metadata related to data from the experimental lifecycle at your facility?

☐ Yes ☑ No

5. Does the metadata enable basic discovery (e.g. does the metadata include bibliographic information such as author, title, date, etc.)?

☐ Yes ☑ No

Can you make multi-faceted, PaN-specific queries (e.g. technique, experimental parameters, instrument, sample)?

☐ Yes ☑ No

Any additional comments on how your metadata enables discovery of data via search?

Users being part of the proposal, can explore metadata of the experiments via free text search.

6. How is the metadata searchable? (select all that apply)

☑	Free text search
☐	Controlled vocabulary

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

☐	Filters (e.g. date, topic, instrument, etc.)

Are there any other ways of searching your metadata?

7. Who can search the metadata (e.g. searching restricted in some way, searchable by anyone)?

Free text search is restricted to users being part of the experiment only.

Standardisation

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- F4. (Meta)data are registered or indexed in a searchable resource.
- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data.
- R1.2 (Meta)data are associated with their provenance

8. Which metadata standard(s) does your facility use for data? (select all that apply)

☐	Dublin Core
☐	DataCite
☐	DCAT
☑	NeXus
☐	B2FIND
☐	OpenAIRE (i.e. DataCite with minor adjustments)

Do you use any other metadata standards for data?

No

9. Does your facility use controlled vocabularies for metadata description?

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

☐ Yes ☑ No

If 'Yes', which one(s) do you use and for which types of metadata (e.g. PaNet ontology for identifying the technique used, list of keywords covering topics or research areas, fixed list of instruments, etc.)?

10. Does your facility link data/metadata with other relevant data/metadata (e.g. do you link data/metadata from your facility to related experiments, resulting publications, calibration data, etc.)?

☐ Yes ☑ No

If 'Yes', do you do this in a way that standardises and captures these relationships formally (e.g. through the use of a 'related resource' metadata field or similar)?

☐ Yes ☐ No

Any additional comments on how your metadata supports formalised links between data/metadata?

11. Does your facility record and make available information about the provenance of the data offered for reuse (e.g. the experiment or research project with which it originated, the processing it has undergone, any curation actions that have been applied)?

☐ Yes ☑ No

If it does, is the provenance information made available in any specific standard form, either for a particular community (e.g. in NeXus for the PaN community) or cross-domain (e.g. the PROV-O language)?

☐ Yes ☐ No

If 'Yes', then please provide further details on the specific standard(s) you use for provenance information:

12. What file formats are used to store and process data at your facility?

Most prominent file formats used are: hdf5, tiff, cbf and ascii

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

13. If your facility employs multiple file formats, how do you make these interoperable within and across your facility?

Partial format conversions are available on request (within limits).

More generally, how do you ensure that the file formats you use are interoperable with standard file formats widely used outside of PaN facilities (e.g. .txt, .csv, .pdf, .png, etc.)?

At the moment, there is no general procedure for that (work in progress).

Indexing and harvesting of metadata by machines

The questions asked in this section relate specifically to the following FAIR principle:

- F4. (Meta)data are registered or indexed in a searchable resource.

14. If your facility maintains a metadata catalogue, can this metadata be queried and retrieved using an API?

☐ Yes ☑ No

15. If your facility maintains a metadata catalogue, are the metadata standards you use compatible with OAI-PMH?

☐ Yes ☑ No

If 'No', do you maintain a mapping from your metadata to the metadata standards (e.g. Dublin Core, DataCite) commonly used by OAI-PMH?

☐ Yes ☑ No

PIDs

The questions asked in this section relate specifically to the following FAIR principles:

- F1. (Meta)data are assigned a globally unique and eternally persistent identifier.
- F3. Metadata specifies the data identifier
- A1. (Meta)data are retrievable by their identifier using a standardised communication protocol

16. What specific persistent identifier service(s) does your facility use? (select all that apply)

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

☑	Digital Object Identifier (DOI)
☐	Handle
☑	Open Researcher and Contributor ID (ORCID)
☐	Research Organization Registry (ROR)
☑	International Generic Sample Number (IGSN)

Does your facility use any other persistent identifier services?

No

17. If your facility has a metadata catalogue, do you include hyperlinked PIDs as part of the metadata provided on landing pages in that catalogue?

☐ Yes ☐ No

18. Do you use any types of PIDs within your facility's data management processes that you don't make available on your metadata catalogue landing pages?

☐ Yes ☑ No

If 'Yes', please provide further details (e.g. we store ORCIDs in our proposal system, but we do not include ORCIDs alongside author names on our catalogue landing pages).

Access to data by users and possibly by machines

The questions asked in this section relate specifically to the following FAIR principles:

- A1. (Meta)data are retrievable by their identifier using a standardised communication protocol
- A1.1 The protocol is open, free, and universally implementable
- A1.2 The protocol allows for an authentication and authorization procedure, where necessary
- R1.1 (Meta)data are released with a clear and accessible data usage license

19. How do potential (re)users of your facility's data know how they can access it? (e.g. this information is described on a web page, provided as part of the metadata, etc.)

Online description for data download and getting access rights is available on a public webpage. Same is the case in regard to accessing the data within the DESY computing facility.

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

20. Are there any authentication and authorisation measures in place to control or monitor who has access to your facility's data?

☑ Yes ☐ No

21. What protocols are employed for accessing the data? Please comment on whether the protocols are open and free (no cost) (e.g. HTTP, FTP), whether they depend on the size or other properties of the dataset, and whether you allow for the possibility of automated access to data through the protocols, as well as human access.

Data download:
- FTP with explicit TLS encryption scheme via DOOR account.
- rsync or globus service via DESY scientific account

Access within DESY computing facility (via DESY scientific account):
- Native GPFS, SMB, NFS4 client (partially Kerberos)

22. Does your facility have license agreements on data that it stores and makes available that apply to the users of that data and make clear how they may reuse it?

☐ Yes ☑ No

23. Are the licences you use standard ones (e.g. Creative Commons 4.0, public domain dedication CC0, etc.)?

☐ Yes ☑ No

Please list the types of licenses you apply to data at your facility.

Project Leader/PI of a beamtime are responsible for the data. Only beamtime participants have initial access.

24. Are the licences you use machine-understandable as well as readable by humans?

☐ Yes ☑ No

Curation of data

The questions asked in this section relate specifically to the following FAIR principles:

- A2. Metadata should be accessible even when the data is no longer available
- R1.2 (Meta)data are associated with their provenance
- R1.3 (Meta)data meet domain-relevant community standards

25. Do you have measures in place at your facility to ensure the integrity of data files against unintentional or unauthorised alteration?

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

☑ Yes ☐ No

If 'Yes', please provide further details of the measures you have in place.

Reminder: please ensure that you do not include any sensitive or confidential information in your response.

Access rights to all data are managed by access control lists (ACLs). Tape copies of data are created after a few days after Beamtime and cannot be modified afterwards. Concepts of snapshots are used to cope accidentally deleted/modified data.

26. Is it ever possible that a dataset generated with its metadata during the experimental lifecycle could be removed from the facility's record?

☑ Yes ☐ No

If 'Yes', is the associated metadata retained when the dataset is removed?

☐ Yes ☑ No

Please provide any additional comments on the deletion or retention of data/metadata at your facility:

27. Do you take steps against the possibility of changes over time (e.g. evolution of community standards) that might affect the reusability of the data your facility holds?

☐ Yes ☑ No

If 'Yes', please provide further details of the steps you take.

Summary reflections

28. What does your facility take away from the FAIR self-assessment exercise, especially in terms of new insight into the FAIRness of your workflows and data management processes and potential avenues for future development?

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

29. Please provide feedback and suggestions for improvement of the approach used for the FAIR self-assessment (e.g. questions asked, facility coordinator role, workshops, explanatory notes provided, links to existing FAIR evaluation models, focus on workflows and processes, etc.).

NOTE: please ensure you do not include any personal identifying information (e.g. names) in your response.

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

B1.3 DLS

Comments on context

0. If you wish, please explain how you have chosen to handle references in the questionnaire to "data" (for example, what type[s] of data are covered by your responses), and/or the situation of multiple beamlines/instruments with different practices (for example, by focusing on only one), and any other context that will help to interpret your responses.

We have answered the questions as a current snapshot, and mainly for the general open data consumer rather than the data-owning facility users.

Existence, completeness, and richness of metadata

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

1. Is metadata associated with the data that are created/collected across the experimental lifecycle at your facility?

☑ Yes ☐ No

2. Looking at the metadata listed as essential in the ExPaNDS metadata framework across all stages of the experimental lifecycle (see Section 6.3, pages 54-56, Essential metadata to capture in each stage of the experimental lifecycle), is all of this metadata captured (i.e. either automatically or manually) at your facility?

☐ Yes ☑ No

If there are any essential metadata that you do not capture, what are they?

Meaningful metadata relating to sample description are largely absent, with sample names being largely used as local identifiers.

3. When data from your facility is accessed, do you provide any metadata that is specifically intended to aid the reuse of that data, i.e. in distinction from metadata for discovery of the data (e.g. do you provide metadata that gives contextual information about how the data was generated)?

☑ Yes ☐ No

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

If so, what is the metadata that you provide to aid reuse?

To some extent this is covered by the NeXus metadata.

Search (flexibility and capability)

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- F4. (Meta)data are registered or indexed in a searchable resource

4. Is it possible to search metadata related to data from the experimental lifecycle at your facility?

☑ Yes ☐ No

5. Does the metadata enable basic discovery (e.g. does the metadata include bibliographic information such as author, title, date, etc.)?

☑ Yes ☐ No

Can you make multi-faceted, PaN-specific queries (e.g. technique, experimental parameters, instrument, sample)?

☑ Yes ☐ No

Any additional comments on how your metadata enables discovery of data via search?

6. How is the metadata searchable? (select all that apply)

☑	Free text search
☐	Controlled vocabulary
☑	Filters (e.g. date, topic, instrument, etc.)

Are there any other ways of searching your metadata?

DLS

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

7. Who can search the metadata (e.g. searching restricted in some way, searchable by anyone)?

Search currently restricted to the data-owning team members.

Standardisation

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- F4. (Meta)data are registered or indexed in a searchable resource.
- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data.
- R1.2 (Meta)data are associated with their provenance

8. Which metadata standard(s) does your facility use for data? (select all that apply)

☐	Dublin Core
☐	DataCite
☐	DCAT
☑	NeXus
☐	B2FIND
☐	OpenAIRE (i.e. DataCite with minor adjustments)

Do you use any other metadata standards for data?

Core Scientific Metadata Model (CSMD).

9. Does your facility use controlled vocabularies for metadata description?

☑ Yes ☐ No

If ‘Yes’, which one(s) do you use and for which types of metadata (e.g. PaNet ontology for identifying the technique used, list of keywords covering topics or research areas, fixed list of instruments, etc.)?

NeXus

DLS

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

10. Does your facility link data/metadata with other relevant data/metadata (e.g. do you link data/metadata from your facility to related experiments, resulting publications, calibration data, etc.)?

☑ Yes ☐ No

If 'Yes', do you do this in a way that standardises and captures these relationships formally (e.g. through the use of a 'related resource' metadata field or similar)?

☑ Yes ☐ No

Any additional comments on how your metadata supports formalised links between data/metadata?

Use of NeXus links, for example.

11. Does your facility record and make available information about the provenance of the data offered for reuse (e.g. the experiment or research project with which it originated, the processing it has undergone, any curation actions that have been applied)?

☐ Yes ☑ No

If it does, is the provenance information made available in any specific standard form, either for a particular community (e.g. in NeXus for the PaN community) or cross-domain (e.g. the PROV-O language)?

☐ Yes ☐ No

If 'Yes', then please provide further details on the specific standard(s) you use for provenance information:

12. What file formats are used to store and process data at your facility?

NeXus (preferred) plus CBF, ascii, TIFF etc.

13. If your facility employs multiple file formats, how do you make these interoperable within and across your facility?

File conversions, mainly from NeXus.

DLS

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

More generally, how do you ensure that the file formats you use are interoperable with standard file formats widely used outside of PaN facilities (e.g. .txt, .csv, .pdf, .png, etc.)?

Standard file conversions.

Indexing and harvesting of metadata by machines

The questions asked in this section relate specifically to the following FAIR principle:

- F4. (Meta)data are registered or indexed in a searchable resource.

