PAPER • **OPEN ACCESS**

DAPHNE4NFDI - Improving Research data management at Synchrotron Facilities

Journal of Physics: Conference Series **3010** (2025) 012133 doi:10.1088/1742-6596/3010/1/012133

DAPHNE4NFDI - Improving Research data management at Synchrotron Facilities

Lisa Amelung[1*], Anton Barty[1], Bridget M. Murphy[2,3], Jan-Dierk Grunwaldt[4], Svenja Hövelmann[2,3], Aliaksandr Leonau[5,6], Sebastian Paripsa[7], Astrid Schneidewind[8], Sebastian Busch[9], Christian Gutt[5], Wiebke Lohstroh[10], Frank Schreiber[11] and Tobias Unruh[12]

[1] FS-Scientific Computing, Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany
[2] Institute of Experimental and Applied Physics, Kiel University, Kiel, Germany
[3] Ruprecht Haensel Laboratory, Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany
[4] Institute for Chemical Technology and Polymer Chemistry and Institute of Catalysis Research and Technology, Karlsruhe Institute of Technology, Karlsruhe, Germany
[5] Department Physik, Universität Siegen, Siegen, Germany
[6] Materials Imaging and Dynamics, European XFEL GmbH, Schenefeld, Germany
[7] Condensed matter - X-ray physics, University of Wuppertal, Wuppertal, Germany
[8] Jülich Center for Neutron Science at Heinz Maier-Leibnitz Zentrum (MLZ), Forschungszentrum Jülich GmbH, Garching, Germany
[9] German Engineering Materials Science Centre (GEMS) at Heinz Maier-Leibnitz Zentrum (MLZ), Helmholtz-Zentrum Hereon GmbH, Garching, Germany
[10] Heinz Maier-Leibnitz Zentrum (MLZ), Technische Universität München, Garching, Germany
[11] Institut für Angewandte Physik, Universität Tübingen, Tübingen Germany
[12] Institute for Crystallography and Structural Physics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

[*]E-mail: lisa.amelung@desy.de

Abstract. Advancements in synchrotron and X-ray free-electron sources and associated developments in instrumentation and techniques offer many new possibilities for researchers. At the same time there is increasing demand and pressure to implement the FAIR data principles by which data should be made Findable, Accessible, Interoperable and Reusable. The consortium DAPHNE4NFDI (DAta from PHoton and Neutron Experiments for NFDI) addresses this challenge within the German National Research Data Infrastructure (NFDI), and also in relation to European / worldwide initiatives. This article gives an overview of our activities and elaborates on our progress, showcasing developments in some of our use cases, including: (1) X-ray reflectivity (XRR), (2) X-ray photon correlation spectroscopy (XPCS) and (3) X-ray absorption spectroscopy (XAS).

1. DAta from PHoton and Neutron Experiments (DAPHNE4NFDI)

DAta from PHoton and Neutron Experiments (DAPHNE4NFDI) is the basis for many scientific discoveries in many different fields and the analytic methods used to obtain experimental results in the field of X-rays and Neutrons (Photons and Neutrons – PaN) extend far beyond the physics or material science community. Instead, the continuously developed instrumentation and

techniques enable and support a different level of research in many different domains. Therefore, the DAPHNE4NFDI research community impacts on global challenges, such as health, energy and environment.

DAPHNE4NFDI (https://www.daphne4nfdi.de) is funded by the German Research Foundation (DFG) as one out of 26 consortia of the National Research Data Infrastructure (NFDI) and compromises 18 funded partners, among them seven large-scale research facilities (LSFs) and 11 universities within Germany, as well as more than 100 participating institutions and individuals within Germany, Europe and across the world.

The consortium is based on a close collaboration between the user community and the photon and neutron facilities and is coordinated by Deutsches Elektronen-Synchrotron (DESY). It brings together IT-specialists, scientists and users (Barty, A. *et al.* 2023).

2. The common needs and challenges – implementation of the FAIR data principles

Advancements in instrumentation and techniques go along with higher complexity and rapidly increasing amounts of data. At the same time, more and more funding agencies ask for the implementation of research data management (RDM) practices to future proposals and make e.g. data management plans a mandatory part of newly funded projects.

To handle these common needs and challenges in the field of PaN data, DAPHNE4NFDI aims at implementing the FAIR data principles thus making research data Findable, Accessible, Interoperable and Reusable - FAIR (Wilkinson, M. D. *et al.* 2016).

