IJDC | *Conference Paper*

Adapting FAIR Evaluation to Photon and Neutron Facilities

Simon Lambert
STFC Scientific Computing, UK

Abigail McBirnie
STFC Scientific Computing, UK

Brian Matthews
STFC Scientific Computing, UK

Abstract

The FAIR principles have become essential in establishing transparent and trustworthy research practices. However, the FAIR principles are guidelines indicating the features expected for data to be FAIR, and do not stipulate evaluation criteria. Consequently, there has been a proliferation of approaches to FAIR evaluation to substantiate claims for FAIR-ness, establish baselines, and measure improvement. Some approaches are focussed on FAIR-ness of individual datasets, others of repositories; some require extensive human evaluation, others use automation. However, within some scientific domains, data generation and management follow well-defined processes that result in datasets annotated with metadata and archived in repositories. Existing FAIR evaluation methods consider in less detail the contribution of the processes used in collecting and analysing data and how these enable FAIR-ness.

We describe the evaluation approach adopted for FAIR self-evaluation for Photon and Neutron Research Infrastructures (PaN RIs). We review selected examples of existing FAIR evaluation frameworks designed to enable assessment at different levels, and outline four dimensions that characterise them. 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, and we developed an approach drawing heavily on the original 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-evaluation a useful and valuable exercise for understanding current levels of FAIR-ness at their facilities and for articulating what implementations they have in progress or planned to support FAIR in future.

Submitted date 9 February 2024 ~ *Accepted* 20 February 2024

Correspondence should be addressed to Brian Matthews, Email: email brian.matthews@stfc.ac.uk

This paper was presented at the International Digital Curation Conference IDCC24, 19-21 February 2024

The *International Journal of Digital Curation* is an international journal committed to scholarly excellence and dedicated to the advancement of digital curation across a wide range of sectors. The IJDC is published by the University of Edinburgh on behalf of the Digital Curation Centre. ISSN: 1746-8256. URL: http://www.ijdc.net/

Introduction

The FAIR principles (Wilkinson et al, 2016) have become essential in establishing transparent and trustworthy research practices. Initiatives such as the European Open Science Cloud (EOSC) now aim that all supported data and services should be FAIR.[1] Thus data providers need to be able to validate the extent to which their data is FAIR ('FAIR-ness'). However, the FAIR principles are guidelines indicating the features expected for data to be FAIR, and do not stipulate evaluation criteria. Consequently, there has been a proliferation of approaches to FAIR evaluation to substantiate claims for FAIR-ness, establish baselines, and measure improvement. Some approaches are focussed on FAIR-ness of individual datasets, others of repositories; some require extensive human evaluation, others use automation.

However we contend that within some scientific domains, something is missing. In these domains, data generation and management follow well-defined processes that result in datasets annotated with metadata and archived in repositories. Existing FAIR evaluation methods focus on metrics for these datasets and repositories, but the crucial contribution of the processes used in collecting and analysing data in ensuring FAIR research outputs has been considered in less detail.

In this paper we consider evaluating the FAIR-ness of data derived from experiments undertaken at facilities operated by Photon and Neutron Research Infrastructures (PaN RIs), such as Synchrotrons, Neutron Sources and Free-Electron Lasers, which have in general well-defined experimental processes with systematic human and IT support. PaN RIs also need to ensure that the use of their facilities leads to transparent and reusable results. Consequently, an appropriate FAIR evaluation method would consider how those processes deliver FAIR data. Further, we would contend that our approach forms an exemplar of the FAIR evaluation of research processes and is generalisable to other domains beyond that of photon and neutron (abbreviated to PaN) facilities.

The evaluation was undertaken in the European project ExPaNDS [2] that had an aim to promote and embed the FAIR principles within the national PaN RIs across Europe by providing suitable guidelines for best practices to deliver FAIR data. The goal of the exercise was to enable ExPaNDS partners to conduct their own FAIR self-evaluation by offering a consistent approach that would be readily usable and would shed light on how their practices and processes support their journey towards FAIR.