14. If your facility maintains a metadata catalogue, can this metadata be queried and retrieved using an API?

☑ Yes ☑ No

15. If your facility maintains a metadata catalogue, are the metadata standards you use compatible with OAI-PMH?

☐ Yes ☑ No

If 'No', do you maintain a mapping from your metadata to the metadata standards (e.g. Dublin Core, DataCite) commonly used by OAI-PMH?

☑ Yes ☐ No

PIDs

The questions asked in this section relate specifically to the following FAIR principles:

- F1. (Meta)data are assigned a globally unique and eternally persistent identifier.
- F3. Metadata specifies the data identifier
- A1. (Meta)data are retrievable by their identifier using a standardised communication protocol

16. What specific persistent identifier service(s) does your facility use? (select all that apply)

☐	Digital Object Identifier (DOI)
☐	Handle
☑	Open Researcher and Contributor ID (ORCID)
☐	Research Organization Registry (ROR)

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

☐	International Generic Sample Number (IGSN)

Does your facility use any other persistent identifier services?

No

17. If your facility has a metadata catalogue, do you include hyperlinked PIDs as part of the metadata provided on landing pages in that catalogue?

☐ Yes ☑ No

18. Do you use any types of PIDs within your facility's data management processes that you don't make available on your metadata catalogue landing pages?

☑ Yes ☐ No

If 'Yes', please provide further details (e.g. we store ORCIDs in our proposal system, but we do not include ORCIDs alongside author names on our catalogue landing pages).

ORCID IDs Diamond will have DOIs once data are open.

DLS

Access to data by users and possibly by machines

The questions asked in this section relate specifically to the following FAIR principles:

- A1. (Meta)data are retrievable by their identifier using a standardised communication protocol
- A1.1 The protocol is open, free, and universally implementable
- A1.2 The protocol allows for an authentication and authorization procedure, where necessary
- R1.1 (Meta)data are released with a clear and accessible data usage license

19. How do potential (re)users of your facility's data know how they can access it? (e.g. this information is described on a web page, provided as part of the metadata, etc.)

Described on web.

20. Are there any authentication and authorisation measures in place to control or monitor who has access to your facility's data?

☑ Yes ☐ No

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

21. What protocols are employed for accessing the data? Please comment on whether the protocols are open and free (no cost) (e.g. HTTP, FTP), whether they depend on the size or other properties of the dataset, and whether you allow for the possibility of automated access to data through the protocols, as well as human access.

HTTP, FTP, Globus

22. Does your facility have license agreements on data that it stores and makes available that apply to the users of that data and make clear how they may reuse it?

☑ Yes ☐ No

23. Are the licences you use standard ones (e.g. Creative Commons 4.0, public domain dedication CC0, etc.)?

☑ Yes ☐ No

Please list the types of licenses you apply to data at your facility.

CC-BY-4.0

24. Are the licences you use machine-understandable as well as readable by humans?

☑ Yes ☐ No

Curation of data

The questions asked in this section relate specifically to the following FAIR principles:

- A2. Metadata should be accessible even when the data is no longer available
- R1.2 (Meta)data are associated with their provenance
- R1.3 (Meta)data meet domain-relevant community standards

25. Do you have measures in place at your facility to ensure the integrity of data files against unintentional or unauthorised alteration?

☑ Yes ☐ No

If ‘Yes’, please provide further details of the measures you have in place.

Reminder: please ensure that you do not include any sensitive or confidential information in your response.

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

Data collected prior to archive by authorized users only. Data archived to tape are immutable.

26. Is it ever possible that a dataset generated with its metadata during the experimental lifecycle could be removed from the facility's record?

☐ Yes ☑ No

If 'Yes', is the associated metadata retained when the dataset is removed?

☐ Yes ☐ No

Please provide any additional comments on the deletion or retention of data/metadata at your facility:

Removal of data is technically possible but requires a manual process.

27. Do you take steps against the possibility of changes over time (e.g. evolution of community standards) that might affect the reusability of the data your facility holds?

☐ Yes ☑ No

If 'Yes', please provide further details of the steps you take.

Summary reflections

28. What does your facility take away from the FAIR self-assessment exercise, especially in terms of new insight into the FAIRness of your workflows and data management processes and potential avenues for future development?

While Diamond has FAIR aspirations, projects in development, and a data policy that allows open data access, it is currently at a very early stage of its FAIR journey.

29. Please provide feedback and suggestions for improvement of the approach used for the FAIR self-assessment (e.g. questions asked, facility coordinator role, workshops, explanatory notes provided, links to existing FAIR evaluation models, focus on workflows and processes, etc.).

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

NOTE: please ensure you do not include any personal identifying information (e.g. names) in your response.

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

B1.4 Elettra

Comments on context

0. If you wish, please explain how you have chosen to handle references in the questionnaire to "data" (for example, what type[s] of data are covered by your responses), and/or the situation of multiple beamlines/instruments with different practices (for example, by focusing on only one), and any other context that will help to interpret your responses.

Existence, completeness, and richness of metadata

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

1. Is metadata associated with the data that are created/collected across the experimental lifecycle at your facility?

☑ Yes ☐ No

2. Looking at the metadata listed as essential in the ExPaNDS metadata framework across all stages of the experimental lifecycle (see Section 6.3, pages 54-56, Essential metadata to capture in each stage of the experimental lifecycle), is all of this metadata captured (i.e. either automatically or manually) at your facility?

☑ Yes ☐ No

If there are any essential metadata that you do not capture, what are they?

In most cases, all 8 essential metadata are captured but in some cases “Proposed Experiment Conditions” may be missing as they are not always relevant to the experiment.

3. When data from your facility is accessed, do you provide any metadata that is specifically intended to aid the reuse of that data, i.e. in distinction from metadata for discovery of the data (e.g. do you provide metadata that gives contextual information about how the data was generated)?

☑ Yes ☐ No

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

If so, what is the metadata that you provide to aid reuse?

Such metadata are often stored in the form of instruction in electronic logbooks associated with the experiment. All of the logbooks are accessible.

Search (flexibility and capability)

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- F4. (Meta)data are registered or indexed in a searchable resource

4. Is it possible to search metadata related to data from the experimental lifecycle at your facility?

☑ Yes ☐ No

5. Does the metadata enable basic discovery (e.g. does the metadata include bibliographic information such as author, title, date, etc.)?

☑ Yes ☐ No

Can you make multi-faceted, PaN-specific queries (e.g. technique, experimental parameters, instrument, sample)?

☐ Yes ☑ No

Any additional comments on how your metadata enables discovery of data via search?

The most popular metadata used in discovery/queries are those associated with the Investigators' team (PIs and co-investigators)

6. How is the metadata searchable? (select all that apply)

☑	Free text search
☐	Controlled vocabulary
☐	Filters (e.g. date, topic, instrument, etc.)

Are there any other ways of searching your metadata?

There are ongoing activities that very soon will enable controlled vocabularies and filters besides the existing free text.

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

7. Who can search the metadata (e.g. searching restricted in some way, searchable by anyone)?

The access is restricted to VUO registered users (the facilities main portal - vuo.elettra.eu). Registration is free.

Standardisation

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- F4. (Meta)data are registered or indexed in a searchable resource.
- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data.
- R1.2 (Meta)data are associated with their provenance

8. Which metadata standard(s) does your facility use for data? (select all that apply)

☐	Dublin Core
☑	DataCite
☐	DCAT
☑	NeXus
☐	B2FIND
☐	OpenAIRE (i.e. DataCite with minor adjustments)

Do you use any other metadata standards for data?

Besides the use of DataCite and NeXus, there are some internal/legacy superset standards that are in use for certain applications.

9. Does your facility use controlled vocabularies for metadata description?

☑ Yes ☐ No

If ‘Yes’, which one(s) do you use and for which types of metadata (e.g. PaNet ontology for identifying the technique used, list of keywords covering topics or research areas, fixed list of instruments, etc.)?

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

Recently the PaNet ontology is examined and used for inspiration. Previous systems dictate ontologies that have been created by the internal controls and data acquisition teams. The data acquisition system is common among many beamlines thus it was the starting point for ontology definition.

10. Does your facility link data/metadata with other relevant data/metadata (e.g. do you link data/metadata from your facility to related experiments, resulting publications, calibration data, etc.)?

☑ Yes ☐ No

If 'Yes', do you do this in a way that standardises and captures these relationships formally (e.g. through the use of a 'related resource' metadata field or similar)?

☑ Yes ☐ No

Any additional comments on how your metadata supports formalised links between data/metadata?

A beta system associates DOIs of publications with DOIs of datasets. Since the publications are formally associated with the Proposal numbers and thus with lots of metadata, any subsequent derived dataset (i.e. for a publication) is associated with the original proposal.

11. Does your facility record and make available information about the provenance of the data offered for reuse (e.g. the experiment or research project with which it originated, the processing it has undergone, any curation actions that have been applied)?

☐ Yes ☑ No

If it does, is the provenance information made available in any specific standard form, either for a particular community (e.g. in NeXus for the PaN community) or cross-domain (e.g. the PROV-O language)?

☐ Yes ☑ No

If 'Yes', then please provide further details on the specific standard(s) you use for provenance information:

12. What file formats are used to store and process data at your facility?

Mostly HDF5 but also RAW, TIFF, text, and blobs in RDBMS

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

13. If your facility employs multiple file formats, how do you make these interoperable within and across your facility?

Automatic converters but also ad-hoc requests for data conversion especially HDF5-to-HDF5 of different structures. The most popular one is a converter of generic “list-mode” generated by the control -to- application specific h5/nexus with datasets of correct dimensions enriched with metadata.

More generally, how do you ensure that the file formats you use are interoperable with standard file formats widely used outside of PaN facilities (e.g. .txt, .csv, .pdf, .png, etc.)?

The as above; automatic converters but also ad-hoc requests for data conversion especially HDF5-to-HDF5 of different structures.

Indexing and harvesting of metadata by machines

The questions asked in this section relate specifically to the following FAIR principle:

- F4. (Meta)data are registered or indexed in a searchable resource.

14. If your facility maintains a metadata catalogue, can this metadata be queried and retrieved using an API?

☑ Yes ☐ No

15. If your facility maintains a metadata catalogue, are the metadata standards you use compatible with OAI-PMH?

☐ Yes ☑ No

If ‘No’, do you maintain a mapping from your metadata to the metadata standards (e.g. Dublin Core, DataCite) commonly used by OAI-PMH?

☑ Yes ☐ No

PIDs

The questions asked in this section relate specifically to the following FAIR principles:

- F1. (Meta)data are assigned a globally unique and eternally persistent identifier.
- F3. Metadata specifies the data identifier
- A1. (Meta)data are retrievable by their identifier using a standardised communication protocol

16. What specific persistent identifier service(s) does your facility use? (select all that apply)

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

☑	Digital Object Identifier (DOI)
☐	Handle
☑	Open Researcher and Contributor ID (ORCID)
☐	Research Organization Registry (ROR)
☐	International Generic Sample Number (IGSN)

Does your facility use any other persistent identifier services?

Over the past years there are experiments and beta deployment of IPFS with original contributions on PIDs. There is also a Python interface now available for handling IPFS CIDs as PID.

17. If your facility has a metadata catalogue, do you include hyperlinked PIDs as part of the metadata provided on landing pages in that catalogue?

☑ Yes ☐ No

18. Do you use any types of PIDs within your facility's data management processes that you don't make available on your metadata catalogue landing pages?

☑ Yes ☐ No

If 'Yes', please provide further details (e.g. we store ORCIDs in our proposal system, but we do not include ORCIDs alongside author names on our catalogue landing pages).

There is more data stored than what is made available. Emails and ORCIDs are among them.

Access to data by users and possibly by machines

The questions asked in this section relate specifically to the following FAIR principles:

- A1. (Meta)data are retrievable by their identifier using a standardised communication protocol
- A1.1 The protocol is open, free, and universally implementable
- A1.2 The protocol allows for an authentication and authorization procedure, where necessary
- R1.1 (Meta)data are released with a clear and accessible data usage license

19. How do potential (re)users of your facility's data know how they can access it? (e.g. this information is described on a web page, provided as part of the metadata, etc.)

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

The instructions are provided in the main portal (VUO) along with basic concepts of FAIR, openness and the facility's data policy.

20. Are there any authentication and authorisation measures in place to control or monitor who has access to your facility's data?

☑ Yes ☐ No

21. What protocols are employed for accessing the data? Please comment on whether the protocols are open and free (no cost) (e.g. HTTP, FTP), whether they depend on the size or other properties of the dataset, and whether you allow for the possibility of automated access to data through the protocols, as well as human access.

