

Christian Gutt, Wiebke Lohstroh, Bridget Murphy

# DAPHNE4NFDI: Data infrastructures for photon and neutron science



The 2024 Nobel Prize in Chemistry honored breakthroughs in AI-based protein structure prediction and highlighted a lesson that reaches far beyond structural biology: transformative scientific advances increasingly rest on decades of shared experimental data. In this case, the Protein Data Bank (PDB) [1], a curated repository of experimentally determined structures, provided much of the structural “ground truth” used to train and benchmark modern machine-learning methods [2]. Built over decades from measurements at large-scale facilities, most prominently synchrotron X-ray crystallography, complemented by neutron methods and, increasingly, cryo-electron microscopy, the PDB demonstrates how sustained data curation can turn individual experiments into a collective resource with long-term scientific impact.

A comparison with everyday practice in physical chemistry highlights how exceptional this situation still is. X-ray absorption spectroscopy for example, is a widely used technique at photon sources, but despite progress many datasets recorded a decade or more ago are rarely reused to address new scientific questions. While the underlying data may still exist, essential information on instrument configurations, calibration procedures, background treatment, and analysis strategies is often no longer preserved in a form that enables reuse. The contrast with the PDB underscores that long-term data reuse is not an automatic consequence of experimental quality, but depends on sustained community standards, shared infrastructures, and continuous stewardship related to all steps of the data capture and analysis as is currently being developed in the RefXAS reference database [3].

The 2024 Nobel Prize in chemistry also illustrates a broader point relevant to physical chemistry and materials research: powerful data-driven methods emerge when high-quality experimental data from large-scale facilities are systematically

curated and made reusable. The ability to design materials, catalysts, and pharmaceuticals with targeted properties increasingly relies on the combination of advanced photon and neutron experiments with modern machine-learning approaches. In this context, next-generation synchrotron and neutron sources such as PETRA IV (DESY) and the European Spallation Source (ESS) represent more than incremental improvements in experimental capability. By delivering data at unprecedented precision and scale, they provide a foundation for data-intensive and AI-assisted approaches that are essential for accelerating materials discovery, supporting energy-transition technologies, and enabling innovation in industrial research, provided that experimental data can be systematically curated and reused.

While next-generation sources will further amplify these possibilities, many of the underlying challenges already exist today. In physical chemistry, materials science, and related disciplines, complex photon and neutron experiments are performed every day, but their results are far less systematically preserved, described, and reused across projects and institutions. This gap becomes increasingly critical as photon and neutron science enters an era of unprecedented data production. Modern synchrotron radiation sources, X-ray free-electron lasers (XFELs), and neutron facilities generate data volumes that exceed traditional workflows. With next-generation sources such as PETRA IV on the horizon, expected to produce data on the scale of hundreds of petabytes per year, the question of how experimental results are efficiently analysed, curated, connected, and reused becomes central to scientific progress.

Photon and neutron experiments often display high experimental complexity. Measurements often involve sophisticated sample environments, operando or time-resolved conditions, and multi-modal data acquisition. Reliable data depend sensitively on beam properties, calibration procedures, data-reduction steps, and analysis choices. Capturing this context in a structured and reusable way is therefore essential if data are to remain interpretable beyond the original beamtime.

Addressing these challenges requires more than additional storage or faster data transfer. What is needed is a systematic approach that allows experimental results to be found, understood, and reused beyond the original experiment, summarized in the FAIR principles [4]. In photon and neutron science, interoperability relies in particular on shared data standards and technique descriptions, such as the continued development of NeXus data formats and the PaNET ontology [5, 6]. Crucially, FAIR must extend beyond experimental data to include analysis

---

Prof. Dr. Christian Gutt (Main author)  
Department Physik, Universität Siegen  
Walter-Flex-Straße 3, 57072 Siegen  
gutt@physik.uni-siegen.de

PD Dr. Bridget Murphy  
Institute of Experimental and Applied Physics, Kiel University,  
Leibnizstraße 19, 24118 Kiel, Germany  
Ruprecht Haensel Laboratory, Deutsches Elektronen-Synchrotron DESY,  
Notkestraße 85, 22607, Hamburg, Germany

Dr. Wiebke Lohstroh  
Heinz Maier-Leibnitz Zentrum (MLZ), Technische Universität München,  
Lichtenbergstraße 1, 85748 Garching, Germany

DOI-Nr.: 10.26125/q620-gw71

software and workflows. End-to-end reproducibility depends on transparent, well-documented pipelines, including versioned code, recorded parameters, and persistent links between data and the software used to process them. Making data and software FAIR also strengthens the verification of results and trust in scientific outcomes [4, 7].

