

White Paper On Metadata Catalogue Systems

Bishoy Hakim¹, Sebastian Busch², Abhijeet Gaur³, Rolf Krah⁴, Johannes Dallmann¹, Christian Felder⁵, Fabio Dall'Antonia⁶, Regina Hinzmann⁷, Sebastian Paripsa⁸, Björn Pedersen⁹, Nicolas Hayen¹⁰, Tobias Unruh¹.

¹Institute of Condensed Matter Physics, Friedrich-Alexander University Erlangen-Nuremberg (FAU), ²German Engineering Materials Science Centre (GEMS) at MLZ, Helmholtz-Zentrum hereon GmbH, ³Institute of Technical Chemistry and Polymer Chemistry, Karlsruhe Institute of Technology (KIT), ⁴Helmholtz-Zentrum Berlin für Materialien und Energie, ⁵Jülich Centre for Neutron Science (JCNS) at MLZ, Helmholtz-Zentrum hereon GmbH, ⁶European XFEL GmbH, ⁷Deutsches Elektronen-Synchrotron (DESY), ⁸Condensed Matter and Experimental Solid State Physics, University of Wuppertal, ⁹Technical University of Munich at Heinz Maier-Leibnitz Zentrum (MLZ), ¹⁰Institute of Experimental and Applied Physics, ¹⁰Kiel university.

DAPHNE4NFDI, part of Germany's National Research Data Infrastructure (NFDI), focuses on implementing FAIR principles for research data from Photon and Neutron (PaN) sources at large-scale facilities, universities, and research institutions. This includes the adoption of SampleDB as a flexible metadata solution, with comparisons to alternatives like ICAT and SciCat. Metadata catalogues are categorized into raw data repositories, institutional databases, and public-access systems, each with distinct roles. Key considerations include the integration of standards like NeXus, user-friendly searchability, and metadata enrichment. Case studies, including RefXAS, illustrate practical implementations and quality assessments. The paper also discusses Authentication and Authorization Infrastructure (AAI) strategies for secure, role-based data access, offering recommendations to enhance collaboration, data quality, and a unified, FAIR-compliant metadata ecosystem for the PaN research community.

The objective of this white paper Catalogue is to pinpoint a catalogue solution suitable for DAPHNE4NFDI partners. It aims to identify catalogue use-cases and the expected nature of searches along with outlining the metadata schema. The contents of this document provide a preliminary yet concise specification of (non-)requirements, a review of the current state and local solution landscape, incorporating lessons learned and a discussion leading to preliminary conclusions and decisions.