Following the data life cycle, we have established six different task areas (TAs) that work hand-in-hand and across facilities/institutions to develop the technical framework and tools (TA1-TA3) for Managing Data Production (TA1), (Meta)data repositories & catalogues (TA2) and Infrastructure for data & software reuse (TA3). In addition, Outreach and Dissemination (TA4) as well as External Communication and Policy (TA5) are important aspects of our work (see Figure 1 for more information). All TAs are supported in terms of Project Management (TA6) and meet online on a monthly basis. Once per year, every technical TA meets for a Face2Face workshop with hands-on sessions. In addition, all project members and collaborators meet in person on an annual basis to secure cross-TA activities.

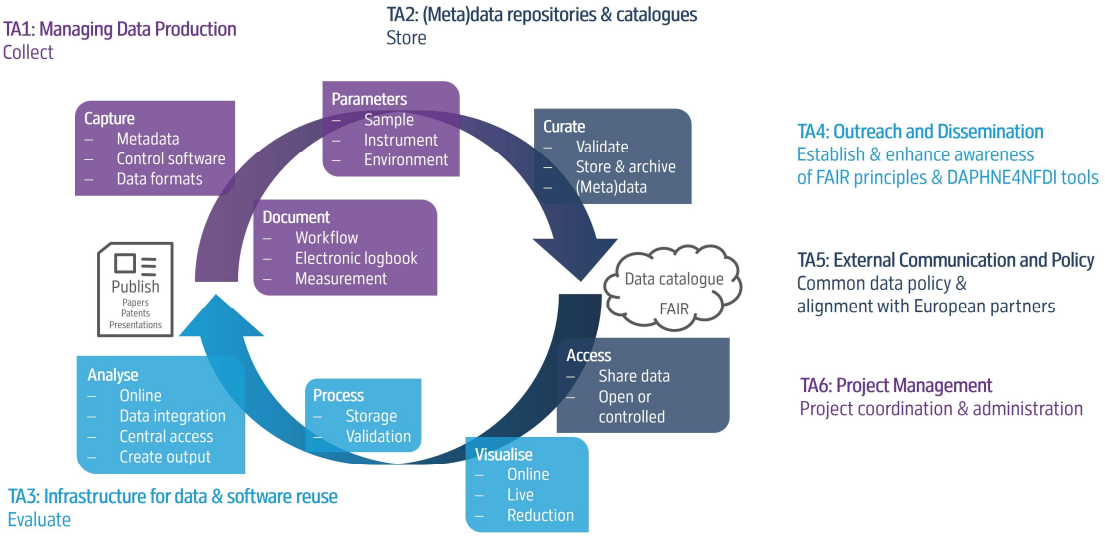

Figure 1. The interaction of the six different task areas (TA)s in DAPHNE4NFDI.

3. Our use cases serve as flagship projects

An important part of the work within DAPHNE4NFDI is realized in the 11 orthogonal use cases (Figure 2) where power users focused on specific scientific methods are an excellent means to engage the community and serve as flagship projects. Working closely with beamline staff, IT experts and the central computing departments, the use cases make significant progress in terms of e.g. standardized metadata formats, searchable catalogues and open data repositories. In general, they are a valuable vehicle to initiate and test FAIR concepts and introduce them to and discuss them with the user community in an early stage of development.

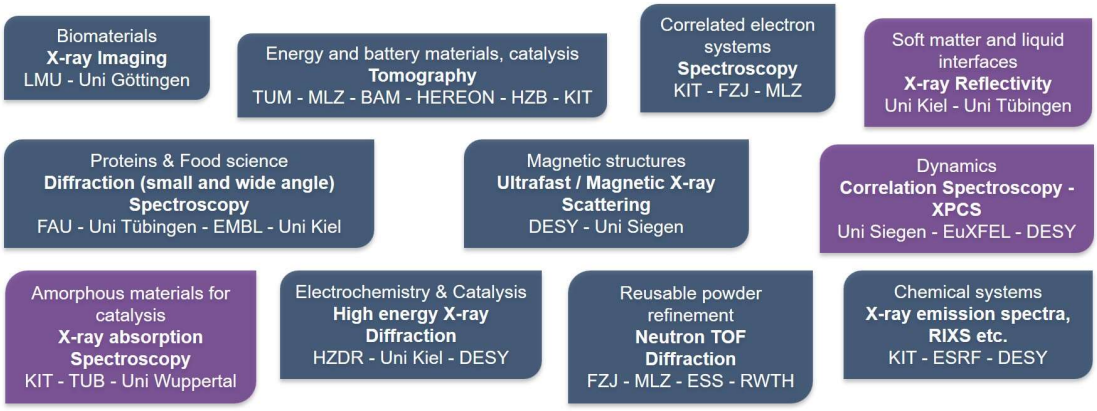

Figure 2. To date, DAPHNE4NFDI has 11 orthogonal use cases. Three of them are highlighted in purple and are explained in detail in the text.

