DAPHNE4NFDI Consortium Progress report

National Research Data Infrastructure

B-1 Progress Report Part 1, for publication

on the basis of **DFG form nfdi140 – 04/24**

Instructions and Template

for Consortia Progress Reports

National Research Data Infrastructure (NFDI)

Project number: 460248799

Hamburg, September 30th 2024

Table of contents

1 General Information

1.1 Name of the consortium

NFDI 40/1 "DAPHNE4NFDI - DAten aus PHoton- und Neutronen Experimenten für NFDI"

1.2 Research domains or research methods addressed by the consortium

According to *DFG Classification of Scientific Disciplines, Research Areas, Review Boards and Subject Areas (2024-2028)*[1] the following research domains and methods are addressed by the consortium:

- 2.11 Basic research in biology and medicine
- 3.13 Physical Chemistry
- 3.16 Polymer Research
- 3.17 Theoretical Chemistry
- 3.21 Condensed Matter physics
- 3.22 Statistical Physics, Nonlinear Dynamics, Complex Systems, Soft and Fluid Matter, Biological Physics
- 3.23 Optics, Quantum Optics, Physics of Atoms, Molecules and Plasmas
- 3.44 Mineralogy, Petrology and Geochemistry
- 4.21 Process Engineering, Technical Chemistry
- 4.32 Materials Science

1.3 URL of the consortium website and repositories used for publishing output

The following URLs are used as consortium website and for publishing output of the consortium:

- DAPHNE4NFDI Consortium website: https://www.daphne4nfdi.de/english/index.php
- DAPHNE4NFDI Zenodo community: https://zenodo.org/communities/daphne4nfdi
- Public Data DESY database: https://public-data.desy.de/
- RefXAS Research database: http://xafsdb.ddns.net/

[1] Deutsche Forschungsgemeinschaft (DFG) (2024) ”DFG Classification of Scientific Disciplines, Research Areas, Review Boards and Subject Areas (2024-2028)” Retrieved 2024-09-15 from https://www.dfg.de/resource/blob/331950/85717c3edb9ea8bd453d5110849865d3/fachsystematik-2024-2028-en-data.pdf.

2 Summary

The main goal of **DA**ta from **PH**oton and **N**eutron **E**xperiments for NFDI (DAPHNE4NFDI) (Barty, A. *et al.* 2023 DOI: 10.5281/zenodo.8040605) is to improve the quality of science for the photon and neutron (PaN) user community and beyond, by making the rapidly increasing data quantities FAIR. The science in these communities is carried out at large-scale research facilities (LSFs) where following a competitive peer review process user groups travel to perform specialized experiments, often complemented with data from their home laboratory. This common point of access provides DAPHNE4NFDI a unique opportunity to inform, provide services, promote and distribute tools for research data management (RDM) and to implement the FAIR data principles in the community.

DAPHNE4NFDI brings the organisations representing both the users and the facilities together within Germany and across Europe and actively integrates the community into the consortium activities. The relevant national user organisations interacting with the German photon and neutron LSFs are KFS (photons) and KFN (neutrons). On the European level, users are represented by the ESUO (photons) and ENSA (neutrons), the facilities join forces in LEAPS and LENS. DAPHNE4NFDI (co-) spokespersons are active in all of the above. The approach of DAPHNE4NFDI is to bring the user perspective into focus, building on previous efforts to make photon and neutron data FAIR, in particular the EU funded projects PaNOSC and ExPaNDS and is a partner for OSCARS. Further we developed close international cooperation on data preparing a European community open data white paper.

Members of DAPHNE4NFDI communicate directly with the user communities through participations in and contributions to facility user meetings, conferences and workshops, e.g. DPG Spring Meetings. Members of DAPHNE4NFDI also actively participate in targeted workshops within academia, including interactions with ErUM-Data and CRCs. DAPHNE4NFDI contributes to the development of the NFDI by cooperating with other consortia, being active in all organs, including the senate and also contributes to the NFDI self-organization (Amelung, L. *et al.* DOI: 10.5281/zenodo.10101412). DAPHNE4NFDI became the second major user of Helmholtz Federated IT Services among the NFDI consortia in 2022 (HIFIS (2022) https://www.hifis.net) and our activities in adopting the Helmholtz ID serve as a use-case within the basic service IAM4NFDI. In the future, we will continue to contribute to a “one NFDI”.