After a brief overview of the context of photon and neutron experiments, we consider a selection of FAIR evaluation methods and discuss their merits for evaluating facilities' experiments. We then describe and outline our development of an approach to FAIR evaluation that adopts and adapts existing models to the context of PaN RIs, enabling them to evaluate the FAIR-ness of their experimental processes and workflows, using a structured questionnaire. We briefly describe the experience of applying this method within the European project ExPaNDS involving ten national PaN RIs across Europe. We conclude with observations on the effectiveness of the method and proposals for further improvement. A complete detailed description of the approach and evaluation can be found in (Lambert et. al. 2022).

Background and context

Large-scale analytic facilities operated by PaN RIs are central to advanced research which requires a detailed investigation into the nature of matter, with applications across disciplines, from physics and chemistry, materials science, to engineering and life sciences. Facilities provide

[1] European Open Science Cloud (EOSC) https://eosc.eu/eosc-about

[2] EOSC Photon and Neutron Data Services (ExPaNDS) project (2019-2023), grant agreement No 857641 of the European Union. http://expands.eu

access to sources of high-intensity particles, typically photons (e.g. X-Rays) or neutrons, that are focussed on samples of materials. The resulting interactions are measured via a variety of instruments supporting different experimental techniques (e.g. diffraction, spectroscopy or tomographic imaging) to provide information on the structure and properties of the sample, from a nano-scale view of a molecular structure, to a macro view of the composition of engineering components. As a valuable and in-demand scientific resource, access to instruments at facilities is managed via a peer-review process which allocates instrument time to end user research groups. While the science disciplines and experimental techniques vary widely, the experimental processes followed by facilities are broadly similar and can be summarised in Figure 1 (Matthews et. al. 2012).

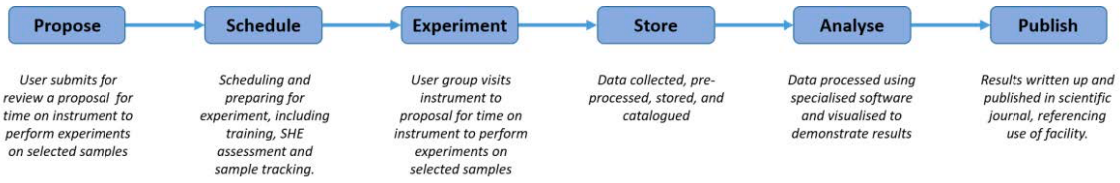

Figure 1: Lifecycle of experiments within PaN facilities

Each stage of this lifecycle is supported by facilities staff and facilities computing systems that can record appropriate information. Facilities have increasingly provided compute infrastructure to manage experimental data in these processes and have produced catalogues to allow users to access data subsequent to the experiment, including ICAT (Flannery et al. 2009) and SciCat.[3] However, PaN RIs also recognise the value of publishing FAIR data from experiments to allow validation and reuse; the computing support and systematic collection of metadata within the experimental lifecycle means that this is a practical prospect.

The ExPaNDS project was a collaboration of ten national PaN RIs from across Europe in partnership with EGI that considered the integration of computing infrastructure provided within and across facilities into the European Open Science Cloud (Boscaro-Clarke et. al. 2023). A complementary project PaNOSC [4] represented European-scale PaN RIs. A significant activity of the ExPaNDS project was to promote and embed the FAIR principles within the national PaN RIs across Europe, by providing recommendations and guidelines, and by practical implementation in the form of tools. The project developed: recommendations on developing FAIR data policy (Matthews et al, 2020, McBirnie et al 2021); guidance on metadata collection and publication to support FAIR data publication and its realisation within facilities tools and information systems (Salvat et al, 2020, Soler et al 2022); a study into the best use of Persistent Identifiers (PIDs) in those tools (Bunakov et al, 2022); and a proposition for methods and tools to support Data Management Planning within facilities processes (Görzig et. al. 2024). A further activity was a FAIR evaluation to provide a means for facilities to evaluate their workflows and processes in a systematic way to better understand what they are doing well and where there may be potential for improvement.