For now it is all free without quotas. In the future, due to sustainability, certain limits may be implemented.

22. Does your facility have license agreements on data that it stores and makes available that apply to the users of that data and make clear how they may reuse it?

☐ Yes ☑ No

23. Are the licences you use standard ones (e.g. Creative Commons 4.0, public domain dedication CC0, etc.)?

☑ Yes ☐ No

Please list the types of licenses you apply to data at your facility.

CC4, MIT, GPL3

24. Are the licences you use machine-understandable as well as readable by humans?

☑ Yes ☐ No

Curation of data

The questions asked in this section relate specifically to the following FAIR principles:

- A2. Metadata should be accessible even when the data is no longer available
- R1.2 (Meta)data are associated with their provenance
- R1.3 (Meta)data meet domain-relevant community standards

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

25. Do you have measures in place at your facility to ensure the integrity of data files against unintentional or unauthorised alteration?

☑ Yes ☐ No

If 'Yes', please provide further details of the measures you have in place.

Reminder: please ensure that you do not include any sensitive or confidential information in your response.

All data stored and published have two distinct classes one of which is strictly read-only after its creation. Not even the PI can edit this through the portal (it can be done only through a special written request). There is a not-read-only part that can be edited by the experiment team.

26. Is it ever possible that a dataset generated with its metadata during the experimental lifecycle could be removed from the facility's record?

☑ Yes ☐ No

If 'Yes', is the associated metadata retained when the dataset is removed?

☐ Yes ☑ No

Please provide any additional comments on the deletion or retention of data/metadata at your facility:

For now and while the resources permit it, no dataset+metadata has been removed. If this will be necessary in the future due to storage constraints, non-essential metadata stored inside data files may be cancelled while any metadata file and metadata stored is DBs/catalogues will be retained.

27. Do you take steps against the possibility of changes over time (e.g. evolution of community standards) that might affect the reusability of the data your facility holds?

☑ Yes ☐ No

If 'Yes', please provide further details of the steps you take.

There are organised activities and ongoing work to ensure the evolution of the initiative. Participation in workshops but also future projects can be of importance.

Elettra

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

Summary reflections

28. What does your facility take away from the FAIR self-assessment exercise, especially in terms of new insight into the FAIRness of your workflows and data management processes and potential avenues for future development?

The FAIR self-assessment has been an important and useful exercise. It helps to highlight what may be missing but also to appreciate the amount of work and achievements of the past years. The overall state is very satisfactory but there are concerns about sustainability and continuation of standard adhesion beyond the end of the ongoing projects. This may be true especially for systems that keep evolving and have not reached maturity and stability yet. Abuse of openness (see data access), security and accountability are also important issues. As the weak point here the reusability constraints should be considered, especially due to software installation and versioning problems. Improvements in this aspect seems to be one of the most important directions of development of FAIR data and future projects.

29. Please provide feedback and suggestions for improvement of the approach used for the FAIR self-assessment (e.g. questions asked, facility coordinator role, workshops, explanatory notes provided, links to existing FAIR evaluation models, focus on workflows and processes, etc.).

NOTE: please ensure you do not include any personal identifying information (e.g. names) in your response.

Recent internal discussions brought up two topics that may not be strictly related to the self-assessment but popped up during filling it out; the first regards data that are not stored in files but completely in DBs. The whole initiative felt too file-oriented but future operations (see AI/ML) may be less file-oriented. The second regards, the role of publications (as in published results) and the storage of the derived/processed sub-datasets (often ending up in Zenodo).

Elettra

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

B1.5 HZB

Comments on context

0. If you wish, please explain how you have chosen to handle references in the questionnaire to "data" (for example, what type[s] of data are covered by your responses), and/or the situation of multiple beamlines/instruments with different practices (for example, by focusing on only one), and any other context that will help to interpret your responses.

We need to consider different kind of data, mainly:

- raw data collected during measurements at HZB instruments,
- curated data publications.

Raw data are generated at large numbers of datasets. Only the metadata collected automatically by the experiment control software is available. No manual curation of these metadata is possible. As a result, we essentially don't have any bibliographic metadata for raw data. There is however the link to the proposal data, but that describes the proposal, not the raw data.

The generation of the raw data may comprise some data processing and even data reduction. If that is done routinely as part of the experimental workflow, the result is still considered raw data according to the HZB Data Policy.

We don't have any particular support for processed data beyond the experimental workflow. If the data processing is done by the user, the curation is entirely in the user's responsibility. So we don't consider processed data in the following.

We provide a separate workflow for the creation of data publications. Data publications are typically done in small numbers for carefully selected data. Usually a new dedicated dataset is created for the publication. The bibliographic metadata are manually curated and of high quality.

The HZB Data Policy is not yet implemented at all instruments. We will consider the raw data from a few example instruments in the following. Since there are at the moment important steps underway, in particular concerning the adoption of PIDs for raw data, we describe in the following the state of the implementation as anticipated as of November 2022.

Existence, completeness, and richness of metadata

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

1. Is metadata associated with the data that are created/collected across the experimental lifecycle at your facility?

☑ Yes ☐ No

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

2. Looking at the metadata listed as essential in the ExPaNDS metadata framework across all stages of the experimental lifecycle (see Section 6.3, pages 54-56, Essential metadata to capture in each stage of the experimental lifecycle), is all of this metadata captured (i.e. either automatically or manually) at your facility?

☐ Yes ☑ No

If there are any essential metadata that you do not capture, what are they?

Proposal
1. Is the Principal investigator declared as part of the metadata fields?
yes (the proposal system uses the term “Proposer”)
2. Are the Co-Investigators declared as part of the metadata fields?
yes (the proposal system uses the term “Co-Proposer”)
3. Is the Instrument requested declared as part of the metadata fields?
yes
4. Is the Sample description declared as part of the metadata fields?
A sample description is provided with the proposal. But that information is not transferred to the metadata of the raw data.
5. Is the Facility where the proposal is submitted declared as part of the metadata fields?
yes
6. Is the Proposal Identifier declared as part of the metadata fields?
yes
7. Is the Experiment Description declared as part of the metadata fields?
yes
8. Are the Proposed Experiment Conditions declared as part of the metadata fields?
The planned experiment is described in detail in the proposal, but this description is not transferred to the metadata of the raw data.

Experiment
9. Is the actual Visiting Experimental Team (people who actually participate during the measurement) declared as part of the metadata fields?
yes
10. Are the Experiment/Measurement dates declared as part of the metadata fields?
Normally yes - there can be exceptions.
11. Does the Samples information provide enough context to understand its structure and characteristics and is declared as part of the metadata fields?
There might be a link to a sample database, but normally it happens that there is not enough information to retrieve the sample provenance
12. Is the Instrument information declared as part of the metadata fields?
Yes
13. Is the Calibration information declared as part of the metadata fields?
Not always
14. Is the produced Dataset information declared as part of the metadata fields?
Yes

HZB

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

Processing (there are normally no processed data in the data catalogue)
15. Is the resulting Data Format declared as part of the metadata fields?
16. Is the Processing information declared as part of the metadata fields?
17. Is the Software package information used for processing declared as part of the metadata fields?
18. Is the Original Data link used for the processing declared as part of the metadata fields?
19. Is the resulting Dataset information declared as part of the metadata fields?

Analysis (there are normally no analysed data in the data catalogue)
20. Is the resulting Data Format of the Analysis declared as part of the metadata fields?
21. Are the Files Identifiers declared as part of the metadata fields?
22. Is the Software package used for the analysis declared as part of the metadata fields?
23. Is the Original Data link used for the analysis declared as part of the metadata fields?
24. Is the resulting Dataset information declared as part of the metadata fields?
Data Publication Record
25. Is the Resource Identity declared as part of the metadata fields?
yes
26. Is the Creator of the record declared as part of the metadata fields?
yes (for data publications; for raw data, the Creator field is left blank, but Proposer, Co-Proposer and Experimentators are listed as Contributors)
27. Is the Publisher of the record declared as part of the metadata fields?
yes
28. Is the Publication year declared as part of the metadata fields?
yes
29. Is the Release date (Embargo due date) declared as part of the metadata fields?
yes
30. Is the Title of the dataset declared as part of the metadata fields?
yes
31. Is the License for usage declared as part of the metadata fields?
yes

HZB

3. When data from your facility is accessed, do you provide any metadata that is specifically intended to aid the reuse of that data, i.e. in distinction from metadata for discovery of the data (e.g. do you provide metadata that gives contextual information about how the data was generated)?

☑ Yes ☐ No

If so, what is the metadata that you provide to aid reuse?

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

Most instruments produce NeXus as the data format for the raw data. The experiment conditions and physical parameters are captured as accurately as possible and stored in the files according to the NeXus standard.

Search (flexibility and capability)

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- F4. (Meta)data are registered or indexed in a searchable resource

4. Is it possible to search metadata related to data from the experimental lifecycle at your facility?

☑ Yes ☐ No

5. Does the metadata enable basic discovery (e.g. does the metadata include bibliographic information such as author, title, date, etc.)?

☑ Yes (Data publications) ☑ No (raw data)

Can you make multi-faceted, PaN-specific queries (e.g. technique, experimental parameters, instrument, sample)?

☑ Yes ☑ No

Any additional comments on how your metadata enables discovery of data via search?

Yes, but limited. There is a text search and a search on experimental parameters. We do not record sample information in the metadata catalogue yet.

6. How is the metadata searchable? (select all that apply)

☑	Free text search
☑	Controlled vocabulary
☑	Filters (e.g. date, topic, instrument, etc.)Proposal

Are there any other ways of searching your metadata?

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

7. Who can search the metadata (e.g. searching restricted in some way, searchable by anyone)?

Open access datasets: everyone Data under embargo: only the experimental team

Standardisation

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- F4. (Meta)data are registered or indexed in a searchable resource.
- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data.
- R1.2 (Meta)data are associated with their provenance

8. Which metadata standard(s) does your facility use for data? (select all that apply)

☑	Dublin Core
☑	DataCite
☐	DCAT
☑	NeXus
☐	B2FIND
☐	OpenAIRE (i.e. DataCite with minor adjustments)

Do you use any other metadata standards for data?

schema.org for data publications

9. Does your facility use controlled vocabularies for metadata description?

☑ Yes ☐ No

If 'Yes', which one(s) do you use and for which types of metadata (e.g. PaNet ontology for identifying the technique used, list of keywords covering topics or research areas, fixed list of instruments, etc.)?

HZB

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

The ICAT schema enforces the use of controlled vocabularies by database relations to the corresponding objects in many places: datasets are are linked with the experimental technique and thus to the term in PaNET, datasets are linked with the instrument, there is a global table parameter types that is used for the physical parameters linked to a dataset. We use a naming scheme inspired by the NeXus standard to populate that table of parameter types.

10. Does your facility link data/metadata with other relevant data/metadata (e.g. do you link data/metadata from your facility to related experiments, resulting publications, calibration data, etc.)?

☑ Yes (Data publications) ☑ No (raw data)

If 'Yes', do you do this in a way that standardises and captures these relationships formally (e.g. through the use of a 'related resource' metadata field or similar)?

☑ Yes ☐ No

Any additional comments on how your metadata supports formalised links between data/metadata?

Using the DataCite metadata schema: related resources are linked with RelatedIdentifier and RelatedItem. PaNET terms are provided as Subject/subjectScheme/valueURI

HZB

11. Does your facility record and make available information about the provenance of the data offered for reuse (e.g. the experiment or research project with which it originated, the processing it has undergone, any curation actions that have been applied)?

☐ Yes ☑ No

If it does, is the provenance information made available in any specific standard form, either for a particular community (e.g. in NeXus for the PaN community) or cross-domain (e.g. the PROV-O language)?

☐ Yes ☐ No

If 'Yes', then please provide further details on the specific standard(s) you use for provenance information:

12. What file formats are used to store and process data at your facility?

NeXus/HDF5 is the preferred data / file format

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

13. If your facility employs multiple file formats, how do you make these interoperable within and across your facility?

We encourage the scientists to have a NeXus or any other in the community widely used version of the files in the data repository.

More generally, how do you ensure that the file formats you use are interoperable with standard file formats widely used outside of PaN facilities (e.g. .txt, .csv, .pdf, .png, etc.)?

We support the scientists converting their data to NeXus.

Indexing and harvesting of metadata by machines

The questions asked in this section relate specifically to the following FAIR principle:

- F4. (Meta)data are registered or indexed in a searchable resource.

