

Adapting FAIR evaluation to Photon and Neutron facilities

Simon Lambert, Abigail McBirnie,
Brian Matthews

*Scientific Computing,
UKRI Science and Technology Facilities Council, Didcot, UK*



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

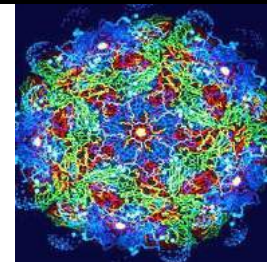
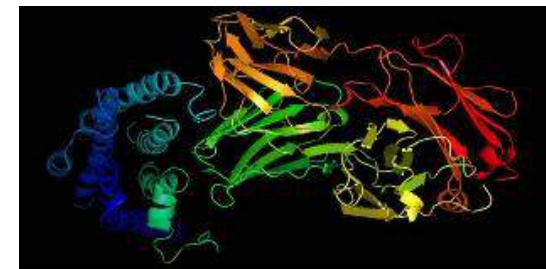


Large-Scale Analytic Facilities



Physics, chemistry, materials, life science, biochemistry, engineering, archaeology ...

- study structure and behavior of matter at small scales in high detail
 - from single atoms (10^{-10} m) to living cells (10^{-6} m) to whole systems (10^{-3} m - 1 m)



High resolution “microscopes” → Intense beams of particles → Specialist sources

These sources require large scale research infrastructures that are beyond the capability of any single university or research group



Photons (X-Rays) “see” electric charge – high atomic number nuclei
Synchrotrons, Free-Electron Lasers

10,000s of users worldwide
Across ~75 institutes
1000s of staff
1000s of experiments
10s of PB of data

Neutrons “see” nucleons – including hydrogen atoms
Reactor sources, Spallation sources



This project has received funding from the *European Union's Horizon 2020 res*



F_{indable} A_{ccessible} I_{nteroperable} R_{eusable} : a recipe for Open Science



The recommendations give a general framework for FAIR

How do we Enable FAIR Data Science within P&N RIs ??

...so we can make handling and sharing data easy

.... for humans **and** machines

... get the most out of Facilities experiments

... support verifiable and reusable science

And help our Users focus on their Science

SCIENTIFIC DATA

Amended: Addendum

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson et al.^a

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have

The FAIR guiding principles: <https://doi.org/10.1038/sdata.2016.18>

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 communication protocol
- A1.1. the protocol is free, open and universally implementable
- A1.2. the protocol allows for an authentication and authorization procedure
- 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 uses 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 data provenance
- R1.3. (meta)data meet domain relevant community standards

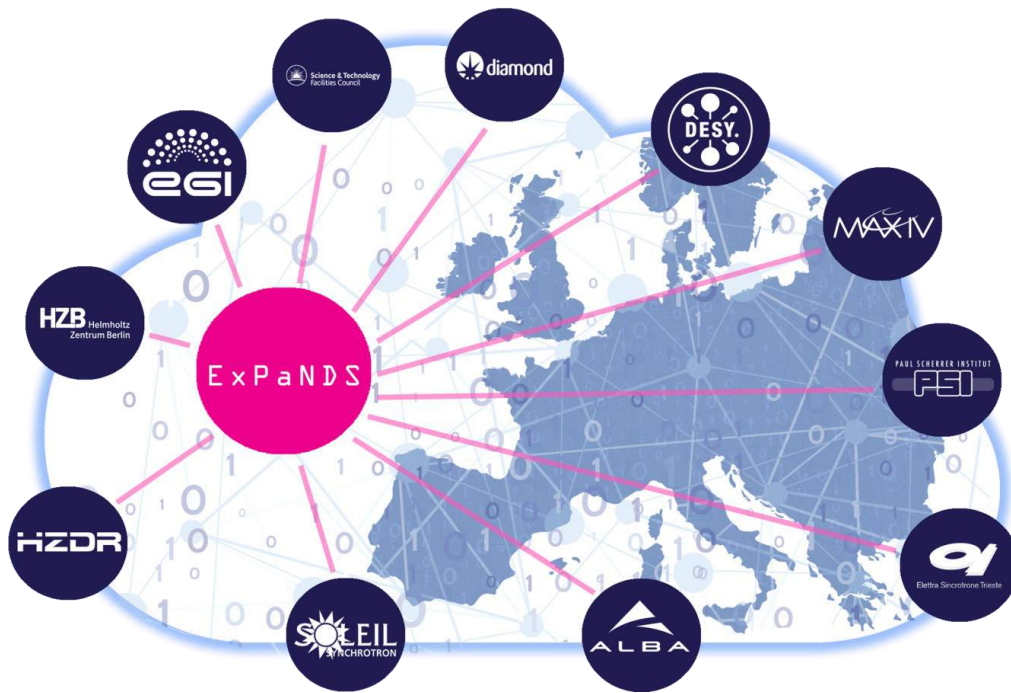
Credit/Source: SangyaPundir [CC BY-SA 4.0](https://creativecommons.org/licenses/by-sa/4.0/) (FAIR logo); Wilkinson, M. D. et al. (2015). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data*, 3:1. <https://doi.org/10.1038/sdata.2016.18> (FAIR Principles)