DAPHNE4NFDI is also directly engaged with active users at the facilities through the 11 use cases which cover a large fraction of the typical techniques offered at PaN facilities. This allows for an agile development and constant readjustment of changing user needs and technical developments. To illustrate the power of the use cases we give two examples: In the X-ray absorption spectroscopy (XAS) use case the whole FAIR data pipeline from the sample, via the measurement and analysis to the publication has been established (Paripsa *et al*. 2024 DOI: 10.1107/S1600577524006751). The soft matter and liquids reflectivity and small angle X-ray scattering use cases focus on implementing a FAIR

workflow and machine learning (Pithan, L. *et al.* 2023 DOI: 10.1107/S160057752300749X) to capture all aspects of an experiment from proposal to publication including samples (Hövelmann, S. C. *et al.* 2024 DOI: 10.1107/s2052252524004032).

In our aim to capture metadata as close to their production as possible and to improve standards to store them for later use, specification of domain specific metadata schemata and vocabulary was carried out through extensive discussions within the community. A set of specifications were agreed-on and published in a white paper (Lohstroh, W. *et al.* 2024 DOI: 10.5281/zenodo.12169110). A test sample persistent digital identifier (PID) service was instigated in collaboration with IGSN. The needs of electronic laboratory notebooks at large scale facilities are complex. Evaluation showed they need to be capable of recording heterogenous user experiments, and allow multiple user strict access protocols. The findings are being captured in a white paper.

(Meta)data repositories and catalogues are invaluable for our community, and we are working to establish interconnected data sets and catalogues that adhere to FAIR principles across PaN sources, universities, and research institutions. This should include all data processing and analysis steps, as well as detailed sample descriptions, to elevate the transparency, quality, and reusability of published data. All facilities have or are working on a metadata catalogue for raw and processed data, and a number of instances of the DAPHNE4NFDI supported SciCat system are running in test user operation. Additionally, a Public Data DESY database (SciCat DESY and DAPHNE4NFDI (2023) https://public-data.desy.de) has been established, and it is currently being populated.

Analysing data is challenging and sharing software can speed up progress. Together with the facilities we develop infrastructure for data and software reuse making data analysis tools and software FAIR. A number of analysis software suits have been developed within DAPHNE4NFDI supporting (meta)data extraction and data analysis workflows (Meinerzhagen, Y. *et al*. 2024 DOI: 10.1107/S1600576724007635), including machine learning tools (Munteanu, V. *et al*. 2024 DOI: 10.1107/s1600576724002115).

Through dissemination and outreach we seek to improve the awareness of FAIR and to develop related skills in our community and beyond. We have a DAPHNE4NFDI homepage, are active on social media and organize (annual) meetings and events. We have and continue to reach out to students by introducing RDM topics into university curricula, holding lecture series aimed at students and are in the process of organising a summer school.

Challenges for DAPHNE4NFDI remain in long-term hosting, authentication, increased cybersecurity and meeting the demands of machine learning.

3 Composition of the consortium

3.1 Applicant institution

Applicant institution	Location	Duration
Deutsches Elektronen-Synchrotron (DESY) Notkestrasse 85 D-22607 Hamburg Germany	Hamburg, Germany	10/21 - current

3.2 Spokesperson

Spokesperson	Institution, location	Duration
Anton Barty DESY	DESY Hamburg, Germany	10/21 - 07/24
Bridget Murphy DESY	DESY Hamburg, Germany	08/24 - current

3.3 Co-applicant institutions

Co-applicant institutions	Location	Duration
Deutsches Elektronen-Synchrotron	Hamburg, Germany	10/21 - current
European Molecular Biology Laboratory	Hamburg, Germany	10/21 - current
European XFEL GmbH	Schenefeld, Germany	10/21 - current
Forschungszentrum Jülich GmbH	Jülich and Garching, Germany	10/21 - current
Friedrich-Alexander-Universität Erlangen-Nürnberg	Erlangen, Germany	10/21 - current
Helmholtz-Zentrum Berlin für Materialien und Energie	Berlin, Germany	10/21 - current
Helmholtz-Zentrum Hereon (former Helmholtz-Zentrum Geesthacht)	Geesthacht and Garching, Germany	10/21 - current
Helmholtz-Zentrum Dresden-Rossendorf	Dresden, Germany	10/21 - current
Karlsruhe Institute of Technology	Karlsruhe, Germany	10/21 - current
Ludwig Maximilian University Munich	Munich, Germany	10/21 - current
RWTH Aachen University	Aachen, Germany	10/21 - current
Technical University of Munich	Munich, Germany	10/21 - current
Technical University of Berlin	Berlin, Germany	10/21 - current
University of Göttingen	Göttingen, Germany	10/21 - current
Kiel University	Kiel, Germany	10/21 - current
University of Siegen	Siegen, Germany	10/21 - current
University of Tübingen	Tübingen, Germany	10/21 - current
University of Wuppertal	Wuppertal, Germany	10/21 - current