14. If your facility maintains a metadata catalogue, can this metadata be queried and retrieved using an API?

☑ Yes ☐ No

15. If your facility maintains a metadata catalogue, are the metadata standards you use compatible with OAI-PMH?

☑ Yes ☐ No

If 'No', do you maintain a mapping from your metadata to the metadata standards (e.g. Dublin Core, DataCite) commonly used by OAI-PMH?

☐ Yes ☐ No

PIDs

The questions asked in this section relate specifically to the following FAIR principles:

- F1. (Meta)data are assigned a globally unique and eternally persistent identifier.
- F3. Metadata specifies the data identifier
- A1. (Meta)data are retrievable by their identifier using a standardised communication protocol

16. What specific persistent identifier service(s) does your facility use? (select all that apply)

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

☑	Digital Object Identifier (DOI)
☑	Handle
☑	Open Researcher and Contributor ID (ORCID)
☑	Research Organization Registry (ROR)
☐	International Generic Sample Number (IGSN)

Does your facility use any other persistent identifier services?

17. If your facility has a metadata catalogue, do you include hyperlinked PIDs as part of the metadata provided on landing pages in that catalogue?

☑ Yes (data publications only) ☐ No

18. Do you use any types of PIDs within your facility's data management processes that you don't make available on your metadata catalogue landing pages?

☐ Yes ☑ No

If 'Yes', please provide further details (e.g. we store ORCIDs in our proposal system, but we do not include ORCIDs alongside author names on our catalogue landing pages).

HZB

Access to data by users and possibly by machines

The questions asked in this section relate specifically to the following FAIR principles:

- A1. (Meta)data are retrievable by their identifier using a standardised communication protocol
- A1.1 The protocol is open, free, and universally implementable
- A1.2 The protocol allows for an authentication and authorization procedure, where necessary
- R1.1 (Meta)data are released with a clear and accessible data usage license

19. How do potential (re)users of your facility's data know how they can access it? (e.g. this information is described on a web page, provided as part of the metadata, etc.)

The metadata catalogue provides a "cart" that the user can add items to while browsing the data. The user may then request to download that cart. The data publication landing pages have a "Request download" button.

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

20. Are there any authentication and authorisation measures in place to control or monitor who has access to your facility's data?

☑ Yes ☐ No

21. What protocols are employed for accessing the data? Please comment on whether the protocols are open and free (no cost) (e.g. HTTP, FTP), whether they depend on the size or other properties of the dataset, and whether you allow for the possibility of automated access to data through the protocols, as well as human access.

HTTPS.
Note that the data is on tape in most cases, so a request to stage that data must be made before being able to download it. Unfortunately, that staging request cannot easily be automated.

22. Does your facility have license agreements on data that it stores and makes available that apply to the users of that data and make clear how they may reuse it?

☑ Yes ☐ No

23. Are the licences you use standard ones (e.g. Creative Commons 4.0, public domain dedication CC0, etc.)?

☑ Yes ☐ No

Please list the types of licenses you apply to data at your facility.

CC0

24. Are the licences you use machine-understandable as well as readable by humans?

☑ Yes ☐ No

Curation of data

The questions asked in this section relate specifically to the following FAIR principles:

- A2. Metadata should be accessible even when the data is no longer available
- R1.2 (Meta)data are associated with their provenance
- R1.3 (Meta)data meet domain-relevant community standards

25. Do you have measures in place at your facility to ensure the integrity of data files against unintentional or unauthorised alteration?

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

☑ Yes ☐ No

If 'Yes', please provide further details of the measures you have in place.

Reminder: please ensure that you do not include any sensitive or confidential information in your response.

Data in the repository is not directly accessible to users, they may only download a copy. We keep multiple copies on tape. We calculate checksums during ingestion.

26. Is it ever possible that a dataset generated with its metadata during the experimental lifecycle could be removed from the facility's record?

☐ Yes ☑ No

If 'Yes', is the associated metadata retained when the dataset is removed?

☐ Yes ☐ No

Please provide any additional comments on the deletion or retention of data/metadata at your facility:

The "No" on the first question depends on what is meant by "experimental lifecycle": we keep all data for *at least* ten years. It has not yet been decided what happens after these ten years. It is very much possible that we'll keep the data much longer than that.

27. Do you take steps against the possibility of changes over time (e.g. evolution of community standards) that might affect the reusability of the data your facility holds?

☐ Yes ☑ No

If 'Yes', please provide further details of the steps you take.

Summary reflections

28. What does your facility take away from the FAIR self-assessment exercise, especially in terms of new insight into the FAIRness of your workflows and data management processes and potential avenues for future development?

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

HZB

The survey is nicely hinting room for improvement.

29. Please provide feedback and suggestions for improvement of the approach used for the FAIR self-assessment (e.g. questions asked, facility coordinator role, workshops, explanatory notes provided, links to existing FAIR evaluation models, focus on workflows and processes, etc.).

NOTE: please ensure you do not include any personal identifying information (e.g. names) in your response.

Maybe it could be considered having a more quantitative approach.

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

B1.6 HZDR

Comments on context

0. If you wish, please explain how you have chosen to handle references in the questionnaire to "data" (for example, what type[s] of data are covered by your responses), and/or the situation of multiple beamlines/instruments with different practices (for example, by focusing on only one), and any other context that will help to interpret your responses.

We distinguish between the published data in our data repository RODARE and the experimental data in the raw or processed state.

First and foremost, the answers refer to the data generated in the course of experiments in the PaN facility ELBE.
ELBE is the generic term for a series of sub-facilities (x ELBE). The sub-facility TELBE is the use case of HZDR, for which we have also reported reference data to ExPaNDS, etc.
Since this use case has the highest level of development with regard to FAIR compared to the other sub-facilities, it represents a realistic standard for a true self-assessment of our facility, as we used TELBE as a blueprint for the whole ELBE.
Nevertheless, for the assessment of ELBE, we have chosen TELBE as the reference model and based on the TELBE responses, we have assessed the other sub-facilities accordingly in order to make a holistic assessment. Ultimately, the developments of our reference facility TELBE will be transferred to the other ELBE facilities in the near future.

Existence, completeness, and richness of metadata

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

1. Is metadata associated with the data that are created/collected across the experimental lifecycle at your facility?

☑ Yes ☐ No

2. Looking at the metadata listed as essential in the ExPaNDS metadata framework across all stages of the experimental lifecycle (see Section 6.3, pages 54-56, Essential metadata to capture in each stage of the experimental lifecycle), is all of this metadata captured (i.e. either automatically or manually) at your facility?

☐ Yes ☑ No

If there are any essential metadata that you do not capture, what are they?

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

We do not capture the processing and analysis specific metadata, because our metadata database is still under development. Our initial SciCat prototype is used to evaluate the automated metadata pipeline to capture most of the metadata generated during the experimental life cycle.

3. When data from your facility is accessed, do you provide any metadata that is specifically intended to aid the reuse of that data, i.e. in distinction from metadata for discovery of the data (e.g. do you provide metadata that gives contextual information about how the data was generated)?

☐ Yes ☑ No

If so, what is the metadata that you provide to aid reuse?

–

Search (flexibility and capability)

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- F4. (Meta)data are registered or indexed in a searchable resource

4. Is it possible to search metadata related to data from the experimental lifecycle at your facility?

☑ Yes ☐ No

5. Does the metadata enable basic discovery (e.g. does the metadata include bibliographic information such as author, title, date, etc.)?

☑ Yes ☐ No

Can you make multi-faceted, PaN-specific queries (e.g. technique, experimental parameters, instrument, sample)?

☐ Yes ☑ No

Any additional comments on how your metadata enables discovery of data via search?

Our open accessible datasets are all searchable via our data repository RODARE, where all datasets implement the DataCite metadata schema. Additional fields are added (e.g. facility, research infrastructure, id in our internal publication system ROBIS and funding grants). The metadata is also available via B2FIND.
The provision of additional PaN-specific metadata via an additional PAN-specific metadata

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

catalogue is currently under development.

6. How is the metadata searchable? (select all that apply)

☑	Free text search
☑	Controlled vocabulary
☑	Filters (e.g. date, topic, instrument, etc.)

Are there any other ways of searching your metadata?

Currently, there are no further options for external searching the metadata at HZDR. For HZDR scientists with access to our IT infrastructure we provide search-interfaces for the metadata available in our internal databases and various sub-systems.

7. Who can search the metadata (e.g. searching restricted in some way, searchable by anyone)?

In the HZDR data repository RODARE, the metadata of all records is always searchable by everyone, regardless of the access right. The metadata for RODARE datasets is also available in B2Find and Open AIRE. The metadata of archived data is only searchable by the owner of the data or the research group/institute until publication in RODARE. The metadata in our databases is only searchable by HZDR scientists after access was granted.

Standardisation

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- F4. (Meta)data are registered or indexed in a searchable resource.
- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data.
- R1.2 (Meta)data are associated with their provenance

8. Which metadata standard(s) does your facility use for data? (select all that apply)

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

☐	Dublin Core
☑	DataCite
☐	DCAT
☐	NeXus
☑	B2FIND
☑	OpenAIRE (i.e. DataCite with minor adjustments)

Do you use any other metadata standards for data?

No, but an internal, on Data Cite oriented standard is in development. NeXus will be used for automated data collection in the future.

9. Does your facility use controlled vocabularies for metadata description?

☑ Yes ☐ No

If 'Yes', which one(s) do you use and for which types of metadata (e.g. PaNet ontology for identifying the technique used, list of keywords covering topics or research areas, fixed list of instruments, etc.)?

We have a controlled vocabulary for our facilities and devices, research grants and projects, access right and licences, authors (including ORCIDs and DOIs), Related identifiers,and upload types, departmental/institutional affiliation (ROBIS), Publication date.

10. Does your facility link data/metadata with other relevant data/metadata (e.g. do you link data/metadata from your facility to related experiments, resulting publications, calibration data, etc.)?

☑ Yes ☐ No

If 'Yes', do you do this in a way that standardises and captures these relationships formally (e.g. through the use of a 'related resource' metadata field or similar)?

☑ Yes ☐ No

Any additional comments on how your metadata supports formalised links between data/metadata?

For published datasets in RODARE it is possible to include references and subjects. Supported identifiers include: DOI, Handle, ARK, PURL, ISSN, ISBN, PubMed ID, PubMed

HZDR

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

Central ID, ADS Bibliographic Code, arXiv, Life Science Identifiers (LSID), EAN-13, ISTC, URNs and URLs.

11. Does your facility record and make available information about the provenance of the data offered for reuse (e.g. the experiment or research project with which it originated, the processing it has undergone, any curation actions that have been applied)?

☐ Yes ☑ No

If it does, is the provenance information made available in any specific standard form, either for a particular community (e.g. in NeXus for the PaN community) or cross-domain (e.g. the PROV-O language)?

☐ Yes ☑ No

If 'Yes', then please provide further details on the specific standard(s) you use for provenance information:

–

12. What file formats are used to store and process data at your facility?

There are no restrictions on suitable file formats - all formats are allowed. Nevertheless, the uploader should choose an open format which is suitable for long-term preservation and accessible without any restrictions.

13. If your facility employs multiple file formats, how do you make these interoperable within and across your facility?

Currently, it cannot be guaranteed that file formats will always be offered interoperably, as they may exist in proprietary formats, for example.
Our data repository RODARE does not allow the curation of the datasets at the moment and our scientists can upload every file format they want. However, the basis of RODARE is currently being gradually changed from Invenio to InvenioRDM.
With the new framework, a curation function will be implemented that will prevent the publication of non-interoperable file formats.

More generally, how do you ensure that the file formats you use are interoperable with standard file formats widely used outside of PaN facilities (e.g. .txt, .csv, .pdf, .png, etc.)?

See above

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

Indexing and harvesting of metadata by machines

The questions asked in this section relate specifically to the following FAIR principle:

- F4. (Meta)data are registered or indexed in a searchable resource.

14. If your facility maintains a metadata catalogue, can this metadata be queried and retrieved using an API?

☑ Yes ☐ No

15. If your facility maintains a metadata catalogue, are the metadata standards you use compatible with OAI-PMH?

☑ Yes ☐ No

If 'No', do you maintain a mapping from your metadata to the metadata standards (e.g. Dublin Core, DataCite) commonly used by OAI-PMH?

