A Framework Enabling FAIR Facilities

Brian Matthews[1], Abigail McBirnie[1], Alejandra Gonzalez-Beltran[1], Simon Lambert[1], Vasily Bunakov[1], Nicolas Soler[2], Heike Görzig[3], Rolf Krahl[3], Majid Ounsy[4], Brigitte Gagey[4], Andrey Vukolov[5], Patrick Fuhrmann[6], Sophie Servan[6], Paul Millar[6]

[1] *UKRI Science and Technology Facilities Council, Didcot, UK*
[2] *ALBA-CELLS Synchrotron, Cerdanyola del Vallès, Spain*
[3]*Helmholtz-Zentrum Berlin für Materialien und Energie, Berlin, Germany*
[4]*Synchrotron SOLEIL, Saint-Aubin, France*
[5]*Elettra Sincrotrone Trieste, Basovizza, Trieste, Italy*
[6]*Deutsches Elektronen-Synchrotron (DESY), Hamburg, Germany*

Introduction

The FAIR principles are recognised as a practical framework for data sharing within research communities. Funders have adopted FAIR principles, notably within the European Open Science Cloud programme of the European Union. However, the principles are couched in broad terms, and interpretation and practical guidance are required for their adoption within different research communities.

Photon and neutron (PaN) science is a cross-disciplinary research community centred around large-scale facilities providing specialised instruments to study the structure of matter using advanced x-ray or neutron sources. Facilities recognise the FAIR principles to enhance the potential for reuse of experimental data and to give users the best opportunity to further their science, especially as the size and complexity of data increases. However, introducing the principles into the practices of facilities requires careful planning.

The EU project ExPaNDS[1] explored how FAIR data can be enabled within PaN facilities, working closely with its sister project PaNOSC[2]. ExPaNDS provides a framework of guidelines that facilities can use to ensure that data arising from experiments is FAIR and suitable for sharing and reuse, as well as easier for the experimenters themselves to use. In this paper, we introduce this framework.

A Framework Enabling FAIR within Facilities

The components of the framework are illustrated in Figure 1, and described below.

FAIR Policy. PaN Facilities issue data policies describing the facility's approach to the management, and publication of experimental data, and the rights and obligations of users. The ExPaNDS framework considers how policies should be expressed so that the experimental data in scope should be *FAIR when it leaves the Facility [1,3]*. This gives a commitment to the facility to uphold the FAIR principles. This is further expanded in a set of 21 policy principles for the facility to consider when drawing up its data policy.
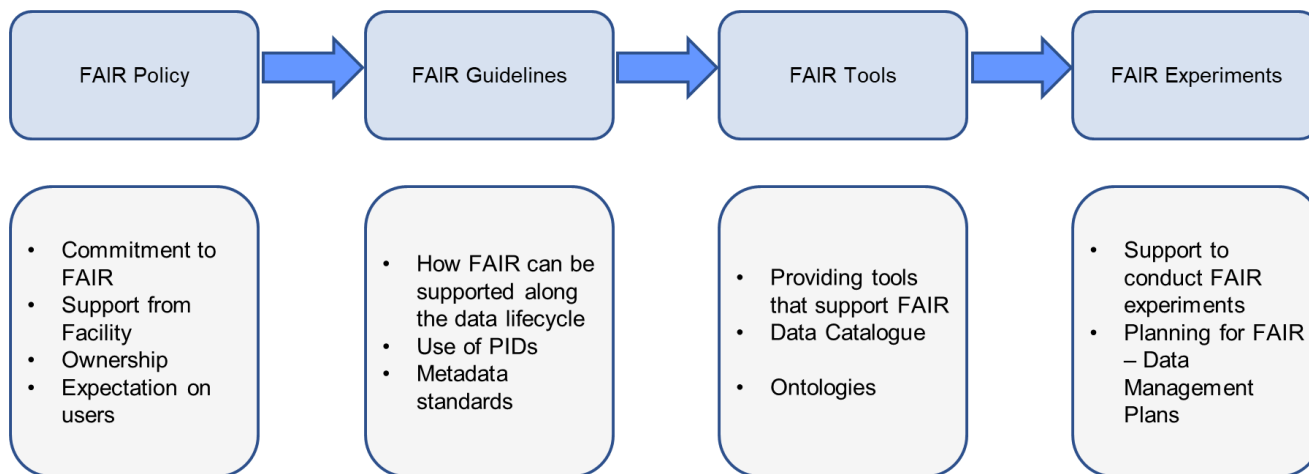
Figure 1: Framework Components