Identifying a Suitable Approach for FAIR Evaluation

PaN science requires the participation of PaN RIs, so their experimental processes are crucial in enabling FAIR. If their processes 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. Any evaluation should provide a means for facilities to examine these

[3] SciCat project scicatproject.github.io
[4] The Photon and Neutron Open Science Cloud (PaNOSC) https://www.panosc.eu/

processes in a systematic way to better understand what they are doing well and where there may be potential for improvement. With this in mind, we reviewed existing methods for FAIR evaluation, looking for an approach that might be suitable for our FAIR evaluation needs.

In our review, we identified that approaches to FAIR evaluation may be characterised in terms of several different dimensions, as shown in Table 1.

Table 1: Dimensions of a characterisation of FAIR evaluation methods

Dimension	Range of options
Subject of the FAIR evaluation	Dataset Other digital object Data Repository Organisation
Purpose of evaluation	'Pass/fail' assessment for certification Measuring progress along journey towards FAIR
Relation to FAIR principles	Direct metrics applied to each principle More general relationship to FAIR
Degree of automation	Extensive human engagement required Automated process

These dimensions are not independent - for example, evaluation at the level of the repository or organisation will obviously require human engagement; conversely the need for scalability implies automation at level of individual datasets. Existing FAIR evaluation methods differ with regard to the positions they adopt in relation to these various dimensions.

The RDA FAIR Data Maturity Model Specification and Guidelines (2020) seek to reduce diversity in interpretation of the FAIR principles by providing '…a common set of core assessment criteria for FAIR-ness…' The specification provides a set of FAIR-ness indicators and priorities that should be considered within an evaluation method. Most FAIR evaluation methods include aspects of this specification. However, this specification still recognises the need for flexibility and is not directly usable for our purposes.

Example approaches

While many approaches to implementing FAIR evaluation have been proposed, in our work, we examined three developed in the FAIRsFAIR project [5] as representative:

- The F-UJI tool,[6] an online service that automates the FAIR evaluation of datasets. Each dataset is evaluated against sixteen FAIR Object Assessment Metrics that draw heavily on the FAIR principles and RDA FAIR Data Maturity Model (Devaraju & Huber, 2020).
- CoreTrustSeal+FAIRenabling seeks to align CoreTrustSeal, a certification approach for trustworthy repositories, with '*repository characteristics that enable FAIR data*'. The resulting

[5] FAIRsFAIR 'Fostering FAIR Data Practices In Europe' project under the European Union's H2020-INFRAEOSC-2018-2020 grant agreement 831558. See https://www.fairsfair.eu/
[6] F-UJI: Automated FAIR Data Assessment Tool https://www.f-uji.net/

method provides an example of FAIR evaluation at the repository level (L'Hours et al, 2021).

- Assessing Capability Maturity and Engagement with FAIR Enabling Practice (ACME-FAIR) is a guide aimed at Research Performing Organisations (RPOs), with the purpose *'… to help managers of Research Data Management … services to self-assess how they are enabling researchers... to put the FAIR data principles into practice…'* (FAIRsFAIR, 2022).

Other methods we considered included FAIR EVA (Aguilar Gómez & Bernal, 2023), FAIRsharing FAIR evaluation services,[7] FAIR Enough data maturity indicators,[8] and FAIR-Checker (Gaignard et. al. 2023); similar to F-UJI, these focus on FAIR-ness of individual datasets through automated assessment.

Common aspects

Several common aspects emerged from our review of these methods. For example, the review suggested a strong leaning towards self-evaluation, with only F-UJI of those discussed above not directly incorporating some method of self-evaluation.