☐ Yes ☐ No

PIDs

The questions asked in this section relate specifically to the following FAIR principles:

- F1. (Meta)data are assigned a globally unique and eternally persistent identifier.
- F3. Metadata specifies the data identifier
- A1. (Meta)data are retrievable by their identifier using a standardised communication protocol

16. What specific persistent identifier service(s) does your facility use? (select all that apply)

☑	Digital Object Identifier (DOI)
☐	Handle
☑	Open Researcher and Contributor ID (ORCID)
☑	Research Organization Registry (ROR)
☐	International Generic Sample Number (IGSN)

Does your facility use any other persistent identifier services?

We use additional URNs for publication landing pages in the HZDR publication repository ROBIS. A handle system to provide open accessible persistent identifiers for all digital objects is currently under development.

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

17. If your facility has a metadata catalogue, do you include hyperlinked PIDs as part of the metadata provided on landing pages in that catalogue?

☑ Yes ☐ No

18. Do you use any types of PIDs within your facility's data management processes that you don't make available on your metadata catalogue landing pages?

☐ Yes ☑ No

If 'Yes', please provide further details (e.g. we store ORCIDs in our proposal system, but we do not include ORCIDs alongside author names on our catalogue landing pages).

Access to data by users and possibly by machines

The questions asked in this section relate specifically to the following FAIR principles:

- A1. (Meta)data are retrievable by their identifier using a standardised communication protocol
- A1.1 The protocol is open, free, and universally implementable
- A1.2 The protocol allows for an authentication and authorization procedure, where necessary
- R1.1 (Meta)data are released with a clear and accessible data usage license

19. How do potential (re)users of your facility's data know how they can access it? (e.g. this information is described on a web page, provided as part of the metadata, etc.)

In our data publication repository RODARE all files can be accessed under open, embargoed, restricted or closed access. This information on access regulation is displayed on the respective landing page of the record in the research data repository and provided as part of the metadata. A simple way to request access is available and will be forwarded to the owner.
For all datasets the HZDR is overall owner and if there is a legitimate interest in accessing data, then authorised groups must decide on access according to the hierarchy level.

20. Are there any authentication and authorisation measures in place to control or monitor who has access to your facility's data?

☑ Yes ☐ No

21. What protocols are employed for accessing the data? Please comment on whether the protocols are open and free (no cost) (e.g. HTTP, FTP), whether they depend on the size or other properties of the dataset, and whether you allow for the possibility of automated access to data through the protocols, as well as human access.

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

Published datasets and files can be accessed via HTTP protocol and REST APIs without any limitations.

22. Does your facility have license agreements on data that it stores and makes available that apply to the users of that data and make clear how they may reuse it?

☑ Yes ☐ No

23. Are the licences you use standard ones (e.g. Creative Commons 4.0, public domain dedication CC0, etc.)?

☑ Yes ☐ No

Please list the types of licenses you apply to data at your facility.

For data we recommend:
Attribution 4.0 International (CC-BY-4.0)
Attribution-NonCommercial 4.0 International (CC-BY-NC-4.0)

Other licences from the spdx licence list are also possible for (research) software and other digital objects.

24. Are the licences you use machine-understandable as well as readable by humans?

☑ Yes ☐ No

Curation of data

The questions asked in this section relate specifically to the following FAIR principles:

- A2. Metadata should be accessible even when the data is no longer available
- R1.2 (Meta)data are associated with their provenance
- R1.3 (Meta)data meet domain-relevant community standards

25. Do you have measures in place at your facility to ensure the integrity of data files against unintentional or unauthorised alteration?

☑ Yes ☐ No

If 'Yes', please provide further details of the measures you have in place.

Reminder: please ensure that you do not include any sensitive or confidential information in your response.

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

For RODARE we have background jobs checking the integrity of the published datasets in regular intervals. Changes of published data require the creation of a new version and we can keep track of the changes. Unintentional changes can be undone.

26. Is it ever possible that a dataset generated with its metadata during the experimental lifecycle could be removed from the facility's record?

☑ Yes ☐ No

If 'Yes', is the associated metadata retained when the dataset is removed?

☑ Yes ☐ No

Please provide any additional comments on the deletion or retention of data/metadata at your facility:

Metadata of deleted records is retained and can be reconstructed internally.
The data itself is permanently deleted, when there is no copy in our internal data archive.
After deletion, the DOI is resolved to a tombstone site, which explains what was there before and for what reason it was deleted.

27. Do you take steps against the possibility of changes over time (e.g. evolution of community standards) that might affect the reusability of the data your facility holds?

☐ Yes ☑ No

If 'Yes', please provide further details of the steps you take.

Summary reflections

28. What does your facility take away from the FAIR self-assessment exercise, especially in terms of new insight into the FAIRness of your workflows and data management processes and potential avenues for future development?

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

We are in the process of establishing a metadata catalogue for the experiment and processing/analysing relevant metadata to provide comprehensible research experiments which respect the FAIR principles. Our datasets are publicly accessible, but we need to extend the metadata beyond the pure DataCite metadata with additional metadata fields as shown in the ExPaNDS deliverable D2.2. Our future solution is based on SciCat with NeXus metadata schema.
We will continue the development of a sample catalogue and, if possible, collect all metadata in suitable databases and catalogues and make them available according to FAIR criteria. In this context, particular consideration will be given to the use of computational workflows in order to enable traceability and data provenance. Here we will pay particular attention to the DataCite Metadata Schema 4.5 and the metadata listed as essential in the ExPaNDS metadata framework at all stages of the experimental life cycle.
Further specific recommendations for processes and trained persons who understand and can classify experiments and initiate or prevent corresponding workflows will be set up. As much as possible should be done automatically and less should be added manually by the scientist.

29. Please provide feedback and suggestions for improvement of the approach used for the FAIR self-assessment (e.g. questions asked, facility coordinator role, workshops, explanatory notes provided, links to existing FAIR evaluation models, focus on workflows and processes, etc.).

NOTE: please ensure you do not include any personal identifying information (e.g. names) in your response.

HZDR

We need more comment fields below (every) question to clarify the situation at our RI.
Especially RIs with a higher heterogeneity in research fields and facilities.
Specify instructions if only one self-assessment document is to be completed for the entire centre, or how to deal with a higher degree of heterogeneity.

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

B1.7 ISIS

Comments on context

0. If you wish, please explain how you have chosen to handle references in the questionnaire to "data" (for example, what type[s] of data are covered by your responses), and/or the situation of multiple beamlines/instruments with different practices (for example, by focusing on only one), and any other context that will help to interpret your responses.

This response covers ISIS Neutron and Muon Source.

Please see here for how this facility interpret data: https://www.isis.stfc.ac.uk/Pages/Data-Policy.aspx and https://zenodo.org/record/6411387

ISIS has 30+ beamlines. Unless otherwise stated the responses below will be beamline agnostic.

Existence, completeness, and richness of metadata

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

1. Is metadata associated with the data that are created/collected across the experimental lifecycle at your facility?

☑ Yes ☐ No

2. Looking at the metadata listed as essential in the ExPaNDS metadata framework across all stages of the experimental lifecycle (see Section 6.3, pages 54-56, Essential metadata to capture in each stage of the experimental lifecycle), is all of this metadata captured (i.e. either automatically or manually) at your facility?

☐ Yes ☑ No

If there are any essential metadata that you do not capture, what are they?

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

With ref to section 6.3:

Metadata fields which are stored:

1. Principal Investigator
2. Co-Investigators
5. Facility
6. Proposal identifier (RB and DOI, so internal and external)
7. Experiment description
9. Actual visiting experimenters (no separation from co-investigators)
10. Experiment + measurement dates
12. Instrument
14. Dataset
15. Data format
26. Creator of record
29. Release/embargo date
30. Dataset title

Metadata field we don't store:
3. Instrument requested
8. Proposed experiment conditions
4. Sample description
16. Software used
28. Publication year

Partially/indirectly/unsure::
11. Enough sample information to understand structure + characteristics - some samples are named as their elemental components, but not all
13. Calibration - we store calibration datasets separately, so it is available, but there's no link to other investigations.
16. Processing information - unclear what this is
18 - 24. Original data link, dataset information, analysis data format - we only store raw data
25. Resource identity - not sure what this is. If it means DOIs for datafiles/datasets then no
27. Publisher of record - not sure what this is
31. Licence for usage - not on individual data, but covered by the data policy

ISIS

3. When data from your facility is accessed, do you provide any metadata that is specifically intended to aid the reuse of that data, i.e. in distinction from metadata for discovery of the data (e.g. do you provide metadata that gives contextual information about how the data was generated)?

☑ Yes ☐ No

If so, what is the metadata that you provide to aid reuse?

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

Descriptive metadata is captured with raw data (raw data, meaning the neutron and muon data and neutron image data collected from detectors and image cameras). This includes the capture of experimental conditions such as sample parameters (e.g. temperature) and beamline parameters (such as positions of detectors). The specifics of what descriptive metadata is stored varies with beamline. These metadata are intended to aid the facility users analyse their data during and after experiments, and for people with access to the ISIS Data Catalogue to reuse the data.

Search (flexibility and capability)

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- F4. (Meta)data are registered or indexed in a searchable resource

4. Is it possible to search metadata related to data from the experimental lifecycle at your facility?

☑ Yes ☐ No

5. Does the metadata enable basic discovery (e.g. does the metadata include bibliographic information such as author, title, date, etc.)?

☑ Yes ☐ No

Can you make multi-faceted, PaN-specific queries (e.g. technique, experimental parameters, instrument, sample)?

☐ Yes ☑ No

Any additional comments on how your metadata enables discovery of data via search?

6. How is the metadata searchable? (select all that apply)

☑	Free text search
☐	Controlled vocabulary
☑	Filters (e.g. date, topic, instrument, etc.)

Are there any other ways of searching your metadata?

ISIS

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

7. Who can search the metadata (e.g. searching restricted in some way, searchable by anyone)?

DOI metadata is public and searchable by everyone. (non-commercial) Experimental data are restricted to the experimental team for the embargo period (3-years).

Standardisation

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- F4. (Meta)data are registered or indexed in a searchable resource.
- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data.
- R1.2 (Meta)data are associated with their provenance

8. Which metadata standard(s) does your facility use for data? (select all that apply)

☐	Dublin Core
☑	DataCite
☐	DCAT
☑	NeXus
☐	B2FIND
☐	OpenAIRE (i.e. DataCite with minor adjustments)

Do you use any other metadata standards for data?

Specific file formats created for data treatment software developed follows the guideline “Liaise with upstream stakeholders to foster standard data formats and full metadata capture”, see https://doi.org/10.3233/jnr-220002 Also, there are plans to make available metadata via B2FIND and OpenAIRE in the near future (currently available on the development server).

9. Does your facility use controlled vocabularies for metadata description?

☐ Yes ☑ No

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

If 'Yes', which one(s) do you use and for which types of metadata (e.g. PaNet ontology for identifying the technique used, list of keywords covering topics or research areas, fixed list of instruments, etc.)?

We are planning to add PaNET.

10. Does your facility link data/metadata with other relevant data/metadata (e.g. do you link data/metadata from your facility to related experiments, resulting publications, calibration data, etc.)?

☑ Yes ☐ No

If 'Yes', do you do this in a way that standardises and captures these relationships formally (e.g. through the use of a 'related resource' metadata field or similar)?

☑ Yes ☐ No

Any additional comments on how your metadata supports formalised links between data/metadata?

Our platform supports linking to other data/metadata but there is more work needed on curation to provide these links.

11. Does your facility record and make available information about the provenance of the data offered for reuse (e.g. the experiment or research project with which it originated, the processing it has undergone, any curation actions that have been applied)?

☑ Yes ☐ No

If it does, is the provenance information made available in any specific standard form, either for a particular community (e.g. in NeXus for the PaN community) or cross-domain (e.g. the PROV-O language)?

☑ Yes ☐ No

If 'Yes', then please provide further details on the specific standard(s) you use for provenance information:

Via NeXus and in the automated minting process of DOIs that links observed data with information from the proposal We are not currently using a cross-domain representation of the provenance.

12. What file formats are used to store and process data at your facility?

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

Now NeXus. Historically another format was also used until recently.

13. If your facility employs multiple file formats, how do you make these interoperable within and across your facility?

To make different file formats (old and new) interoperable ISIS uses the Mantid data reduction framework: http://dx.doi.org/10.1016/j.nima.2014.07.029 .

More generally, how do you ensure that the file formats you use are interoperable with standard file formats widely used outside of PaN facilities (e.g. .txt, .csv, .pdf, .png, etc.)?