3.4 Co-spokespersons

Co-spokespersons ORCID ID	Institution, location	Task area(s)	Duration
Anton Barty n.a.	Deutsches Elektronen-Synchrotron	TA1, TA2, **TA3** and **TA6** (both as lead until 07/24)	10/21 - current
Bridget Murphy 0000-0002-1354-2381	Kiel University, Deutsches Elektronen-Synchrotron (since 08/24)	**TA1**, TA2, TA3, TA4, **TA6** (as lead since 08/24)	10/21 - current

Co-spokespersons ORCID ID	Institution, location	Task area(s)	Duration
Astrid Schneidewind 0000-0002-7239-9888	Forschungszentrum Jülich GmbH	TA1, TA3, **TA4, TA5**	10/21 - current
Jan-Dierk Grunwaldt 0000-0003-3606-0956	Karlsruhe Institute of Technology	TA1, TA2, **TA4**	10/21 - current
Wiebke Lohstroh 0000-0001-8404-2109	Technical University of Munich	**TA1**, TA2, TA3, TA4	10/21 - current
Christian Gutt 0000-0002-0051-8542	University of Siegen	TA2, TA4, **TA5**	10/21 - current
Sebastian Busch 0000-0002-9815-909X	Helmholtz-Zentrum Hereon (former Helmholtz-Zentrum Geesthacht)	**TA2**, TA3, TA5	10/21 - current
Tobias Unruh 0000-0002-8903-4850	Universität Erlangen-Nürnberg	**TA2**	10/21 - current
Frank Schreiber 0000-0003-3659-6718	University of Tübingen	**TA3**	10/21 - current
Fabio Dall'Antonia	European XFEL GmbH	**TA3** (since 08/24)	08/24 - current
Thomas Schneider 0000-0001-6955-7374	European Molecular Biology Laboratory	TA1, TA2	10/21 - current
Jörg Hammel 0000-0002-6744-6811	Helmholtz-Zentrum Hereon (former Helmholtz-Zentrum Geesthacht)	TA1, TA3	10/21 - current
Thomas Kluge 0000-0003-4861-5584	Helmholtz-Zentrum Dresden-Rossendorf	TA1, TA2	10/21 - current
Michael Bussmann 0000-0002-8258-3881	Helmholtz-Zentrum Dresden-Rossendorf	TA1, TA2, TA3	10/21 - current
Ingo Manke 0000-0001-9795-5345	Helmholtz-Zentrum Berlin für Materialien und Energie	TA1, TA4	10/21 - current
Heike Görzig 0000-0001-9121-8643	Helmholtz-Zentrum Berlin für Materialien und Energie	TA1, TA2	10/21 - current
Frank Weber 0000-0002-4256-1354	Karlsruhe Institute of Technology	TA1, TA3, TA4	10/21 - current
Sarah Köster 0000-0002-0009-1024	University of Göttingen	TA1, TA3	10/21 - current
Paola Coan 0000-0003-1399-2398	Ludwig Maximilian University Munich	TA1, TA3	10/21 - current
Birgit Kanngießer 0000-0001-6508-0150	Technische Universität Berlin	TA2, TA3	10/21 - current
Dirk Lützenkirchen-Hecht 0000-0002-7605-4350	University of Wuppertal	TA2, TA3	10/21 - current
Andreas Houben 0000-0002-4918-6251	RWTH Aachen University	TA3	10/21 - current
Luca Gelisio 0000-0001-7832-6201	European XFEL GmbH	TA3	10/21 - current

Task area leaders are highlighted in bold face.

3.5 Participating institutions

Participating institutions	Location	Duration
Bundesanstalt für Materialforschung und -prüfung (BAM)	Berlin	10/21 - current
Deutsche Physikalische Gesellschaft (DPG)	Bad Honnef	10/21 - current
Max-Born-Institut Berlin, TU Berlin (MBI)	Berlin	10/21 - current
Physikalisch-Technische Bundesanstalt (PTB)	Braunschweig	10/21 - current
European Synchrotron Radiation Facility (ESRF)	Grenoble, FR	10/21 - current
European Spallation Source (ESS)	Lund, SE	10/21 - current
Institute Laue-Langevin (ILL)	Grenoble, FR	10/21 - current