The review also highlighted that existing methods support FAIR evaluation at different levels: the dataset, the repository, and the organisation. However, for our case, the lack of consideration of research processes was an important gap in coverage.

Link to FAIR Principles

Importantly, given that the FAIR principles act as the foundation, the methods do all link back in some way to those principles. However, the methods articulate these links differently. Both F-UJI and Core TrustSeal+FAIRenabling link evaluation criteria directly back to specific FAIR principles. In contrast, ACME-FAIR acknowledges the FAIR principles in the collective sense but makes no reference to specific principles. This lack of explicit linking is an important omission as it is the organisational level methods that focus most strongly on evaluating practices and processes. In practice, improvements to processes depend on which details of the process need to be added, left out, or changed to enhance the FAIR-ness of the research output. In the PaN context, where facility processes are crucial to data generation, storage, and analysis, a FAIR evaluation that includes details linking back to specific principles would help PaN RIs to be the most effective in making improvements to support FAIR data.

Purpose

Existing models suggest that FAIR evaluation serves two main purposes: to provide 'pass/fail' indicators for the four components of FAIR; and to measure progress along a 'journey' towards FAIR. Given that PaN RIs are at different stages of the FAIR journey, the second approach seems most valuable and promising. Likewise, a self-evaluation approach seems appropriate for the same reason. However, no existing approach offers the means to evaluate FAIR with respect to processes of the sort that feature in the context of PaN RIs.

Consequently, it was necessary for us to select, combine, and adapt existing methods to tailor an evaluation method suited to PaN RIs.

[7] FAIRsharing and FAIR metrics groups FAIR Evaluation Services. https://fairsharing.github.io/FAIR-Evaluator-FrontEnd

[8] FAIR Enough data maturity indicators https://fair-enough.semanticscience.org/collections/fair-enough- data

Developing the FAIR Self-evaluation Questionnaire

We describe the systematic method we applied to develop a FAIR self-evaluation questionnaire for PaN RIs. To ensure that our adaptation of existing FAIR evaluation models did not simply cherry-pick what seemed most relevant, but had a more systematic basis, we applied four principles, requiring that our approach must:

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

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

This focus on self-evaluation meant that we chose to avoid formal indicators that generated a metric for external assessment. Instead, we took a more qualitative approach that encouraged self-reflection on the FAIR-readiness of the facility.

In line with these four principles, we chose CoreTrustSeal+FAIRenabling as our starting point. Though aimed at repositories, it features a strong self-evaluation component related to capability maturity, reflecting the existence of reliable processes within the organisation, and thus it could be sensibly adapted to the processes of PaN RIs. CoreTrustSeal+FAIRenabling also retains explicit links back to both the RDA FAIR Data Maturity Model and the FAIR principles. These links ensured that our method related explicitly to the FAIR principles and helped to bring necessary objectivity to our question development.

Principle 4 requires the consistent use of terminology familiar to the PaN domain: terms such as 'facilities', 'metadata catalogues', and 'experimental lifecycle'. Further, it was necessary to be aware of the distinction between the PaN RI as a whole and individual instruments within the RI, as responses might differ for each case. This ensured that the respondents felt confident in interpreting the questions.

The four-step process used to develop the questionnaire is illustrated in Figure 2.