3.1 X-ray reflectivity (XRR)

Neutron and X-ray reflectivity methods allow us to study surface properties such as layer thickness, density and roughness on the nanoscale. The XRR use case focuses on creating a FAIR data workflow for X-ray and Neutron reflectivity measurements studying soft matter and liquid interfaces (Figure 3).

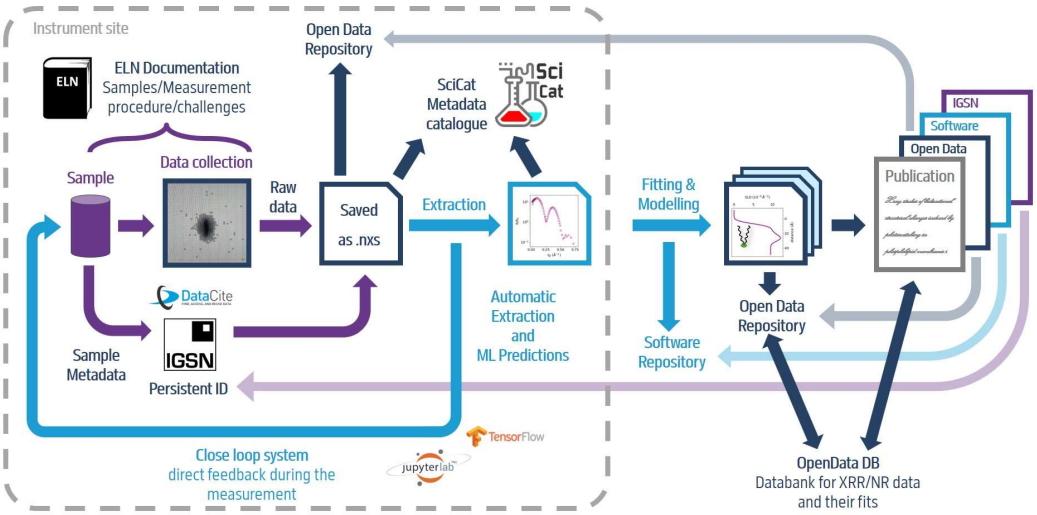

Figure 3. FAIR workflow for the data life cycle as implemented at P08 for the X-ray reflectivity use case (more information for a BioSAXS use case in: Hövelmann, S. C. *et al*. 2024).

In close collaboration with the beamline staff at PETRA III beamline P08 and the Experimental Control group at DESY, integration of electronic logbooks and metadata catalogues with automatic data ingestion are implemented. In addition, persistent sample identifiers (IGSN) are employed to identify samples and their associated datasets (more information for a BioSAXS use case in: Hövelmann, S. C. *et al*. 2024).

Aiming for a fully traceable and FAIR data workflow, analysis software developed within this use case is publicly available such as the ML-data analysis packages mlreflect (Greco, A. *et al.* 2019) and reflecttorch (Munteanu, V. *et al.* 2024). Implementing the ML-base analysis at the beamline allows fast online data processing and enables closed loop in situ measurements with a direct feedback to the beamline and experiment control software based on the analysis results (Pithan, L. *et al.* 2023). Further, a reflectivity reference database is under construction (at public-data.desy.de) in which publish data and their results are findable and reusable, for example for validation of analysis routines and ML model training .

3.2 X-ray photon correlation spectroscopy (XPCS)

The current use case is focused on X-ray Photon Correlation Spectroscopy (XPCS) – the experimental technique, allowing to access protein and molecular dynamics on the length scales from Angstroms to micrometres with a temporal resolution starting in sub-microsecond time domain (Figure 4).