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 8576



ExPaNDS WP2: Enabling FAIR data PaNOSC WP2: Data Policy and Stewardship



www.expands.eu

- ESFRI Cluster project: PaNOSC (2018-22)
- Thematic project for national institutions: ExPaNDS (2019-23)
- Review and recommend the policies, practises, standards and tools which would develop best practise for FAIR data generation and use in the National Photon and Neutron RIs.
 - In the policies of the RIs
 - In the data-generation, collection and analysis process
 - In Data Management Planning
- Raising awareness and competence in FAIR data of our scientific communities.

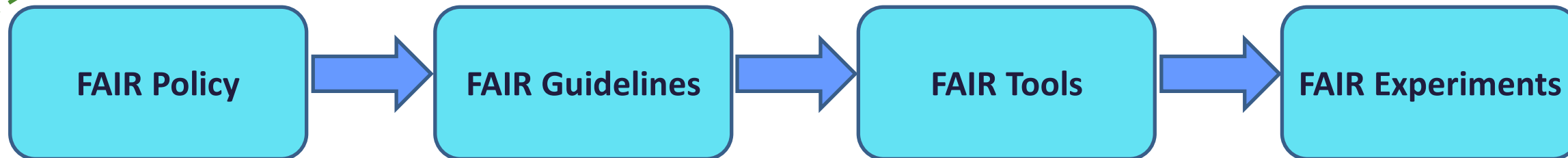


www.panosc.eu

To guide services to support FAIR-ness



A FAIR Framework for Photon and Neutron Facilities



Commitment to FAIR

How FAIR can be supported throughout the data lifecycle

Providing tools that support FAIR

How to conduct experiments FAIR-ly

FAIR Evaluation
How are we doing?



PaN Experimental Lifecycle



Constructing FAIR through an Experiment

“FAIR when it leaves the facility”