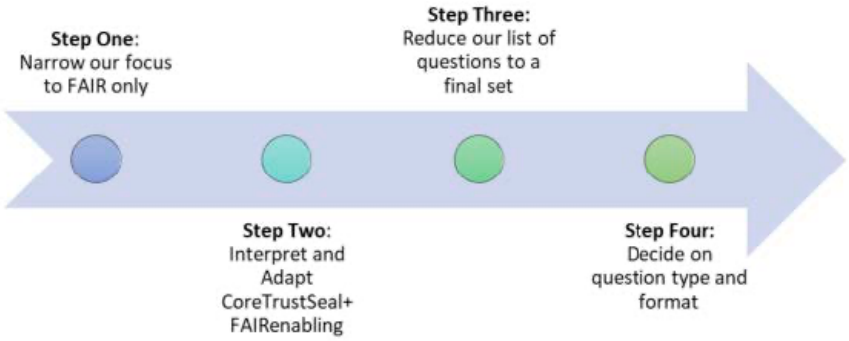

Figure 2: The process of developing the self-evaluation questionnaire

Step One: Focus on the +FAIRenabling components of CoreTrustSeal+FAIRenabling

The aim was to develop an evaluation specifically for FAIR rather than of general institutional practices so we concentrated on the seven CoreTrustSeal requirements that included a +FAIRenabling element:

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

CoreTrustSeal+FAIRenabling associates each of these elements with one or more FAIR principles. For example, *R2:Licenses* is associated with the FAIR principle: *R1.1 (Meta)data are released with a clear and accessible data usage license*. The seven +FAIRenabling elements address all fifteen FAIR principles, and are also correlated with indicators from the RDA FAIR Data Maturity Model.

Step Two: Adapt CoreTrustSeal+FAIRenabling.

The second step was to match the method to the PaN domain. This entailed the use of domain terms, and reworking the requirements and questions from CoreTrustSeal+FAIRenabling. Formulating these PaN-specific questions necessarily required judgement to some extent, drawing on our experience of FAIR, the PaN experimental lifecycle, and the processes of PaN RIs.

The factors that were taken into account in this step included:

- understanding what the CoreTrustSeal+FAIRenabling requirements are really probing, and ensuring that is reflected in the adapted question set;
- checking coverage against the FAIR principles (this was already noted above) and the RDA indicators, and determining whether there are any 'missing' indicators to be introduced;
- taking account of the operation of the PaN facilities, including any local variations in practice at the level of a particular instrument, for example: organisational context, processes and workflows, data and metadata;
- framing the questions for the intended respondents, bearing in mind that the questions needed to be suitable for self-evaluation, not too open-ended, and with clear relevance to the respondents' local environment and practices.

As a specific example, consider the case of R13: Data discovery and identification. Among the questions intended to guide submission of evidence for CoreTrustSeal is:

> 'What persistent identifier systems does the repository use?'

After the processes above, several related PaN-oriented questions were produced. These are shown in Figure 3, along with annotations explaining the reasoning that led to the expansion and adaptation of the one initial question.

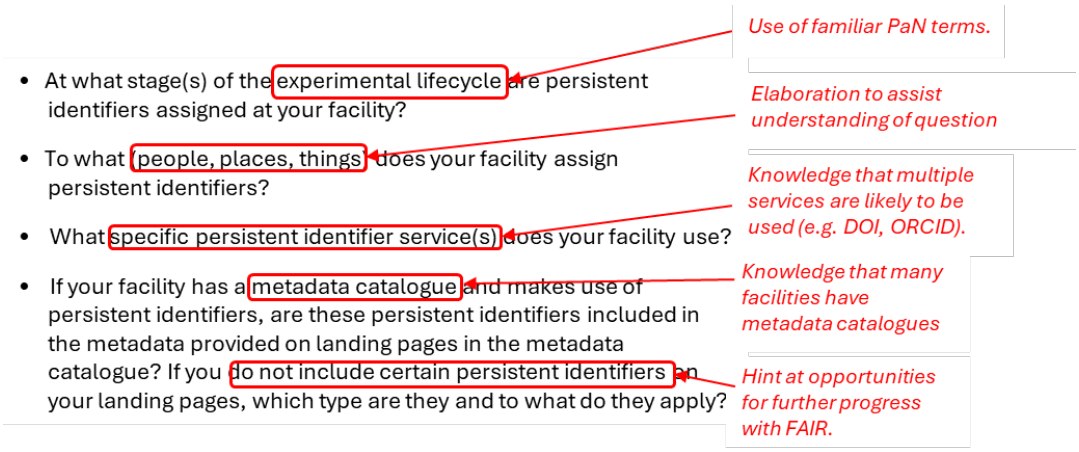

Figure 3: Adapting questions for the domain

Step Three: Reduce the Question Set

Step two produced a large set of questions, which needed to be reduced to a more focused set. Questions were omitted or combined if they did not address a FAIR Principle (for example, questions about data citation from CoreTrustSeal); or if they were addressing the same RDA indicators. A final set of 27 questions resulted, plus two extra for reflection on the self-evaluation itself.

