

Several partners – joint effort: RDM synergies in large scale research

Astrid Schneidewind¹[\[https://orcid.org/0000-0002-7239-9888\]](https://orcid.org/0000-0002-7239-9888), Kilian Schwarz²[\[https://orcid.org/0000-0002-0800-2743\]](https://orcid.org/0000-0002-0800-2743),
Thomas Schörner²[\[https://orcid.org/0000-0002-7213-0352\]](https://orcid.org/0000-0002-7213-0352), Bridget M. Murphy¹[\[https://orcid.org/0000-0002-1354-2381\]](https://orcid.org/0000-0002-1354-2381),
and Martin Erdmann⁴ [\[https://orcid.org/0000-0002-1653-1303\]](https://orcid.org/0000-0002-1653-1303)

¹ JCNS at MLZ, FZ Jülich, Germany

² Deutsches Elektronen-Synchrotron DESY, Germany

³ Institut für Experimentelle und Angewandte Physik, Kiel University, Kiel, Germany

⁴ Physikalisches Institut III A, RWTH Aachen, Aachen, Germany

Abstract. We report about the synergies of NFDI, ErUM-Data, and related partners in RDM for large scale research in universe and matter.

Keywords: RDM, infrastructures, synergies

Research at large-scale research infrastructures (RI), combined with external use (user services), does not only require adequate funding, but also dedicated services – including data collection and data management. Scientists in the fields of research with ions, neutrons and photons, of high-energy particle and astro-particle physics, of the physics of hadrons and nuclei, of astronomy and accelerator physics – together called the “ErUM” communities (German for “Research on Universe and Matter”) – therefore join their efforts to meet the challenges of what is commonly referred to as the digital transformation.

These challenges arise on one hand from increasing data rates and volumes – data that need to be transported (e.g. radio astronomy – SKA project), or rapidly analyzed in situ / operando (photon and neutron science) or distributed within a collaboration for independent analysis (e.g. high-energy physics), and the storage of raw data needs to be curated. On the other hand, the dedicated infrastructures limit the storage of raw data. There are in addition aspects of open and FAIR data – often already respected within concepts of international collaboration – and of user-friendliness considering diverse needs, as well as different cultures and traditions across DIG-UM.

Within NFDI two consortia are based with in the “ErUM” communities : PUNCH4NFDI¹ is the consortium of particle, astro-, astroparticle, hadron and nuclear physics, representing scientists from universities, the Max Planck society, the Leibniz Association, and the Helmholtz Association. The goal is to setup a federated and "FAIR" science data platform, offering the infrastructures and interfaces necessary for access to and use of data and computing resources of the involved communities and beyond. DAPHNE4NFDI² is a photons and neutrons (PaN) user-driven consortium which brings scientists from universities and research institutes together with RIs, to develop digital based research data management with the aim of embed-

¹ <https://www.punch4nfdi.de/>

² <https://www.daphne4nfdi.de/english/index.php>

ding FAIR data practices in the community by developing and providing services for data capture, storage, curation and analysis. It serves a wide range of scientific disciplines across physics, chemistry, biology, health and environment sciences partially represented in other NFDI consortia. Although there is a diversity of scientific and technical topics synergy between PUNCH4NFDI and DAPHNE4NFDI include the tradition of long-standing international collaborations and the dependencies and interactions of users and facilities in data collection and delivery and the need to develop artificial intelligence methods.

Base4NFDI as an overarching initiative of all NFDI consortia for the development of basic services supporting the joint efforts of PUNCH and DAPHNE and strengthens the cooperation by coherent work. There is a common interest of both consortia in user-friendliness and compatibility to internationally established standards of basic services to be agreed within the NFDI landscape.

An independent request for a closer data-related cooperation within the ErUM communities is expressed by BMBF, which resulted in the ErUM-Data³ action plan. This referred to the outcome of a BMBF workshop, as a white paper from the ErUM communities entitled "Opportunities of Digital Transformation in Fundamental Research on Universe and Matter"⁴. In a self-organized manner DIG-UM works on five identified topics: Federated infrastructures, User interfaces, Research Data management, Big Data Analytics and Knowledge distribution. Correspondingly five working groups were established in this self-organized manner, to enable the joint scientific communities (DIG-UM) working together⁵. Similar conclusions and working structures were established with the first NFDI consortia. There is, by definition, topical overlap between NFDI and ErUM-Data; it must however be noted that the ErUM-Data comprises a larger community than DAPHNE4NFDI and PUNCH4NFDI.

Still more players exist in the ErUM data landscape with representatives being continuously invited to common discussions and activities:

A number of RIs, experiments and computing infrastructures are operated by the Helmholtz Association (HGF) that in turn, in the form of its centres, contributes to e.g. RDA, NFDI and EOSC. In addition to the established infrastructures, HGF develops its own strategy for exploiting digital techniques to improve its scientific work: the Helmholtz Incubator Framework for Information and Data Science. It contains five platforms: Metadata, AI, Imaging, Federated IT Services, and the Helmholtz information & data Science Academy. Moreover and important for ErUM, within the HGF's research field "Matter", the topic "Data Management and Analysis" (DMA) brings together scientists from facilities either run by Helmholtz centres or operated with Helmholtz involvement with topical experts in computational and data science along with numerous users from diverse communities. The Helmholtz project oriented funding scheme (POF) allows for long-term and strategic support of DMA and its goals.