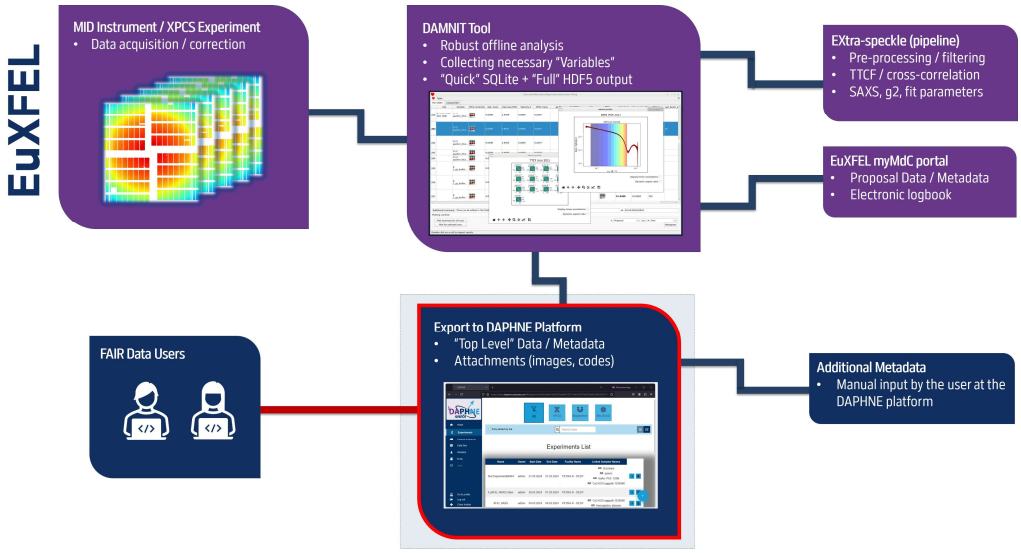

Figure 4. FAIR data and metadata collection strategy for the XPCS use case.

One of the key features of XPCS experiments at EuXFEL is the large amount of the collected data, which can reach the volumes up to several Petabytes during a single beamtime and its processing may become a tough challenge. Along with that the users need fast and reliable feedback from the data processing pipelines, providing, on the one hand, the possibility to flexibly tune up the experimental parameters during carrying out the XPCS experiment, and, on the other hand, exploit the facility's infrastructure for post-processing and refining of the data afterwards. In order to achieve this, the robust user-oriented open-source software (EXtra-speckle) is being developed and successfully integrated with the generic DAMNIT tool of EuXFEL (https://damnit.readthedocs.io/en/latest/), providing one to perform the near-online analysis of

the experimental data and its convenient visualization. In order to make the data and metadata FAIR, efforts are made to establish secure bridging between the DAMNIT tool and the external DAPHNE4NFDI database (XPCS platform), which can be used as a portal for public users. Similar activities are performed with the experimental (meta)data collected during the XPCS experiments at P10 beamline of PETRA III at DESY.

3.3 X-ray absorption spectroscopy (XAS)

In a third use case, the focus is on X-ray absorption spectroscopy (XAS), which is important for functional materials like catalysts or sensor and in actinide research. In this case, it is especially important to compare spectra and therefore a reference database is developed first for X-ray absorption near edge structure (XANES) spectra as well as for extended X-ray absorption fine structure (EXAFS) data (Figure 5).

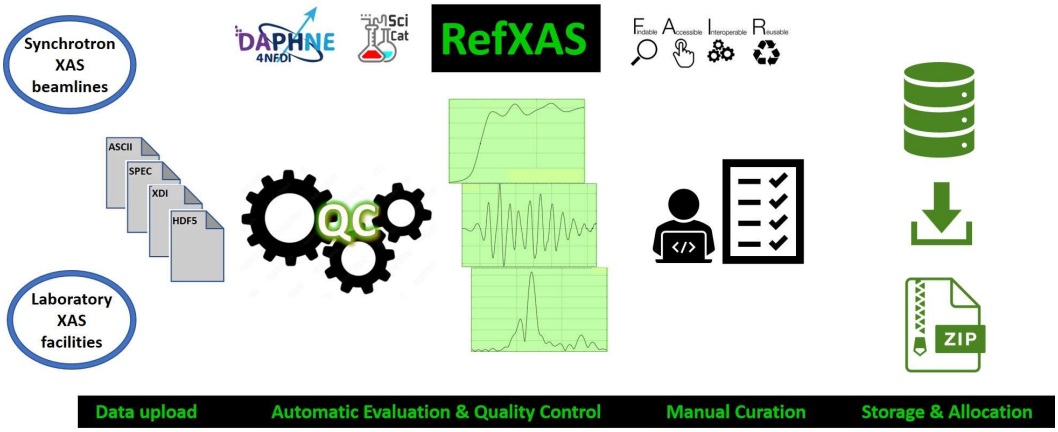

Figure 5. Workflow implementation in the XAS use case with RefXAS as dedicated repository of reference spectra (more information in: Gaur, A. *et al.* 2023; Paripsa, S. *et al.* 2024).

In order to achieve this, depositing of results, quality assurance and processing/analysis software are developed (Gaur, A. *et al.,* 2023; Paripsa, S. *et al.* 2024). Metadata fields and the quality criteria are important for re-use. This use case, once being established will be a prototype and will be extended to other reference databases, e.g. for X-ray emission spectra.