Step Four: Question Type and Format

A careful balance was needed between binary versus free text questions. Too many free text answers risked overburdening respondents, but excessive use of yes/no or multiple-choice options might restrict reflective and thoughtful responses. The questions were grouped under seven broad topics 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.

Figure 4 is an extract from the final questionnaire, showing some questions under the 'search' grouping, with yes/no answers but also an expandable text box for additional information.

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

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

☐ Yes ☐ No

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

Figure 4: An extract from self-evaluation questionnaire

The FAIR evaluation in practice

During the period July – September 2022, each ExPaNDS partner PaN RI undertook a FAIR self-evaluation of their facility's processes, using our questionnaire.

Because of the nature of experiments at PaN RIs, the processes and workflows involved are often spread across different teams and systems (i.e. local instrument team and systems, facility-wide computing resources). As such, a number of people would need to input to the questionnaire at each facility. Thus, the self-evaluation was led in each facility by a volunteer 'facility coordinator' with the responsibility to gather responses from multiple colleagues across their facility. Two workshops and one-to-one support offered opportunities for engagement with these facility coordinators and other colleagues during the exercise.

Each facility provided their responses to the questionnaire using the same reporting template. There was a 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 insights and potential avenues for future development.

As the FAIR evaluation was intended to be open in nature, we published these reports in a public Deliverable (Lambert et. al. 2022, Appendix B); we refer the interested reader to the Deliverable for details of the results. As part of the exercise, we asked for feedback on the evaluation process itself. We reflect on that feedback below.

Overall impressions

Overall feedback from the ExPaNDS partners strongly indicated that they found the FAIR self-evaluation a useful undertaking. Many participants commented that the results provided a good baseline for understanding their facility's current level of FAIR-ness.

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-evaluation on an annual basis for internal use at their facilities.

Partners emphasised that they felt that the FAIR self-evaluation could serve a dual purpose. On the one hand, they felt the self-evaluation provided good information to establish a baseline on their current state of FAIR-ness, highlighting what was in place and what was not; on the other hand, partners were able to use the self-evaluation to better understand their planned journey towards increasing FAIR-ness and to articulate what was in progress and how far along that work was. This observation echoes the two different evaluation perspectives (i.e. 'pass/fail' and 'measuring progress') supported by existing FAIR evaluation tools.

What Worked Well

ExPaNDS partners felt that several aspects of the FAIR self- evaluation 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.

A key concern in the design of the FAIR self-evaluation 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; partners felt that the level of commitment required and the time and resource needed to complete the FAIR evaluation was appropriate, but that at the same time, the exercise was useful and successful.

Participants in the FAIR self-evaluation 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 evaluation coordinators by the team leading for the FAIR self-evaluation task was helpful. The workshops offered additional opportunities for help with any outstanding issues.

As explained above, we avoided the use of formal indicators in the FAIR self-evaluation. We sought to promote the exercise as an opportunity for partners to reflect on issues related to FAIR and their progress on addressing these. Feedback from partners suggested that they, too, felt not including formal indicators in the evaluation was the right course of action, especially as this was the first FAIR self-evaluation 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, partners welcomed the free text options and flexibility of the template and felt that these encouraged reflection on FAIR-ness levels, whereas the use of formal indicators could have been discouraging at this stage.

Suggestions for Changes and Improvements

The FAIR evaluation 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.