The combined efforts of the partners described above will be reinforced by data competence centres (DKZ⁶), initialized by BMBF, these will act as sites for data-related education, research and networking, in close cooperation with NFDI. They shall support the cultural change in using data and the resulting data-based innovation.

Given such a manifold of partners with similar challenges and at least partly overlapping actors, it appears quite natural that work on synergies and efficient common structures was started quite early. Educational activities and information channels are already at least partly

³ https://www.bmbf.de/SharedDocs/Publikationen/de/bmbf/7/31640_Aktionsplan_ErUM-Data.pdf?__blob=publicationFile&v=7

⁴ http://www.physik.uni-siegen.de/x-ray/publications/erumdata_dina4_30.04.2019_druck.pdf

⁵ <https://erumdatahub.de/en/dig-um/>

⁶ <https://www.bmbf.de/bmbf/shareddocs/bekanntmachungen/de/2022/06/2022-06-21-Bekanntmachung-Datenkompetenzzentren.html>

shared between the various players. Obviously, the ErUM schools (e.g. on Deep Learning – basic as well as Train-the-Trainer) and the expert workshops (e.g. on analysis tools, sustainability in digital transformation) multiplex the network and interaction within the ErUM communities, and create joint knowledge. A special series of workshops on synergies is being established as a forum for information exchange and networking. Within this format, also Helmholtz activities can interact with the NFDI partners: Helmholtz-DMA views both the NFDI and ErUM-Data as natural partners and strives to support common solutions and complementary contributions to the data ecosystem. Beginning collaboration of DAPHNE4NFDI and PUNCH4NFDI with HMC guarantees not only increasing awareness on the importance of metadata, but concurrent metadata standards and agreed formats. In addition, topical connections to other NFDI consortia occur naturally.

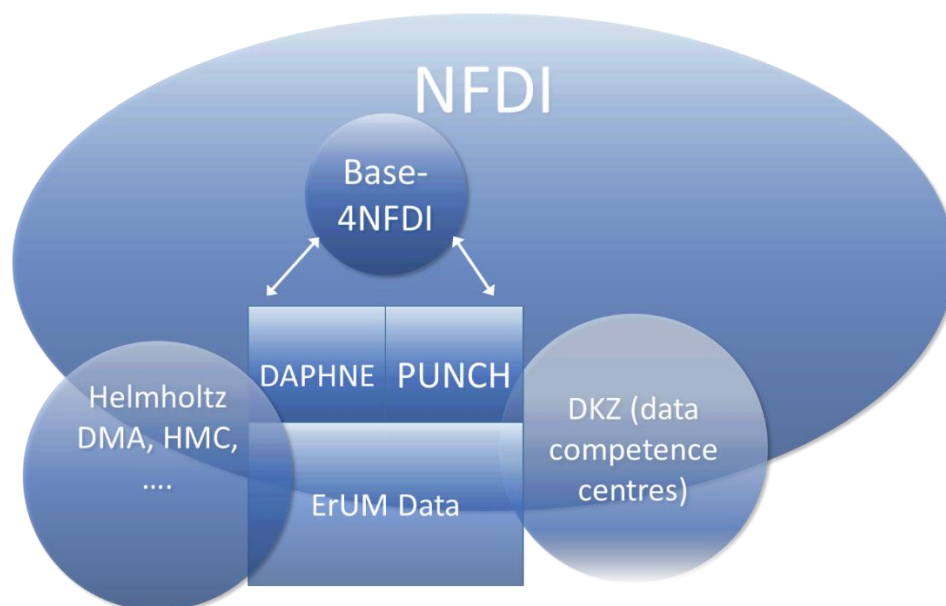


Fig.1: A view point of the NFDI initiatives with their connections to relevant efforts for mastering digital transformation in fundamental and applied research.

The exchange on detailed technical solutions, tools, workflows and strategies is another important aspect of the cooperation between ErUM-Data, NFDI, Helmholtz and others. – The effectiveness of such exchange can e.g. be seen in products like SciCat⁷ or Rucio⁸.

SciCat as a metadata catalogue was created especially for the needs of large and diverse user communities in using data from RIs – the challenge being e.g. to search, find and cite these data even when you are not a member of the original experimental team. University groups, in close contact with the main SciCat developers, now created a reduced SciCat version that allows the handling of data collected in small labs in a way similar to the data from large RIs, in close contact to the main developers.

Rucio provides services and associated libraries for allowing scientific collaborations to manage large volumes of data spread across numerous facilities at multiple institutions and organisations. It was originally developed to meet the requirements of the high-energy physics experiment ATLAS, and now is continuously extended to support the LHC experiments and other diverse scientific communities. DAPHNE4NFDI will test the usability of Rucio for the experiments in photon and neutron science.

Furthermore, joint working groups go for on PIDs, sustainable research software, education and metadata standards.

⁷ <https://scicatproject.github.io/>

⁸ <https://rucio.cern.ch/>

Author contributions

AS wrote the original draft. AS, KS, TSS, BM and ME reviewed and edited the article.

Competing interests

The authors declare that they have no competing interests.

Acknowledgement

We kindly acknowledge the in-kind support of FZJ, DESY, Kiel University and RWTH for the presented work. DAPHNE4NFDI is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 460248799. PUNCH4NFDI is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 460248186. ErUM-Data Hub is funded by BMBF (German Ministry for Education and Research), Förderkennzeichen 05D21PA1.