FAIR Guidelines. Experiments within facilities are incredibly diverse in terms of instruments and techniques used, science discipline and in the volume and nature of data collected; this makes comprehensive FAIR data guidelines difficult. However, facilities follow a broadly similar experimental process of users submitting proposals, performing experiments during a visit, and analysing and publishing results. Each stage is supported by IT systems, giving the opportunity for the facility to collect metadata during the process. The ExPaNDS guidelines [2,7] recommend the mandatory and desirable metadata that should be collected at each stage to ensure a FAIR data description of the experimental data, as illustrated in Figure 2.
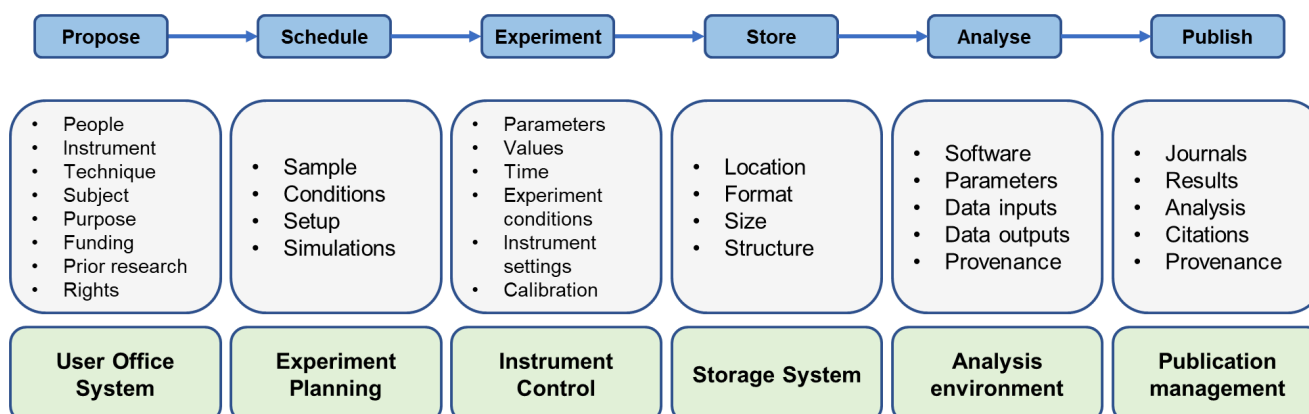

Figure 2: Experimental process and metadata within information systems

The guidelines describe mappings to standards, such as the DataCite schema, and how these mappings can be used for aggregators like B2Find, which enhances data Findability.

Further, Persistent Identifiers are a key aspect of FAIR data, and the framework includes recommendations on use of Persistent Identifiers (PIDs) [5], for data, but also emerging standards for resources including instruments and samples that are of value to the PaN community.

FAIR Tools. FAIR data practices need to be embedded in software tools. Facilities have developed common approaches to data cataloging in the ICAT[3] and SciCAT[4] toolkits used in several European facilities. ExPaNDS extends these to support data federation via a common search API and OAI-PMH interface, conforming to FAIR framework recommendations [10]. Additionally, new shared

vocabularies were identified to support FAIR data, including PaN Experimental Techniques (PaNET), an ontology of analytic techniques supported by facilities [9]. However, techniques are very diverse, and we need additional information to fully describe data to make it FAIR.