Several ExPaNDS partners remarked on the binary 'yes/no' approach used for many of the questions, echoing aspects of a 'pass/fail' evaluation method. While partners did acknowledge the value of simple 'yes/no' questions for a summary of status, and see their potential usefulness, for example, in establishing a baseline, they also felt they 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 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 interesting because the evaluation design had sought to reduce the burden on participants by not requiring too many free text answers. Importantly though, a careful balance needs to be struck. There also remains the open question as to whether partners, if they had not had to choose from the binary option, would have as clearly been able to identify the aspects of FAIR that needed further attention.

An example where a 'yes/no' answer was not felt to be appropriate was on 'essential' metadata, that is core metadata that is seen to be necessary for a FAIR documentation of the experiment. This question took the form:

> 'Looking at the metadata listed as essential in the ExPaNDS metadata framework across all stages 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]'

Because the answer applied to all 31 essential fields specified in the ExPaNDS metadata framework (Salvat et. al. 2020) 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 FAIR self-evaluation, and indeed only one facility answered 'yes' and that was qualified in the free text box. Further, participants found it hard to refer back to the definitions of the essential metadata; this emphasises the need to link the evaluation method to associated documentation relevant to the domain.

For PaN RIs, the Context is Not Always PaN

The FAIR self-evaluation task was designed to focus on PaN RIs. However, ExPaNDS partners operate in a wider disciplinary context, for example, where a PaN facility is a part of a wider science institution or domain community with their own standards and conventions. In such cases, issues such as what metadata is required by the parent science institution's data repository may 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-evaluation 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-evaluation.

Concluding Remarks

With the completion of the ExPaNDS project, the ongoing maintenance and re-application of the evaluation method has been passed back to the PaN community, as part of the PaN FAIR implementation framework to be taken further by initiatives such as LEAPS, a consortium of European advanced photon sources.[9]

[9] LEAPS – the League of European Accelerator-based Photon Sources https://leaps-initiative.eu

The overall purpose of the PaN FAIR self-evaluation 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 processes. A key aspect used to judge the success of the FAIR self-evaluation exercise was its usefulness (or not) to ExPaNDS partner facilities.

For some facilities at an early stage in their FAIR journey, the exercise provided a baseline from which to evaluate in future; for others, further along in their journey, the exercise served as a way to assess progress. In all cases, the self-evaluation was 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 FAIR-ness and highlighting possibilities for future development. 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 FAIR-ness: to establish a baseline and to measure progress.

More broadly, we hope that our experience will make a significant contribution to the still developing area of FAIR evaluation. Given the multiplicity of approaches that already exist, it might be doubted that another is needed. However, the simplicity and attractiveness of the FAIR principles does not translate into self-evident mechanisms for evaluating FAIR-ness, and there are diverse needs and expectations for what such evaluation might offer. We believe that our approach addresses an important gap and could be adapted to other domains which have well- defined experimental workflows and experimental lifecycles where machines and automated processes play a major role. It should be possible to follow the steps presented in Figure 2, taking account of the state of FAIR advancement in the domain and any particular characteristics of its practices and processes, leading to a self-evaluation that will be as illuminating as the one we conducted for the PaN RIs.

The contribution of this work concerns the integration of FAIR into the processes of science, which are often long-established and familiar within particular domains. Where processes are the focal point, there seems no purpose in undertaking a FAIR evaluation if it cannot be useful in some practical way and lead to definite actions that could be taken to enhance the value of those processes. In this context, details matter, as does a considered reflection about what might need to be changed, and how that would be done. The usefulness of a FAIR evaluation is directly tied to its end product, with the combination of actionable detail and useful insight as the important outcomes.

Acknowledgements

This work was supported by the EOSC Photon and Neutron Data Services (ExPaNDS) project, grant agreement No 857641 of European Union's Horizon 2020 research and innovation programme. We would like to thank our partners and contributors from this project who participated in the self-assessment exercise.