Within the framework of the National Research Data Infrastructure (NFDI), different consortia address the data-management needs of distinct scientific communities. DAPHNE4NFDI represents those researching at large-scale photon and neutron sources, including synchrotron radiation facilities, XFELs, and neutron scattering centres, and supports a broad spectrum of experimental techniques [8, 9, 10]. It supports the research-data-management needs of a large and diverse scientific community. In Germany alone, the photon and neutron user landscape comprises approximately 7 000 registered facility users, who carry out thousands of experiments each year and collectively produce more than 2 000 scientific publications annually. Beyond academia, industrial and applied-research communities increasingly rely on photon and neutron techniques, ranging from catalysis and pharmaceuticals to engineering, energy technologies, and fusion-related research.

The strength of the German photon and neutron community lies in the close cooperation between large-scale research facilities, universities, and democratically organised user representations. Facilities provide the experimental infrastructure and methodological expertise required for state-of-the-art measurements, while universities play a central role in providing domain knowledge as well as education and training, ensuring that the next generation of scientists encounters modern data practices early in their careers. Elected community bodies such as the Komitee Forschung mit Synchrotronstrahlung (KFS) and the Komitee Forschung mit Neutronen (KFN) articulate scientific needs and priorities from the user perspective.

DAPHNE4NFDI is a joint effort of users and facilities, built on a science-driven, bottom-up approach. Developments begin with concrete scientific use cases and real experimental needs, and solutions are shaped through close interaction between users, beamline scientists, method developers, and university-based researchers. Building on a highly successful first funding period, the current proposed phase of DAPHNE4NFDI focuses on moving from demonstrator solutions to routine, sustainable operation, accompanied by training activities, documentation, and community exchange formats that support broad uptake [8, 9].

The scientific use cases driving these developments span a wide range of photon and neutron techniques central to physical chemistry and materials research. They include operando and time-resolved studies of catalytic reactions and functional materials, coherent scattering and X-ray photon correlation spectroscopy for soft and biological matter, X-ray absorption spectroscopy, as well as neutron and X-ray diffraction, reflectometry, spectroscopy and imaging. Many of these experiments are multimodal and depend on near-online feedback during beamtime, placing high demands on structured metadata and reproducible analysis workflows.

To support these use cases, DAPHNE4NFDI provides a growing portfolio of user-facing services designed to integrate into experimental workflows at large-scale facilities leading to FAIR publications [11]. These include electronic laboratory notebook concepts tailored to photon and neutron experiments, automated metadata capture at beamlines, facility-wide data catalogues based on SciCat and ICAT [12, 13], and curated reference databases such as RefXAS [3]. Persistent identifiers for samples and datasets strengthen provenance across the experimental chain, while FAIR-oriented software packaging and workflow descriptions enable reproducible analysis. Near-online inspection and metadata extraction tools such as DAMNIT demonstrate how analysis workflows can be coupled directly to data acquisition handling up to PBs of data per experiment [14].

Beyond the national context, DAPHNE4NFDI is embedded in the European research-data ecosystem. Its services are integrated with the German National Node of the European Open Science Cloud (EOSC) and the Photon and Neutron topical node, linking facility-based research in Germany to a broader European open-science infrastructure [15].

Many photon and neutron experiments already rely on automation, including online data reduction and adaptive scanning strategies. As curated datasets and reference collections grow, they increasingly enable data-driven and machine-learning-based approaches to experiment control and analysis. These developments point toward more autonomous and adaptive experiments. DAPHNE4NFDI provides the data and software solutions that make such developments feasible and reproducible [16].

Photon and neutron science is entering an era of unprecedented data abundance. The central challenge is no longer data generation, but sustained stewardship: ensuring that valuable experimental results remain accessible, interpretable, and reusable over decades. We need highly curated resources such as the Protein Data Bank, established across the photon and neutron techniques. By translating bottom-up scientific needs into interoperable and routinely available services, DAPHNE4NFDI helps transform individual experiments into a collective scientific memory. Simultaneously, the community is actively discussing how such research-data-management infrastructures can be maintained and evolved in the long term, beyond individual project phases, to ensure continuity, reliability, and lasting scientific value.