FAIR Experiments. For a particular experiment we need the details of the environment and instrument taking the measurements, the data produced, the post-processing undertaken, and this information needs to be supplied by the experimental team. The ExPaNDS framework advocates Data Management Plans (DMPs) to identify additional attributes to make data FAIR. This can be burdensome: the DMP questionnaire proposed by PaNOSC identified 115 questions to give a complete picture of experimental data [11]. However, in practice many of these depend on the instrument and computing infrastructure, and standardised responses can be supplied, drastically reducing the user input. The ExPaNDS framework discusses this within guidelines on the implementation and use of DMPs [4,8].

FAIR Assessment

ExPaNDS provides guidelines on the implementation of FAIR within PaN facilities [6]. To measure the progress of facilities towards achieving FAIR data, ExPaNDS developed a FAIR assessment method. This considered existing methods, such as the RDA FAIR Data Maturity model [12]. However, these methods focus on a dataset's conformance to the principles. Facilities generate 1000s of datasets per year, and we would like to ensure that every experiment generates FAIR data. Thus the method tailors best practice to evaluate whether the data-generation *process* results in FAIR data, using 29 self-assessment questions.

Conclusions

ExPaNDS has developed a framework of guidelines to enable FAIR data within PaN facilities ranging from data policy to experimental data. This has had clear impact, with all ExPaNDS facilities planning to implement FAIR-related actions as a response, including modifying Data Policies to promote FAIR, extending the use of PIDs, and regularising the metadata collected in catalogues. The use of the PaNET ontology offers the prospect of shared domain specific descriptions. The development of the framework is still active. It is being reworked into a set of simpler guidelines within the context of the FAIR-Impact[5] FAIR Implementation Framework and further collaboration is being proposed to maintain the framework within the PaN community.

Acknowledgements

References

[1] ExPaNDS Deliverable D2.1 Draft extended data policy framework for Photon and Neutron RIs (Sep 2020). https://doi.org/10.5281/zenodo.4014811

---

[5] https://fair-impact.eu/

[2] ExPaNDS Deliverable D2.2 Draft recommendations for FAIR Photon and Neutron data management (Dec 2020). https://doi.org/10.5281/zenodo.4312825

[3] ExPaNDS Deliverable D2.3 Final data policy framework for Photon and Neutron RIs (Aug 2021). https://doi.org/10.5281/zenodo.5205825

[4] ExPaNDS Deliverable D2.4 DMPs for Photon and Neutron RIs (Nov 2021). https://doi.org/10.5281/zenodo.5636096

[5] ExPaNDS Deliverable D2.5 Advanced infrastructure for PIDs in Photon and Neutron RIs (Mar 2022). https://doi.org/10.5281/zenodo.5905351

[6] ExPaNDS Deliverable D2.6 Self-evaluation Photon and Neutron RIs for FAIR data certification (Dec 2022). https://doi.org/10.5281/zenodo.7246802

[7] ExPaNDS Deliverable D2.7 Final recommendations for FAIR Photon and Neutron data management (Jul 2022). https://doi.org/10.5281/zenodo.6821676

[8] ExPaNDS Deliverable D2.8 Active DMPs for Photon and Neutron RIs (Dec 2022). https://doi.org/10.5281/zenodo.7223438

[9] ExPaNDS Deliverable D3.2 ExPaNDS ontologies v1.0 (Jun 2021) https://doi.org/10.5281/zenodo.4806026

[10] ExPaNDS Deliverable D3.3 Demonstrate ICAT and SciCat released with APIs compatible with ExPaNDS federated EOSC services (Mar 2022) https://doi.org/10.5281/zenodo.6363591

[11] PaNOSC Deliverable D2.2 DMP Template for facility users (Nov 2021) https://doi.org/10.5281/zenodo.5639428

[12] Research Data Alliance Recommendation FAIR Data Maturity Model: specification and guidelines (Jun 2020) https://doi.org/10.15497/rda00050