References

Aguilar Gómez, F., Bernal, I. (2023). FAIR EVA: Bringing institutional multidisciplinary repositories into the FAIR picture. Sci Data 10, 764 https://doi.org/10.1038/s41597-023-02652-8

Boscaro-Clarke, I., Fuhrmann, P., Marauska, J., Matthews, B., McBirnie, A., Ashton, A., … Millar, P. (2023). ExPaNDS: Laying the Foundations for Achieving Open Science for Everyone. Synchrotron Radiation News, 36(1), 25–28. https://doi.org/10.1080/08940886.2023.2186664

Bunakov, V., Krahl, R., Matthews, B., Vizcaino, N., & Vukolov, A., (2022). Advanced infrastructure for PIDs in Photon and Neutron RIs. ExPaNDS deliverable D2.5. https://doi.org/10.5281/zenodo.5905351

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

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

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

Flannery, D. et al. (2009). ICAT: Integrating Data Infrastructure for Facilities Based Science. In Fifth IEEE International Conference on e-Science, Oxford, UK, 2009, pp. 201- 207, https://doi.org/10.1109/e-Science.2009.36

Gaignard, A., Rosnet, T., De Lamotte, F. et al. (2023). FAIR-Checker: supporting digital resource findability and reuse with Knowledge Graphs and Semantic Web standards. J Biomedical Semantics 14, 7. https://doi.org/10.1186/s13326-023-00289-5

Görzig, H., Gonzalez-Beltran, A., Engel, F., & Matthews, B. (2024). A Framework for Active DMPs in Photon and Neutron Science Large-Scale Facilities. Data Science Journal, 23: 4, pp. 1–18. DOI: https://doi.org/10.5334/dsj-2024-004

Lambert, S., McBirnie, A., Matthews, B., Collins, S., Fiedler, M., Gagey, B., Görzig, H., Knodel, O., Kourousias, G., Minotti, C., Markvardsen, A., Ounsy, M., Pozsa, K., Schön, S., Soler, N., Spruce, D., & Vukolov, A. (2022). Self-evaluation Photon and Neutron RIs for FAIR data certification. ExPaNDS Deliverable 2.6. Zenodo. https://doi.org/10.5281/zenodo.7246802

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

Matthews, B., Kourousias, G., Yang, E., & Griffin, T. (2012). Model of the data continuum in Photon and Neutron Facilities. PaN-Data Open Data Infrastructure Deliverable D6.1. https://doi.org/10.5281/zenodo.3897190

Matthews, B., McBirnie, A., Vukolov, A., Ashtun, A., Collins, S., Da Graca Ramos, S., Gagey, B., Gonzalez-Beltran, A., Johnsson, M., Krahl, R., Ousny, M., & Van Daalen, M., (2020). Draft extended data policy framework for Photon and Neutron RIs. ExPaNDS deliverable D2.1 https://doi.org/10.5281/zenodo.4014811

McBirnie, A., Matthews, B., Gagey, B., Minotti, C., Salvat, D., Schlünzen, F., & Vukolov, A. (2021) Final data policy framework for Photon and Neutron RIs. ExPaNDS deliverable D2.3 https://doi.org/10.5281/zenodo.5205825

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

Salvat, D, Gonzalez-Beltran, A., Görzig, H., Matthews, B., McBirnie, A., Ousny, M., Da Graca Ramos, S., & Vukolov, A., (2020). Draft recommendations for FAIR Photon and Neutron data management. ExPaNDS deliverable D2.2. https://doi.org/10.5281/zenodo.4312825

Soler, N, McBirnie, A., Gonzalez-Beltran, A., Vukolov, A., Minotti, C., Görzig, H., Pozsa, C., Matthews, B., &. Ousny, M., (2022). Final recommendations for FAIR Photon and Neutron data management. ExPaNDS Deliverable D2.7 https://doi.org/10.5281/zenodo.6821676

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018. https://doi.org/10.1038/sdata.2016.18