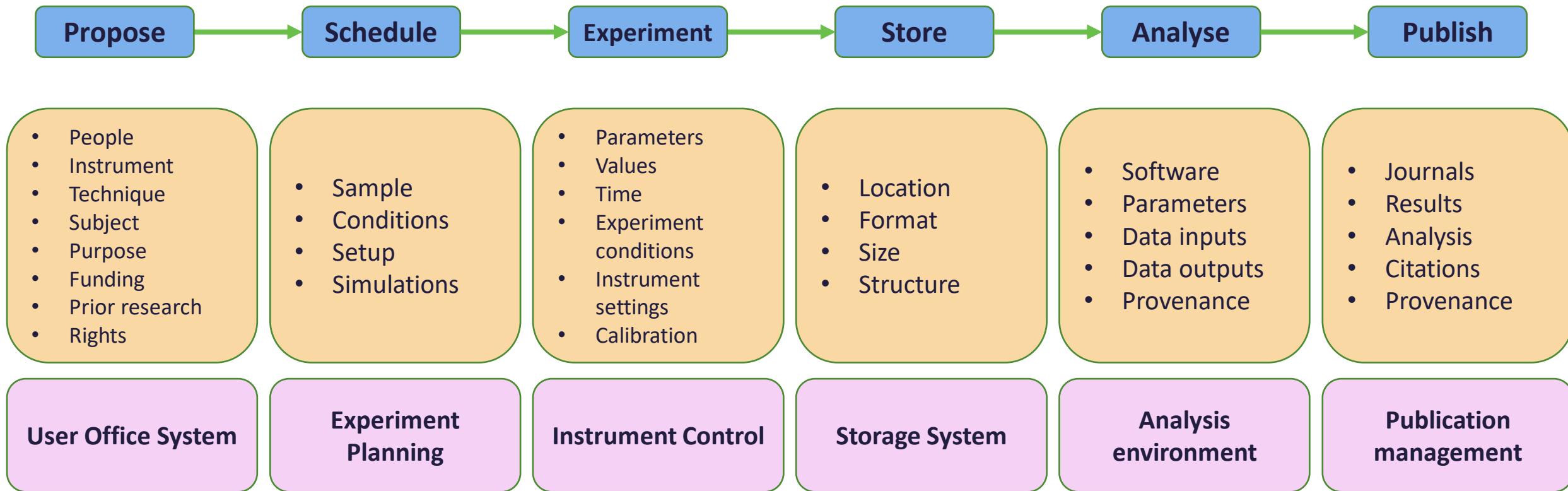


This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857641

Source: Matthews, B., Kourousias, G., Yang, E., Griffin, T. (2012). Model of the data continuum in Photon and Neutron Facilities. PaN-data 05 Deliverable 6.1. <https://doi.org/10.5281/zenodo.3897910>



FAIR Guidelines: FAIR at every step



Collect, Connect, Curate

Recommendations on Mandatory and Desirable metadata to be collected
 And how that metadata can be published with PIDs so that it is accessible in public catalogues.

Active DMPs guide how to record these for a particular experiment


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



Checking our Progress: FAIR Evaluation



- How do we know that the policies, standards and tools we have put in place will mean we achieve data which people can find, access and use?
- FAIR Evaluation
 - Test your data to see if it satisfies the FAIR Principles
- But we want to be sure that *every* experiment results in FAIR data:
 - Test your *process* to see if it includes the right things to make the data FAIR



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

Document Control Information

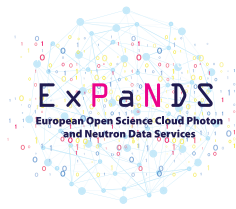
Settings	Value
Document Identifier:	D2.6
Project Title:	ExPaNDS
Work Package:	WP2
Work Package Lead	UKRI
Document Author(s):	Simon Lambert (UKRI), Abigail McBimie (UKRI), Brian Matthews (UKRI), Stephen Collins (DLS), Maik Fiedler (HZDR), Brigitte Gagey (SOLEIL), Heike Görzig (HZB), Oliver Knodel (HZDR), 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)

<https://doi.org/10.5281/zenodo.7246802>

A self evaluation of FAIR-ness for facilities



Selecting a FAIR Evaluation Method



Dimension	Range of options
Subject of the FAIR evaluation	<ul style="list-style-type: none">• Dataset• Other digital object• Data Repository• Organisation
Purpose of evaluation	<ul style="list-style-type: none">• ‘Pass/fail’ assessment for certification• Measuring progress along journey towards FAIR
Relation to FAIR principles	<ul style="list-style-type: none">• Direct metrics applied to each principle• More general relationship to FAIR
Degree of automation	<ul style="list-style-type: none">• Extensive human engagement required• Automated process