This is a build-in requirement for how ISIS develops data treatment software, and in agreement with other European Neutron facilities, as e.g. express in https://doi.org/10.3233/jnr-220002

Indexing and harvesting of metadata by machines

The questions asked in this section relate specifically to the following FAIR principle:

- F4. (Meta)data are registered or indexed in a searchable resource.

14. If your facility maintains a metadata catalogue, can this metadata be queried and retrieved using an API?

☑ Yes ☐ No

15. If your facility maintains a metadata catalogue, are the metadata standards you use compatible with OAI-PMH?

☐ Yes ☑ No

If 'No', do you maintain a mapping from your metadata to the metadata standards (e.g. Dublin Core, DataCite) commonly used by OAI-PMH?

☑ Yes ☐ No

PIDs

The questions asked in this section relate specifically to the following FAIR principles:

- F1. (Meta)data are assigned a globally unique and eternally persistent identifier.
- F3. Metadata specifies the data identifier

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

- A1. (Meta)data are retrievable by their identifier using a standardised communication protocol

16. What specific persistent identifier service(s) does your facility use? (select all that apply)

☑	Digital Object Identifier (DOI)
☐	Handle
☐	Open Researcher and Contributor ID (ORCID)
☐	Research Organization Registry (ROR)
☐	International Generic Sample Number (IGSN)

Does your facility use any other persistent identifier services?

ISIS uses so-called RB numbers to uniquely identify approved proposals. RB numbers are used with internal systems, communicated with users and searchable via the ISIS Data Catalogue. Furthermore these RB numbers forms the basis for creating memorisable DOIs, see https://www.isis.stfc.ac.uk/Pages/Digital-Object-Identifiers-(DOIs)-for-ISIS-Data.aspx

17. If your facility has a metadata catalogue, do you include hyperlinked PIDs as part of the metadata provided on landing pages in that catalogue?

☑ Yes ☐ No

18. Do you use any types of PIDs within your facility's data management processes that you don't make available on your metadata catalogue landing pages?

☐ Yes ☑ No

If 'Yes', please provide further details (e.g. we store ORCIDs in our proposal system, but we do not include ORCIDs alongside author names on our catalogue landing pages).

Access to data by users and possibly by machines

The questions asked in this section relate specifically to the following FAIR principles:

- A1. (Meta)data are retrievable by their identifier using a standardised communication protocol
- A1.1 The protocol is open, free, and universally implementable
- A1.2 The protocol allows for an authentication and authorization procedure, where necessary

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

- R1.1 (Meta)data are released with a clear and accessible data usage license

19. How do potential (re)users of your facility's data know how they can access it? (e.g. this information is described on a web page, provided as part of the metadata, etc.)

Users with successful ISIS proposals get emailed information concerning the ISIS Data Catalogue, https://data.isis.stfc.ac.uk/datagateway . A page on how to access ISIS data is available on the ISIS web site https://www.isis.stfc.ac.uk/Pages/ICAT.aspx

20. Are there any authentication and authorisation measures in place to control or monitor who has access to your facility's data?

☑ Yes ☐ No

21. What protocols are employed for accessing the data? Please comment on whether the protocols are open and free (no cost) (e.g. HTTP, FTP), whether they depend on the size or other properties of the dataset, and whether you allow for the possibility of automated access to data through the protocols, as well as human access.

There are two main APIs, one SOAP, the other REST, both of which are over HTTP. There's also an OAI-PMH (https://www.openarchives.org/pmh/) server. It's all possible to access in an automated way.

ISIS

22. Does your facility have license agreements on data that it stores and makes available that apply to the users of that data and make clear how they may reuse it?

☑ Yes ☐ No

23. Are the licences you use standard ones (e.g. Creative Commons 4.0, public domain dedication CC0, etc.)?

☑ Yes ☐ No

Please list the types of licenses you apply to data at your facility.

CC-BY license

24. Are the licences you use machine-understandable as well as readable by humans?

☑ Yes ☐ No

Curation of data

The questions asked in this section relate specifically to the following FAIR principles:

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

- A2. Metadata should be accessible even when the data is no longer available
- R1.2 (Meta)data are associated with their provenance
- R1.3 (Meta)data meet domain-relevant community standards

25. Do you have measures in place at your facility to ensure the integrity of data files against unintentional or unauthorised alteration?

☑ Yes ☐ No

If 'Yes', please provide further details of the measures you have in place.

Reminder: please ensure that you do not include any sensitive or confidential information in your response.

To date, all ISIS data are archived, and once a file is stored it is read only

26. Is it ever possible that a dataset generated with its metadata during the experimental lifecycle could be removed from the facility's record?

☐ Yes ☑ No

If 'Yes', is the associated metadata retained when the dataset is removed?

☐ Yes ☐ No

Please provide any additional comments on the deletion or retention of data/metadata at your facility:

27. Do you take steps against the possibility of changes over time (e.g. evolution of community standards) that might affect the reusability of the data your facility holds?

☑ Yes ☐ No

If 'Yes', please provide further details of the steps you take.

The entrypoint software for reading ISIS data file formats is called Mantid, http://dx.doi.org/10.1016/j.nima.2014.07.029. It has automated tests that are run at least daily, which checks that Mantid can read all ISIS data file formats (past or present), across multiple operations systems.

Summary reflections

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

28. What does your facility take away from the FAIR self-assessment exercise, especially in terms of new insight into the FAIRness of your workflows and data management processes and potential avenues for future development?

ISIS Neutron and Muon Source together with other STFC departments (most significantly the STFC SCD department) have been engaging with FAIR together with other European facilities for some time.

29. Please provide feedback and suggestions for improvement of the approach used for the FAIR self-assessment (e.g. questions asked, facility coordinator role, workshops, explanatory notes provided, links to existing FAIR evaluation models, focus on workflows and processes, etc.).

NOTE: please ensure you do not include any personal identifying information (e.g. names) in your response.

I liked the explanations, which aided in filling in the form.

There was a few places where more short explanations may also have been added:

- In question 2:
  - 25. Resource identity
  - 27. Publisher of record
  - 16. Processing information
- Question 6: Controlled vocabulary
- Question 8 and 16: a one line sentence explanation of each of the items in the tables

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

B1.8 MAX IV

Comments on context

0. If you wish, please explain how you have chosen to handle references in the questionnaire to "data" (for example, what type[s] of data are covered by your responses), and/or the situation of multiple beamlines/instruments with different practices (for example, by focusing on only one), and any other context that will help to interpret your responses.

The information in the below answers applies to the majority of the beamlines at MAX IV. Currently, 12 out of 16 beamlines at MAX IV, have the capability to save raw data files to a centrally managed storage solution, and collect and preserve structured information about the data in a metadata catalogue.The acquired metadata in the catalogue may differ from beamline to beamline, and even from experiment to experiment, as MAX IV offers the ability for the beamlines to configure themselves which instrument data to collect from the control system.

Existence, completeness, and richness of metadata

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

1. Is metadata associated with the data that are created/collected across the experimental lifecycle at your facility?

☑ Yes ☐ No

2. Looking at the metadata listed as essential in the ExPaNDS metadata framework across all stages of the experimental lifecycle (see Section 6.3, pages 54-56, Essential metadata to capture in each stage of the experimental lifecycle), is all of this metadata captured (i.e. either automatically or manually) at your facility?

☐ Yes ☑ No

If there are any essential metadata that you do not capture, what are they?

MAX IV captures most of the essential Proposal, Experiment and Record metadata automatically. For derived data, we rely on the researcher to manually provide information about Processing and Analysis with greater uncertainty about what is registered as a result.

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

3. When data from your facility is accessed, do you provide any metadata that is specifically intended to aid the reuse of that data, i.e. in distinction from metadata for discovery of the data (e.g. do you provide metadata that gives contextual information about how the data was generated)?

☐ Yes ☑ No

If so, what is the metadata that you provide to aid reuse?

Search (flexibility and capability)

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- F4. (Meta)data are registered or indexed in a searchable resource

4. Is it possible to search metadata related to data from the experimental lifecycle at your facility?

☑ Yes ☐ No

5. Does the metadata enable basic discovery (e.g. does the metadata include bibliographic information such as author, title, date, etc.)?

☑ Yes ☐ No

Can you make multi-faceted, PaN-specific queries (e.g. technique, experimental parameters, instrument, sample)?

☑ Yes ☐ No

Any additional comments on how your metadata enables discovery of data via search?

The MAX IV metadata catalogue supports OAI-PMH. The search API is linked to the PaN data Search Portal.

6. How is the metadata searchable? (select all that apply)

☑	Free text search
☐	Controlled vocabulary

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

☑	Filters (e.g. date, topic, instrument, etc.)

Are there any other ways of searching your metadata?

Conditions can be added to a search field.

7. Who can search the metadata (e.g. searching restricted in some way, searchable by anyone)?

Open datasets and their associated metadata are searchable by anyone, other datasets and metadata are only searchable by members of the research team who participate somehow on the experiment/proposal.

Standardisation

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- F4. (Meta)data are registered or indexed in a searchable resource.
- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data.
- R1.2 (Meta)data are associated with their provenance

8. Which metadata standard(s) does your facility use for data? (select all that apply)

☑	Dublin Core
☑	DataCite
☐	DCAT
☑	NeXus
☐	B2FIND
☐	OpenAIRE (i.e. DataCite with minor adjustments)

Do you use any other metadata standards for data?

9. Does your facility use controlled vocabularies for metadata description?

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

☐ Yes ☑ No

If 'Yes', which one(s) do you use and for which types of metadata (e.g. PaNet ontology for identifying the technique used, list of keywords covering topics or research areas, fixed list of instruments, etc.)?

10. Does your facility link data/metadata with other relevant data/metadata (e.g. do you link data/metadata from your facility to related experiments, resulting publications, calibration data, etc.)?

☐ Yes ☑ No

If 'Yes', do you do this in a way that standardises and captures these relationships formally (e.g. through the use of a 'related resource' metadata field or similar)?

☐ Yes ☐ No

Any additional comments on how your metadata supports formalised links between data/metadata?

Occasionally (but not always), we store calibration data in the NeXus file. Future plans to link to experimental notes in an electronic logbook.

MAX IV

11. Does your facility record and make available information about the provenance of the data offered for reuse (e.g. the experiment or research project with which it originated, the processing it has undergone, any curation actions that have been applied)?

☐ Yes ☑ No

If it does, is the provenance information made available in any specific standard form, either for a particular community (e.g. in NeXus for the PaN community) or cross-domain (e.g. the PROV-O language)?

☐ Yes ☐ No

If 'Yes', then please provide further details on the specific standard(s) you use for provenance information:

12. What file formats are used to store and process data at your facility?

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

We aspire to use NeXus/HDF5 as the facility standard, but there are exceptions such as *.txt, *.dat *.zip, *.pxp, *.ibw, *.sp2,*.sle, *.xy, *.vms, *.xml

13. If your facility employs multiple file formats, how do you make these interoperable within and across your facility?

There are initiatives at MAX IV to agree on interoperable formats between beamlines using similar techniques, where there is no user community practices.

More generally, how do you ensure that the file formats you use are interoperable with standard file formats widely used outside of PaN facilities (e.g. .txt, .csv, .pdf, .png, etc.)?

Convertors and io libraries are widely available in relevant communities (i.e. users can usually convert hdf5 to xy, tiff etc.).

Indexing and harvesting of metadata by machines

The questions asked in this section relate specifically to the following FAIR principle:

- F4. (Meta)data are registered or indexed in a searchable resource.

14. If your facility maintains a metadata catalogue, can this metadata be queried and retrieved using an API?

☑ Yes ☐ No

15. If your facility maintains a metadata catalogue, are the metadata standards you use compatible with OAI-PMH?

☑ Yes ☐ No

If 'No', do you maintain a mapping from your metadata to the metadata standards (e.g. Dublin Core, DataCite) commonly used by OAI-PMH?

☐ Yes ☐ No

PIDs

The questions asked in this section relate specifically to the following FAIR principles:

- F1. (Meta)data are assigned a globally unique and eternally persistent identifier.
- F3. Metadata specifies the data identifier
- A1. (Meta)data are retrievable by their identifier using a standardised communication protocol

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

16. What specific persistent identifier service(s) does your facility use? (select all that apply)

☑	Digital Object Identifier (DOI)
☑	Handle
☐	Open Researcher and Contributor ID (ORCID)
☐	Research Organization Registry (ROR)
☐	International Generic Sample Number (IGSN)

Does your facility use any other persistent identifier services?

17. If your facility has a metadata catalogue, do you include hyperlinked PIDs as part of the metadata provided on landing pages in that catalogue?