The integration of the NeXus data format is being explored to improve standardization and interoperability within the XAS database. This approach aligns with the FAIR principles by making the database more accessible and machine-readable, which supports advanced computational methods, such as machine learning. As part of DAPHNE4NFDI, this XAS use case demonstrates a cohesive research data management strategy at synchrotron radiation facilities, facilitating data sharing and analysis.

4. Further information and Acknowledgement

The slides of the talk given at SRI2024 conference "DAPHNE4NFDI - Improving Research data management at Synchrotron Facilities" can be accessed online (Amelung, L. *et al.* 2024).

This publication was written in the context of the work of the consortium DAPHNE4NFDI in association with the German National Research Data Infrastructure (NFDI) e.V. NFDI is financed by the Federal Republic of Germany and the 16 federal states and the consortium is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - project number 460248799. The authors would like to thank for the funding and support. Furthermore, thanks go to all institutions and actors who are committed to the association and its goals.

5. References

Amelung, L., Barty, A., Murphy, B. M., Grunwaldt, J.-D., Schneidewind, A., Busch, S., Gutt, C., Lohstroh, W., Schreiber, F. and Unruh, T. (2024). "DAPHNE4NFDI – Improving Research data management at Synchrotron Facilities." DOI: 10.5281/zenodo.13638030.

Barty, A., Gutt, C., Lohstroh, W., Murphy, B., Schneidewind, A., Grunwaldt, J.-D., Schreiber, F., Busch, S., Unruh, T. and Bussmann, M., Fangohr, H., Görzig, H., Houben, A., Kluge, T., Manke, I., Lützenkirchen-Hecht, D., Schneider, T., Weber, F., Bruno, G., Einsle, O., Felder, C., Herzig, E., Konrad, U., Markötter, H., Rossnagel, K., Sheppard, T., and Turchinovich, D. (2023). "DAPHNE4NFDI - Consortium Proposal." DOI: 10.5281/zenodo.8040605.

Gaur, A., Paripsa, S., Förste, F., Doronkin, D., Malzer, W., Schlesiger, C., Kanngießer, B., Lützenkirchen-Hecht, D., Welter, E. and Grunwaldt, J.-D. (2023). "Metadata Fields and Quality Criteria - XAS Reference Database under DAPHNE4NFDI." *Proceedings of the Conference on Research Data Infrastructure* **1**. DOI: 10.52825/cordi.v1i.258.

Greco, A., Starostin, V., Karapanagiotis, C., Hinderhofer, A., Gerlach, A., Pithan, L., Liehr, S., Schreiber, F. and Kowarik, S. (2019). "Fast fitting of reflectivity data of growing thin films using neural networks." *Journal of Applied Crystallography* 52(6): 1342-1347. DOI: 10.1107/S1600576719013311.

Hövelmann, S. C., Dieball, E., Kuhn, J., Dargasz, M., Giri, R. P., Reise, F., Paulus, M., Lindhorst, T. K. and Murphy, B. M. (2024). "Photoinduced bidirectional mesophase transition in vesicles containing azobenzene amphiphiles." *International Union of Crystallography Journal* **11**(Pt 4): 486-493. DOI: 10.1107/s2052252524004032.

Munteanu, V., Starostin, V., Greco, A., Pithan, L., Gerlach, A., Hinderhofer, A., Kowarik, S. and Schreiber, F. (2024). "Neural network analysis of neutron and X-ray reflectivity data incorporating prior knowledge." *Journal of Applied Crystallography* 57(2): 456-469. DOI: 10.1107/S1600576724002115.

Paripsa, S., Gaur, A., Forste, F., Doronkin, D. E., Malzer, W., Schlesiger, C., Kanngiesser, B., Welter, E., Grunwaldt, J.-D. and Lützenkirchen-Hecht, D. (2024). "RefXAS: an open access database of X-ray absorption spectra." *Journal of Synchrotron Radiation* 31(5): 1105-1117. DOI: 10.1107/S1600577524006751.

Pithan, L., Starostin, V., Marecek, D., Petersdorf, L., Volter, C., Munteanu, V., Jankowski, M., Konovalov, O., Gerlach, A., Hinderhofer, A., Murphy, B., Kowarik, S. and Schreiber, F. (2023). "Closing the loop: autonomous experiments enabled by machine-learning-based online data analysis in synchrotron beamline environments." *Journal of Synchrotron Radiation* 30(6): 1064-1075. DOI: 10.1107/S160057752300749X.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. and Mons, B. (2016). "The FAIR Guiding Principles for scientific data management and stewardship." *Scientific Data* 3(1): 160018. DOI: 10.1038/sdata.2016.18.