Existing FAIR Evaluation Methods

- **RDA FAIR Data Maturity Model Specification and Guidelines (2020)**
 - *'...a common set of core assessment criteria for FAIR-ness...'*
 - A set of FAIR-ness indicators and priorities that should be considered within an evaluation method.
- **Example approaches:** three developed in the FAIRsFAIR project as representative (in 2022):
 - The F-UJI tool,
 - an online service that automates the FAIR evaluation of **datasets**.
 - CoreTrustSeal+FAIRenabling
 - seeks to align CoreTrustSeal, a certification approach for trustworthy repositories,
 - FAIR evaluation at the **repository** level.
 - Assessing Capability Maturity and Engagement with FAIR Enabling Practice (ACME-FAIR)
 - is a guide aimed at **Research Performing Organisations** (RPOs),
 - *'... to help managers of Research Data Management ... services to self-assess how they are enabling researchers... to put the FAIR data principles into practice...'*
- However, none were seen as directly usable
 - *Particularly they do not consider the Data Generation Process*

Defining our method

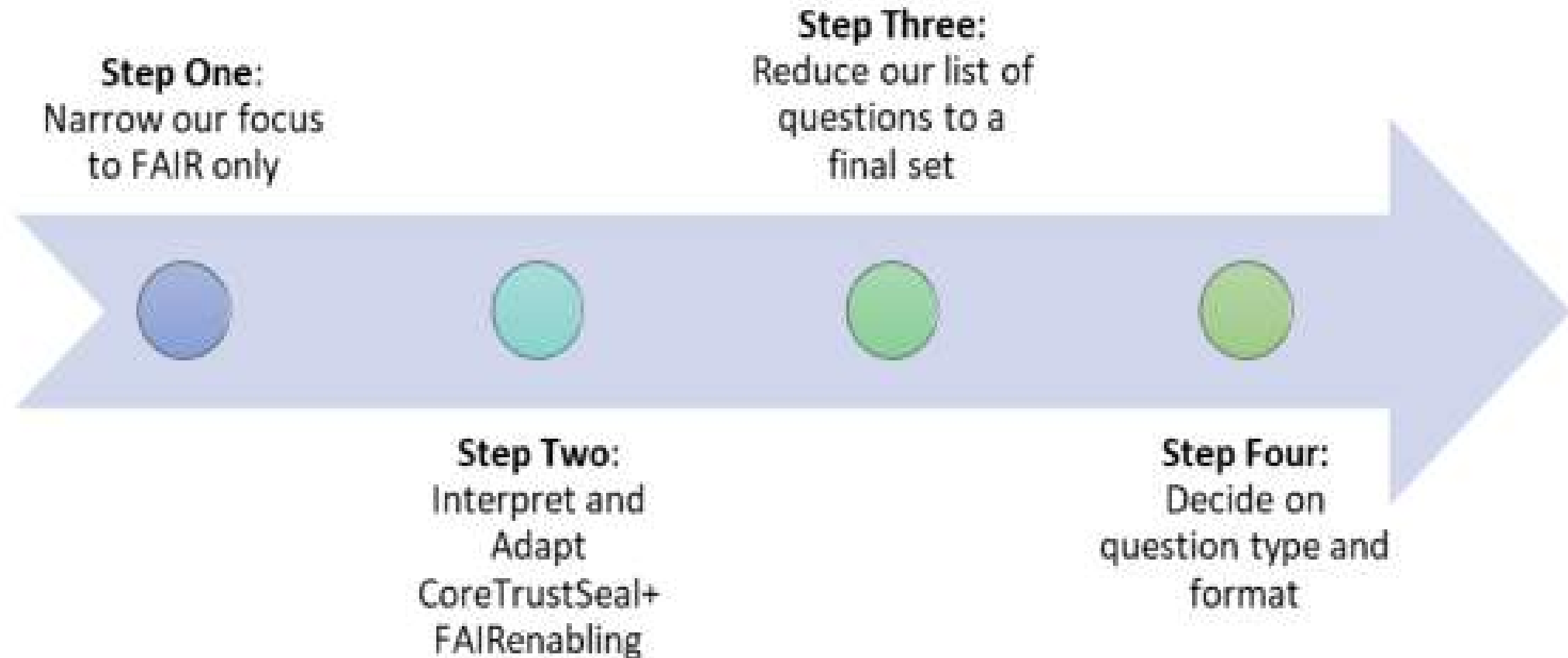
Four principles, requiring that our approach must:

1. take advantage of what existing FAIR evaluation methods have to offer;
2. link directly back to the FAIR principles;
3. take into account the relationships between existing FAIR evaluation approaches;
4. relate clearly to the processes and practices of PaN RIs.