☑ Yes ☐ No

18. Do you use any types of PIDs within your facility's data management processes that you don't make available on your metadata catalogue landing pages?

☐ Yes ☑ No

If 'Yes', please provide further details (e.g. we store ORCIDs in our proposal system, but we do not include ORCIDs alongside author names on our catalogue landing pages).

Access to data by users and possibly by machines

The questions asked in this section relate specifically to the following FAIR principles:

- A1. (Meta)data are retrievable by their identifier using a standardised communication protocol
- A1.1 The protocol is open, free, and universally implementable
- A1.2 The protocol allows for an authentication and authorization procedure, where necessary
- R1.1 (Meta)data are released with a clear and accessible data usage license

19. How do potential (re)users of your facility's data know how they can access it? (e.g. this information is described on a web page, provided as part of the metadata, etc.)

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

The information is provided in the MAX IV public web page and in an internal wiki.

20. Are there any authentication and authorisation measures in place to control or monitor who has access to your facility's data?

☑ Yes ☐ No

21. What protocols are employed for accessing the data? Please comment on whether the protocols are open and free (no cost) (e.g. HTTP, FTP), whether they depend on the size or other properties of the dataset, and whether you allow for the possibility of automated access to data through the protocols, as well as human access.

MAX IV only allow access to data via Globus (GridFTP). The protocol is open and free, Globus clients are free. Machine access automation is not possible for MAX IV at the moment due to infrastructure constraints, but possible with Globus e.g. by using a REST-api.

22. Does your facility have license agreements on data that it stores and makes available that apply to the users of that data and make clear how they may reuse it?

☐ Yes ☑ No

23. Are the licences you use standard ones (e.g. Creative Commons 4.0, public domain dedication CC0, etc.)?

☐ Yes ☐ No

Please list the types of licenses you apply to data at your facility.

Curation of data

The questions asked in this section relate specifically to the following FAIR principles:

- A2. Metadata should be accessible even when the data is no longer available
- R1.2 (Meta)data are associated with their provenance
- R1.3 (Meta)data meet domain-relevant community standards

25. Do you have measures in place at your facility to ensure the integrity of data files against unintentional or unauthorised alteration?

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

☑ Yes ☐ No

If 'Yes', please provide further details of the measures you have in place.

Reminder: please ensure that you do not include any sensitive or confidential information in your response.

Data files that have been made openly accessible are copied to a special segment in the MAX IV storage system, with a highly controlled and restricted permission schema that will prevent unintentional or unauthorised alteration. In addition, raw data from detectors is saved in a folder locked for changes.

26. Is it ever possible that a dataset generated with its metadata during the experimental lifecycle could be removed from the facility's record?

☑ Yes ☐ No

If 'Yes', is the associated metadata retained when the dataset is removed?

☑ Yes ☐ No

Please provide any additional comments on the deletion or retention of data/metadata at your facility:

MAX IV is currently expanding its long-term storage so that a 7-year minimum data retention policy can be offered to the users but with the strive to store data +10 years. Metadata will be stored indefinitely.

27. Do you take steps against the possibility of changes over time (e.g. evolution of community standards) that might affect the reusability of the data your facility holds?

☐ Yes ☑ No

If 'Yes', please provide further details of the steps you take.

Summary reflections

28. What does your facility take away from the FAIR self-assessment exercise, especially in terms of new insight into the FAIRness of your workflows and data management processes and potential avenues for future development?

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

The facility needs to intensify efforts to meet FAIR and particularly interoperability and reuse of data. More interoperability requires greater ability to manage and comply with multiple formats and standards, but also major efforts that needs the researcher's involvement and commitment to FAIR. This is a challenge when researcher's engagement in FAIR generally is low.
For improved possibility for reuse of data and research output, the facility clearly needs to improve the data management support in several phases of the experimental lifecycle. Among first and foremost to improve are:

- Support for ontologies to refine discovery of the data
- Structured curation of data from analysis and processing phases
- The ability to preserve contextual information about an experiment and linking to digital objects such as calibration data, experimental notes, software used for analysing and processing of taken data, etc.
- Machine accessibility and understandability of the data

29. Please provide feedback and suggestions for improvement of the approach used for the FAIR self-assessment (e.g. questions asked, facility coordinator role, workshops, explanatory notes provided, links to existing FAIR evaluation models, focus on workflows and processes, etc.).

NOTE: please ensure you do not include any personal identifying information (e.g. names) in your response.

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

B1.9 PSI

Comments on context

0. If you wish, please explain how you have chosen to handle references in the questionnaire to "data" (for example, what type[s] of data are covered by your responses), and/or the situation of multiple beamlines/instruments with different practices (for example, by focusing on only one), and any other context that will help to interpret your responses.

Existence, completeness, and richness of metadata

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

1. Is metadata associated with the data that are created/collected across the experimental lifecycle at your facility?

☑ Yes ☐ No

2. Looking at the metadata listed as essential in the ExPaNDS metadata framework across all stages of the experimental lifecycle (see Section 6.3, pages 54-56, Essential metadata to capture in each stage of the experimental lifecycle), is all of this metadata captured (i.e. either automatically or manually) at your facility?

☐ Yes ☑ No

If there are any essential metadata that you do not capture, what are they?

The following fields are missing:

- Proposal:
  - Sample description
  - Proposed experiment conditions
- Experiment:
  - Sample information

3. When data from your facility is accessed, do you provide any metadata that is specifically intended to aid the reuse of that data, i.e. in distinction from metadata for discovery of the data (e.g. do you provide metadata that gives contextual information about how the data was generated)?

PSI

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

☑ Yes ☐ No

If so, what is the metadata that you provide to aid reuse?

In the case of derived datasets, there is a direct link to the raw dataset(s) that it was derived from. Scientists can also add any other information they see fit.

Search (flexibility and capability)

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- F4. (Meta)data are registered or indexed in a searchable resource

4. Is it possible to search metadata related to data from the experimental lifecycle at your facility?

☐ Yes ☑ No

5. Does the metadata enable basic discovery (e.g. does the metadata include bibliographic information such as author, title, date, etc.)?

☑ Yes ☐ No

Can you make multi-faceted, PaN-specific queries (e.g. technique, experimental parameters, instrument, sample)?

☑ Yes ☐ No

Any additional comments on how your metadata enables discovery of data via search?

Even though the catalogue supports the search-API and all its functionalities, not every query returns results, as some further curation of the metadata is still required.

6. How is the metadata searchable? (select all that apply)

☑	Free text search
☐	Controlled vocabulary
☑	Filters (e.g. date, topic, instrument, etc.)

Are there any other ways of searching your metadata?

PSI

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

No.

7. Who can search the metadata (e.g. searching restricted in some way, searchable by anyone)?

Anyone for public datasets.

Standardisation

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- F4. (Meta)data are registered or indexed in a searchable resource.
- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data.
- R1.2 (Meta)data are associated with their provenance

8. Which metadata standard(s) does your facility use for data? (select all that apply)

☑	Dublin Core
☐	DataCite
☐	DCAT
☐	NeXus
☑	B2FIND
☐	OpenAIRE (i.e. DataCite with minor adjustments)

Do you use any other metadata standards for data?

No.

9. Does your facility use controlled vocabularies for metadata description?

☐ Yes ☑ No

If ‘Yes’, which one(s) do you use and for which types of metadata (e.g. PaNet ontology for identifying the technique used, list of keywords covering topics or research areas, fixed list of instruments, etc.)?

PSI

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

We are planning to introduce the PaNet ontology.

10. Does your facility link data/metadata with other relevant data/metadata (e.g. do you link data/metadata from your facility to related experiments, resulting publications, calibration data, etc.)?

☑ Yes ☐ No

If 'Yes', do you do this in a way that standardises and captures these relationships formally (e.g. through the use of a 'related resource' metadata field or similar)?

☑ Yes ☐ No

Any additional comments on how your metadata supports formalised links between data/metadata?

11. Does your facility record and make available information about the provenance of the data offered for reuse (e.g. the experiment or research project with which it originated, the processing it has undergone, any curation actions that have been applied)?

☑ Yes ☐ No

If it does, is the provenance information made available in any specific standard form, either for a particular community (e.g. in NeXus for the PaN community) or cross-domain (e.g. the PROV-O language)?

☐ Yes ☑ No

If 'Yes', then please provide further details on the specific standard(s) you use for provenance information:

12. What file formats are used to store and process data at your facility?

The main ones are:
- hdf5/nexus
- tiff and other image format files (e.g. png)
- ascii
- csv
- pdf

PSI

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

13. If your facility employs multiple file formats, how do you make these interoperable within and across your facility?

We don't. It is up to the dataset owner how much information they provide to help interoperability.

More generally, how do you ensure that the file formats you use are interoperable with standard file formats widely used outside of PaN facilities (e.g. .txt, .csv, .pdf, .png, etc.)?

We use well-known file formats only.

Indexing and harvesting of metadata by machines

The questions asked in this section relate specifically to the following FAIR principle:

- F4. (Meta)data are registered or indexed in a searchable resource.

14. If your facility maintains a metadata catalogue, can this metadata be queried and retrieved using an API?

☑ Yes ☐ No

15. If your facility maintains a metadata catalogue, are the metadata standards you use compatible with OAI-PMH?

☐ Yes ☑ No

If 'No', do you maintain a mapping from your metadata to the metadata standards (e.g. Dublin Core, DataCite) commonly used by OAI-PMH?

☑ Yes ☐ No

PIDs

The questions asked in this section relate specifically to the following FAIR principles:

- F1. (Meta)data are assigned a globally unique and eternally persistent identifier.
- F3. Metadata specifies the data identifier
- A1. (Meta)data are retrievable by their identifier using a standardised communication protocol

16. What specific persistent identifier service(s) does your facility use? (select all that apply)

PSI

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

☑	Digital Object Identifier (DOI)
☐	Handle
☑	Open Researcher and Contributor ID (ORCID)
☐	Research Organization Registry (ROR)
☐	International Generic Sample Number (IGSN)

Does your facility use any other persistent identifier services?

No.

17. If your facility has a metadata catalogue, do you include hyperlinked PIDs as part of the metadata provided on landing pages in that catalogue?

☐ Yes ☑ No

18. Do you use any types of PIDs within your facility's data management processes that you don't make available on your metadata catalogue landing pages?

☑ Yes ☐ No

If 'Yes', please provide further details (e.g. we store ORCIDs in our proposal system, but we do not include ORCIDs alongside author names on our catalogue landing pages).

We do not show ORCIDs on the PSI Public Data Repository.

Access to data by users and possibly by machines

The questions asked in this section relate specifically to the following FAIR principles:

- A1. (Meta)data are retrievable by their identifier using a standardised communication protocol
- A1.1 The protocol is open, free, and universally implementable
- A1.2 The protocol allows for an authentication and authorization procedure, where necessary
- R1.1 (Meta)data are released with a clear and accessible data usage license

19. How do potential (re)users of your facility's data know how they can access it? (e.g. this information is described on a web page, provided as part of the metadata, etc.)

Provided as part of the metadata for public datasets.

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

20. Are there any authentication and authorisation measures in place to control or monitor who has access to your facility's data?

☑ Yes ☐ No

21. What protocols are employed for accessing the data? Please comment on whether the protocols are open and free (no cost) (e.g. HTTP, FTP), whether they depend on the size or other properties of the dataset, and whether you allow for the possibility of automated access to data through the protocols, as well as human access.

HTTPS
SSH (e.g. rsync)
GridFTP (e.g. globusOnline)

All are free to use for end users. There are no size or other restrictions.

22. Does your facility have license agreements on data that it stores and makes available that apply to the users of that data and make clear how they may reuse it?

☐ Yes ☑ No

23. Are the licences you use standard ones (e.g. Creative Commons 4.0, public domain dedication CC0, etc.)?

☑ Yes ☐ No

Please list the types of licenses you apply to data at your facility.

CC BY-SA 4.0

24. Are the licences you use machine-understandable as well as readable by humans?

☑ Yes ☐ No

Curation of data

The questions asked in this section relate specifically to the following FAIR principles:

- A2. Metadata should be accessible even when the data is no longer available
- R1.2 (Meta)data are associated with their provenance
- R1.3 (Meta)data meet domain-relevant community standards

25. Do you have measures in place at your facility to ensure the integrity of data files against unintentional or unauthorised alteration?

PSI

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

☑ Yes ☐ No

If 'Yes', please provide further details of the measures you have in place.

Reminder: please ensure that you do not include any sensitive or confidential information in your response.