## References

- [1] Berman, H. M. *et al.* *The Protein Data Bank. Nucleic Acids Research* **28**, 235–242 (2000).
- [2] Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
- [3] Paripisa, R. *et al.* *RefXAS: a reference database for X-ray absorption spectroscopy. Journal of Synchrotron Radiation* **31**, 1105 (2024).
- [4] Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**, 160018 (2016).

- [5] Könnecke, M. et al. The NeXus data format. *Journal of Applied Crystallography* **48**, 301–305 (2015).
- [6] Collins, S.; Millar, P.; Servan, S. *PaNET Ontology Sustainability Sheet*; Zenodo, 2023. <https://doi.org/10.5281/ZENODO.7684470>.
- [7] Helliwell, J.R. (2024). *The Scientific Truth, the Whole Truth and Nothing but the Truth* (1st ed.). CRC Press. <https://doi.org/10.1201/9781003405399>
- [8] Murphy, B. M. et al. *FAIR data – the photon and neutron communities move together towards open science 2025*, IUCR **12**, 8 (2025).
- [9] Barty, A. et al. *DAPHNE4NFDI - Consortium Proposal*, Zenodo, <https://doi.org/10.5281/zenodo.8040606> (2023).
- [10] Dolcet, P. et al. *DAPHNE4NFDI Annual Meeting 2023*. *Synchrotron Radiation News* **36**(6), 55–57 (2023).
- [11] Hovelmann, S. C., Dieball, E., Kuhn, J., Dargasz, M., Giri, R. P., Reise, F., Paulus, M., Lindhorst, T. K. & Murphy, B. M. (2024). *IUCr*, **11**, 486-493.
- [12] <https://www.scicatproject.org>
- [13] Flannery, D. et al. *ICAT: Integrating data infrastructure for facilities-based science*. *J. Phys.: Conf. Ser.* **513**, 042038 (2014).
- [14] Leonau, A. et al. A pipeline for megahertz X-ray photon correlation spectroscopy at the European XFEL (DAMNIT), *arXiv:2506.08668* (2025).
- [15] European Open Science Cloud (EOSC). Photon and Neutron Topical Node.
- [16] Pithan, L. et al. *Closing the loop: autonomous experiments at large-scale facilities*. *Journal of Synchrotron Radiation* **30**, 1064 (2023).

### Prof. Dr. Christian Gutt

Christian Gutt studied physics at Kiel University and completed his PhD in 2001. He specializes in coherent X-ray scattering using synchrotron and XFEL sources to study dynamic phenomena in condensed matter. After postdoctoral research at TU Dortmund and the University of California, San Diego, he worked as a senior scientist at DESY in Hamburg. Since 2013, he has been Professor of Condensed Matter Physics at the University of Siegen. He currently chairs the Committee for Research with Synchrotron Radiation (KFS) and is co-speaker of the DAPHNE4NFDI consortium.



### PD Dr. Bridget Murphy

Bridget Murphy studied applied physics at Dublin City University and completed her PhD in 2004 at Kiel University on X-ray diffraction at solid surfaces. Her research focuses on synchrotron-based X-ray scattering of interfaces and functional materials. She held beamline positions at Daresbury synchrotron light source, and was visiting scientist at Brookhaven, European synchrotron radiation facility (ESRF) and the Institute Laue-Langevin. Since 2012, she has been Group Leader of the LISA Group at Kiel University and, since 2024, DAPHNE Group Leader for Photon Science at DESY. She is currently Vice President of the European synchrotron and free electron laser user organisation (EUSO), and an elected member of the NFDI Senate. She is the speaker of the DAPHNE4NFDI consortium.



### Dr. Wiebke Lohstroh

Wiebke Lohstroh studied physics at the University of Göttingen, where she completed her PhD on neutron reflectometry of magnetic multilayers. After postdoctoral research at the University of Oxford and Vrije Universiteit Amsterdam, she worked on hydrogen storage materials at the Karlsruhe Institute of Technology. Since 2011, she has been an instrument scientist at the time-of-flight spectrometer TOFTOF at the Heinz Maier-Leibnitz Zentrum (FRM II/MLZ) in Garching. Her research focuses on quasielastic neutron scattering of complex hydrides, hydrogen storage systems, and energy materials. Since 2019, Wiebke Lohstroh is head of the Instrument Group, responsible for instrumentation and data management at the neutron research user facility (FRM II/MLZ). She is co-speaker of the DAPHNE4NFDI consortium.