Our method needed to be open in nature and take the form of a self-evaluation.

- avoid formal indicators that generated a metric for external assessment.
 - qualified approach that encouraged self-reflection on the FAIR-readiness of the facility.
-
- We chose **CoreTrustSeal+FAIRenabling** as our starting point.
 - Emphasis on Self-evaluation
 - CoreTrustSeal brings a emphasis on *capability maturity*,
 - reflecting the existence of reliable processes within the organisation

Developing a Questionnaire



Developing a Questionnaire

Step One:
Narrow our focus
to FAIR only

Step 1: Focus on FAIR

Narrow down to the questions which cover all the FAIR principles.

- R2: Licenses;
 - R7: Data integrity and authenticity;
 - R10: Preservation plan;
 - R13: Data discovery and identification;
 - R14: Reuse;
 - R15: Technical infrastructure;
 - R16: Security.
- *E.g. **R2:Licenses** is associated with the FAIR principle: **R1.1 (Meta)data are released with a clear and accessible data usage license.***

Developing a Questionnaire

Step One:
Narrow our focus
to FAIR only



Step Two:
Interpret and
Adapt
CoreTrustSeal+
FAIRenabling

Step Two: adapt to the PaN domain

- In language and terminology,
- In relationship to the experimental process
- e.g. in R13: Data discovery and identification:
‘What persistent identifier systems does the repository use?’

- At what stage(s) of the **experimental lifecycle** are persistent identifiers assigned at your facility? Use of familiar PaN terms
- To what **(people, places, things)** does your facility assign persistent identifiers? Elaboration to assist understanding of question
- What **specific persistent identifier service(s)** does your facility use? Knowledge that multiple services are likely to be used (e.g. DOI, ORCID)
- 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? Knowledge that many facilities have metadata catalogues
Hint at opportunities for further progress with FAIR

Developing a Questionnaire

Step 3: Reduce the question set

Produced a large set of questions
Remove those not related to FAIR or duplicates
27 questions + 2 for feedback

Step Three:
Reduce our list of questions to a final set



Step Four:
Decide on question type and format

Developing a Questionnaire

Step 4: Question type and format

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

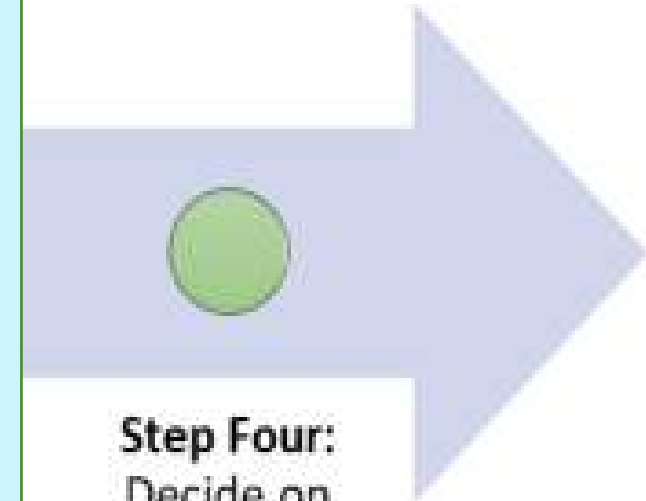
5. Does the metadata enable basic discovery (e.g. does the metadata include bibliographic information such as author, title, date, etc.)?

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?



Step Four:
Decide on
question type and
format

Balance
between
“yes/no” and
free text

Grouping and
rubric

Conducting the Evaluation

Self-evaluation is time-consuming and needs expert input.