Access to all non-published data requires a username and password.

When datasets are archived, the checksum of each file in the dataset is also calculated and stored. During retrieval, the checksums are then compared to detect any alterations or corruption in any of the files.

The data catalogue provides audit tracking on the changes in the metadata. Only authenticated and authorized users can change metadata.

26. Is it ever possible that a dataset generated with its metadata during the experimental lifecycle could be removed from the facility's record?

☑ Yes ☐ No

If 'Yes', is the associated metadata retained when the dataset is removed?

☑ Yes ☐ No

Please provide any additional comments on the deletion or retention of data/metadata at your facility:

Scientists have to explicitly mark the datasets they want to retain.

27. Do you take steps against the possibility of changes over time (e.g. evolution of community standards) that might affect the reusability of the data your facility holds?

☑ Yes ☐ No

If 'Yes', please provide further details of the steps you take.

We are actively participating in the community and keep an eye on upcoming changes. We do not have any means to prevent changes from happening.
We evaluate the need to update our infrastructure on a case-by-case basis if a change happens that might impact reusability.

PSI

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

Summary reflections

28. What does your facility take away from the FAIR self-assessment exercise, especially in terms of new insight into the FAIRness of your workflows and data management processes and potential avenues for future development?

It is a good exercise to periodically review our processes against a standardized set of criteria and to compare it against other facility responses.
It also helps us to have an active understanding of what metadata we capture compared to the recommended standards.

29. Please provide feedback and suggestions for improvement of the approach used for the FAIR self-assessment (e.g. questions asked, facility coordinator role, workshops, explanatory notes provided, links to existing FAIR evaluation models, focus on workflows and processes, etc.).

NOTE: please ensure you do not include any personal identifying information (e.g. names) in your response.

We miss questions related to the search API standard that was developed as part of PaNOSC.

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

B1.10 SOLEIL

Comments on context

0. If you wish, please explain how you have chosen to handle references in the questionnaire to "data" (for example, what type[s] of data are covered by your responses), and/or the situation of multiple beamlines/instruments with different practices (for example, by focusing on only one), and any other context that will help to interpret your responses.

In all our answers we tried to include the status of SOLEIL and our beamlines with regard to data management taking into account all types of data and metadata.
For some questions where it was difficult to say clearly if it is YES or NO because the situation is in between (in progress, in test phase, under discussion with the SOLEIL management or the beamlines….) we chose to complete the answer with additional comments to explain the context.

Existence, completeness, and richness of metadata

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

1. Is metadata associated with the data that are created/collected across the experimental lifecycle at your facility?

☑ Yes ☐ No

2. Looking at the metadata listed as essential in the ExPaNDS metadata framework across all stages of the experimental lifecycle (see Section 6.3, pages 54-56, Essential metadata to capture in each stage of the experimental lifecycle), is all of this metadata captured (i.e. either automatically or manually) at your facility?

☐ Yes ☑ No

If there are any essential metadata that you do not capture, what are they?

- The main important metadata are in the User office tool (SUNset) or in the NeXus files (for the experimental raw data).
- Calibration information are in the experimental log book.

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

- Dataset Information : yes only for those using NeXus
- Processing and Analysis Metadata : not handled equally at all the beamlines

Record metadata : in progress

3. When data from your facility is accessed, do you provide any metadata that is specifically intended to aid the reuse of that data, i.e. in distinction from metadata for discovery of the data (e.g. do you provide metadata that gives contextual information about how the data was generated)?

☑ Yes ☐ No

If so, what is the metadata that you provide to aid reuse?

Registered contextual metadata : beam intensity, scan trajectory, sample contextual environment (temperature, pressure, ...)

Search (flexibility and capability)

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- F4. (Meta)data are registered or indexed in a searchable resource

4. Is it possible to search metadata related to data from the experimental lifecycle at your facility?

☐ Yes ☑ No

5. Does the metadata enable basic discovery (e.g. does the metadata include bibliographic information such as author, title, date, etc.)?

☑ Yes ☐ No

Can you make multi-faceted, PaN-specific queries (e.g. technique, experimental parameters, instrument, sample)?

☐ Yes ☑ No

Any additional comments on how your metadata enables discovery of data via search?

Answer is NO as we are still in the process of deploying SciCat at SOLEIL

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

6. How is the metadata searchable? (select all that apply)

☑	Free text search
☑	Controlled vocabulary
☑	Filters (e.g. date, topic, instrument, etc.)

Are there any other ways of searching your metadata?

Via the SciCat data catalog (installation in progress)

It is then foreseen is to integrate the PaN Search API and the Federated search API.
The « Free text search » and « filters » will be provided first. Ontologies will come later

7. Who can search the metadata (e.g. searching restricted in some way, searchable by anyone)?

Once SciCat is installed: persons having an account with the right authorisations (limited to the experimental project team during the embargo, open access afterwards, as stated in SOLEIL data policy).

Standardisation

The questions asked in this section relate specifically to the following FAIR principles:

- F2. Data are described with rich metadata
- F4. (Meta)data are registered or indexed in a searchable resource.
- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data.
- R1.2 (Meta)data are associated with their provenance

8. Which metadata standard(s) does your facility use for data? (select all that apply)

☐	Dublin Core
☑	DataCite (in progress)
☐	DCAT
☑	NeXus
☐	B2FIND

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

☐	OpenAIRE (i.e. DataCite with minor adjustments)

Do you use any other metadata standards for data?

B2FIND and OpenAIRE will come after the federated search API integration

9. Does your facility use controlled vocabularies for metadata description?

☐ Yes ☑ No

If 'Yes', which one(s) do you use and for which types of metadata (e.g. PaNet ontology for identifying the technique used, list of keywords covering topics or research areas, fixed list of instruments, etc.)?

10. Does your facility link data/metadata with other relevant data/metadata (e.g. do you link data/metadata from your facility to related experiments, resulting publications, calibration data, etc.)?

☐ Yes ☑ No

If 'Yes', do you do this in a way that standardises and captures these relationships formally (e.g. through the use of a 'related resource' metadata field or similar)?

☐ Yes ☐ No

Any additional comments on how your metadata supports formalised links between data/metadata?

11. Does your facility record and make available information about the provenance of the data offered for reuse (e.g. the experiment or research project with which it originated, the processing it has undergone, any curation actions that have been applied)?

☑ Yes ☐ No

If it does, is the provenance information made available in any specific standard form, either for a particular community (e.g. in NeXus for the PaN community) or cross-domain (e.g. the PROV-O language)?

☑ Yes ☐ No

SOLEIL

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

If 'Yes', then please provide further details on the specific standard(s) you use for provenance information:

In Nexus file: experiment project identifier

12. What file formats are used to store and process data at your facility?

NeXus mainly, EDF

13. If your facility employs multiple file formats, how do you make these interoperable within and across your facility?

Use of a SOLEIL locally implemented software layer, called Common Data Model Access (CDMA), using an abstract interface and multiple plugins to handle different data formats

More generally, how do you ensure that the file formats you use are interoperable with standard file formats widely used outside of PaN facilities (e.g. .txt, .csv, .pdf, .png, etc.)?

CDMA

Indexing and harvesting of metadata by machines

The questions asked in this section relate specifically to the following FAIR principle:

- F4. (Meta)data are registered or indexed in a searchable resource.

14. If your facility maintains a metadata catalogue, can this metadata be queried and retrieved using an API?

☐ Yes ☑ No (Yes, as soon as SciCat is deployed)

15. If your facility maintains a metadata catalogue, are the metadata standards you use compatible with OAI-PMH?

☐ Yes ☑ No (Yes, after federated search API installation)

If 'No', do you maintain a mapping from your metadata to the metadata standards (e.g. Dublin Core, DataCite) commonly used by OAI-PMH?

☐ Yes ☑ No

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

PIDs

The questions asked in this section relate specifically to the following FAIR principles:

- F1. (Meta)data are assigned a globally unique and eternally persistent identifier.
- F3. Metadata specifies the data identifier
- A1. (Meta)data are retrievable by their identifier using a standardised communication protocol

16. What specific persistent identifier service(s) does your facility use? (select all that apply)

☑	Digital Object Identifier (DOI)
☐	Handle
☑	Open Researcher and Contributor ID (ORCID)
☑	Research Organization Registry (ROR)
☐	International Generic Sample Number (IGSN)

Does your facility use any other persistent identifier services?

DIGSN upcoming (Digital LEAPS STARS project)

17. If your facility has a metadata catalogue, do you include hyperlinked PIDs as part of the metadata provided on landing pages in that catalogue?

☐ Yes ☑ No (In progress with SciCat)

18. Do you use any types of PIDs within your facility's data management processes that you don't make available on your metadata catalogue landing pages?

☐ Yes ☑ No

If 'Yes', please provide further details (e.g. we store ORCIDs in our proposal system, but we do not include ORCIDs alongside author names on our catalogue landing pages).

Access to data by users and possibly by machines

The questions asked in this section relate specifically to the following FAIR principles:

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

- A1. (Meta)data are retrievable by their identifier using a standardised communication protocol
- A1.1 The protocol is open, free, and universally implementable
- A1.2 The protocol allows for an authentication and authorization procedure, where necessary
- R1.1 (Meta)data are released with a clear and accessible data usage license

19. How do potential (re)users of your facility's data know how they can access it? (e.g. this information is described on a web page, provided as part of the metadata, etc.)

Web pages to come (landing page, SUNset)
+ provided as part of the metadata

20. Are there any authentication and authorisation measures in place to control or monitor who has access to your facility's data?

☑ Yes ☐ No

21. What protocols are employed for accessing the data? Please comment on whether the protocols are open and free (no cost) (e.g. HTTP, FTP), whether they depend on the size or other properties of the dataset, and whether you allow for the possibility of automated access to data through the protocols, as well as human access.

GLOBUS : https://www.globus.org/data-transfer

High-performance data transfers
Automation still under test at SOLEIL

22. Does your facility have license agreements on data that it stores and makes available that apply to the users of that data and make clear how they may reuse it?

☑ Yes ☐ No

23. Are the licences you use standard ones (e.g. Creative Commons 4.0, public domain dedication CC0, etc.)?

☑ Yes ☐ No

Please list the types of licenses you apply to data at your facility.

CC BY 4.0

24. Are the licences you use machine-understandable as well as readable by humans?

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

☑ Yes ☐ No

Curation of data

The questions asked in this section relate specifically to the following FAIR principles:

- A2. Metadata should be accessible even when the data is no longer available
- R1.2 (Meta)data are associated with their provenance
- R1.3 (Meta)data meet domain-relevant community standards

25. Do you have measures in place at your facility to ensure the integrity of data files against unintentional or unauthorised alteration?

☑ Yes ☐ No

If 'Yes', please provide further details of the measures you have in place.

Reminder: please ensure that you do not include any sensitive or confidential information in your response.

Read-only raw data files, archive history with HSM software Active Circle (https://active-circle.com/), multiple copies.

SOLEIL

26. Is it ever possible that a dataset generated with its metadata during the experimental lifecycle could be removed from the facility's record?

☑ Yes ☐ No

If 'Yes', is the associated metadata retained when the dataset is removed?

☑ Yes ☐ No

Please provide any additional comments on the deletion or retention of data/metadata at your facility:

SOLEIL is committed to keeping them for 5 years, aiming for 10 years in the long term (see SOLEIL experimental data policy). To be noticed that SciCat must be filled in with the metadata of the NeXus files before deleting them

27. Do you take steps against the possibility of changes over time (e.g. evolution of community standards) that might affect the reusability of the data your facility holds?

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

☐ Yes ☑ No

If 'Yes', please provide further details of the steps you take.

Summary reflections

28. What does your facility take away from the FAIR self-assessment exercise, especially in terms of new insight into the FAIRness of your workflows and data management processes and potential avenues for future development?

We will have to take this into account further. We are still in the process of implementing the SOLEIL data policy. The exercise should be done regularly to evaluate progress.

29. Please provide feedback and suggestions for improvement of the approach used for the FAIR self-assessment (e.g. questions asked, facility coordinator role, workshops, explanatory notes provided, links to existing FAIR evaluation models, focus on workflows and processes, etc.).

NOTE: please ensure you do not include any personal identifying information (e.g. names) in your response.

The topics addressed in the questionnaire are taken into account by SOLEIL, but for some of them, our state of progress in the installation of the tools did not yet allow us to answer Yes to the questions. Suggestion to put intermediate choices as: "Planned (P)" and "Work in progress (WIP)"

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