- During the period July – September 2022, each ExPaNDS partner PaN RI undertook a FAIR self-evaluation
 - Dedicated coordinators at each facility
 - Supported by expert staff, framed by introductory and reporting workshops
- Reports in a public Deliverable (Lambert et. al. 2022, Appendix B)

Feedback

- Overall, the evaluation was well-received and conducted in an open and serious manner
 - Seen as a valuable contribution to good practice of Facilities
 - Balance of thoroughness vs. effort required.
 - Provide a baseline of FAIR-ness and identify areas of improvement.
- Areas of improvement
 - Facility vs. Instrument level assessment
 - Yes/no vs. Free-text questions
 - Indicators and metrics in the future?
 - Relationship to the wider science context



Are we getting FAIR-er ?

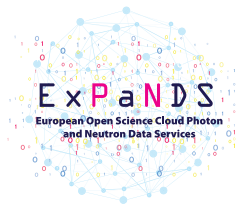
Yes!

All ExPaNDS facilities are planning to implement FAIR-related actions as a direct effect of ExPaNDS

- Developing Data Policies which promote FAIR and are regularly updated
- Assigning DOIs to Data and also other PIDs, e.g. ORCIDs
- Reviewing and extending metadata collected in metadata catalogues
- Starting to use ontologies to annotate data
- Adding data licences
- Beginning to use DMPs

So the Facilities are on the way for their FAIR journey!

Conclusions and next steps



An exercise of real and practical value

- About the right level and tailored for facilities practices and processes.
- Facilities are changing their practises to become more FAIR

Run again in future as a comparator

Available for individual facilities or via a community body

LEAPS: League of European Accelerator-based Photon Sources

Handbook on FAIR for Facilities

Method applicable to wider experimental sciences

- Provides an exemplar of deriving a method

The screenshot shows a digital interface for a handbook. On the left is a table of contents titled 'FAIR ASSESSMENT FOR PAN RIS' with a list of items, each preceded by a green checkmark. The item 'Undertaking FAIR self-evaluations at PaN RIs' is highlighted. On the right is a preview of the chapter 'Undertaking FAIR self-evaluations at PaN RIs', which includes a title, a paragraph of text, and navigation buttons for 'Previous' and 'Next'.

FAIR ASSESSMENT FOR PAN RIS	
<input checked="" type="checkbox"/>	FAIR evaluation in practice
<input checked="" type="checkbox"/>	Generic approaches
<input checked="" type="checkbox"/>	Within PaN RIs
<input checked="" type="checkbox"/>	Undertaking FAIR self-evaluations at PaN RIs
<input checked="" type="checkbox"/>	Why assess FAIRness?
<input checked="" type="checkbox"/>	Who to involve?
<input checked="" type="checkbox"/>	When to evaluate?
<input checked="" type="checkbox"/>	What to assess?
<input checked="" type="checkbox"/>	How to evaluate?
<input checked="" type="checkbox"/>	FAIR self-evaluation questionnaire and reporting template
<input checked="" type="checkbox"/>	After the evaluation

Undertaking FAIR self-evaluations at PaN RIs

It is important that FAIR self-assessment at PaN RIs focuses on facility workflows and processes. There is also an inherent recognition with such evaluation that the outcomes of the exercise will differ for each individual PaN RI. What is useful is what facilities take away for themselves, especially in terms of new insight and potential avenues for future development.

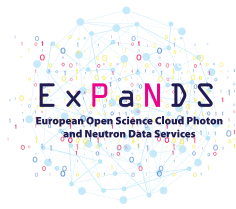
In this section of the handbook, we consider the *why*, *who*, *when*, *what* and *how* of FAIR evaluation at PaN RIs. We also emphasise the importance of taking time to reflect after the evaluation.

← Developing a PaN RI-specific evaluati... Previous

Next Why assess FAIRness? →



This project has received funding from the *European Union's Hori*



Thank You

Brian Matthews

Brian.Matthews@stfc.ac.uk

<https://expands.eu>



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

