Data & Metadata in the ROCK-IT Project

DAPHNE4NFDI Annual Meeting 2025 | 24-26 March 2025 | Helmholtz-Zentrum Berlin

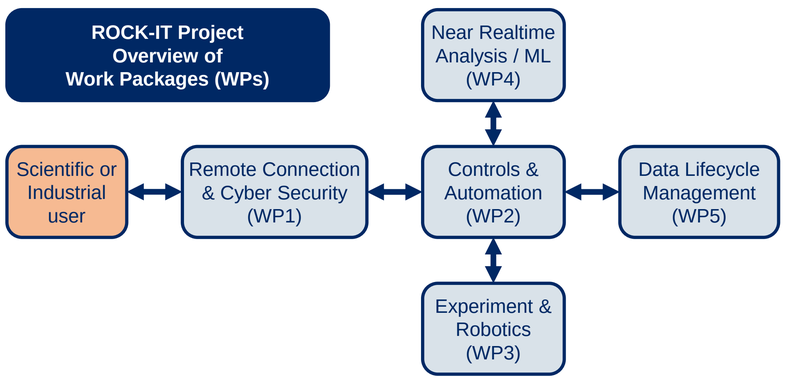
ABSTRACT: For a ROCK-IT thermal catalysis experiment, the (meta)data lifecycle in the facility begins with the initial proposal and extends through the entire investigation, culminating in the archiving and cataloguing of the (meta)data. At all stages, coherence between data and metadata is vital. We are using a combination of established systems and new initiatives to achieve a comprehensive framework aligned with FAIR principles.

K. Rial[2], J.-D. Grunwaldt[4], A. Rothkirch[1], H. Görzig[2], U. Konrad[3], A. Aggarwal[1], G. Guenther[2], R. Krahl[2], H. Perez Ponce[2], A. Velazquez Sanchez[2], T. Gruber[3], V. Biniyaminov[4], E. Bründermann[4], A. Gaur[4], J. Gethmann[4]

[1] Deutsches Elektronen-Synchrotron DESY, Notkestr 85, 22607 Hamburg
[2] Helmholtz-Zentrum Berlin für Materialien und Energie, Hahn-Meitner-Platz 1, 14109 Berlin
[3] Helmholtz-Zentrum Dresden Rossendorf, Bautzner Landstr 400, 01328 Dresden
[4] Karlsruhe Institute for Technology, Kaiserstr 12, 76131 Karlsruhe

Introduction

ROCK-IT is a collaborative project between four Helmholtz Centres (DESY, HZB, HZDR, KIT) that aims to standardise and automate the process of performing thermal catalysis experiments to allow remote operation regulated by AI/machine learning. The project is split into work packages (WPs). WP5 is responsible for data management.



The four centres already collect data and metadata from experiments, using their own unique setups. ROCK-IT is not proposing to replace these, rather the challenge is to continue using the existing systems whilst standardising the metadata requirements and the data formats used. In doing so, it is important to also conform to FAIR data principles.

These various requirements necessitate the creation and/or introduction of new elements to cover specific areas that are not currently robust. This is largely centre-specific, as it complements the existing infrastructure, yet must align to a universal framework across the project.

Here we outline how data management is being handled across ROCK-IT, and consider the new tools being developed and implemented by WP5 members and their colleagues across the four centres.

Metadata requirements

Metadata is vital to give data context. For ROCK-IT, the necessary metadata can be categorised as:

Anything needed…	Groups to specify needs
… to run the experiment	WP2, WP3, WP4
… for analysis	User, Beamline scientist
… for searchability	User, Community
… for accountability	WP1, Registration/user portal
… for reproducibility	User, WP2, WP3, WP4

This highlights the need for involvement from users and the wider catalysis community in identifying the metadata that should be collected.

This is complicated by the abundance of terms in use. For ROCK-IT, we are building a cross-reference of terms used by resources such as DAPHNE4NFDI, DataCite and the NFDI4Cat Voc4Cat ontology, as well as those used internally by bluesky, NeXus NXxas and the control systems.

Common formats

Within ROCK-IT, data will be stored in HDF5 files using the NeXus format. This allows the structure of the data output to be standardised, regardless of which formats the various instruments favour. To achieve this, HZB are using a conversion tool developed in-house, NeXusCreator. At DESY Nexus Writer from ESRF performs this function, as it integrates with their control software.

Similarly, it is helpful if incoming data is standardised. LinkML provides a way to convert data from a wide variety of ELNs (electronic lab notebooks) into a common format that can be used by the ROCK-IT centres. In particular, this will enable the collection of sample data and the import of data into the lab or beamline ELNs in use at the centres: SciLog (DESY), ICAT+ and Nomad (HZB), MediaWiki (HZDR), and LabIMotion (KIT).

Implementing new systems

Where necessary, ROCK-IT is implementing new tools. In some cases, this standardises existing data or procedures, for example introducing ophyd as an intermediary instrument control software layer.

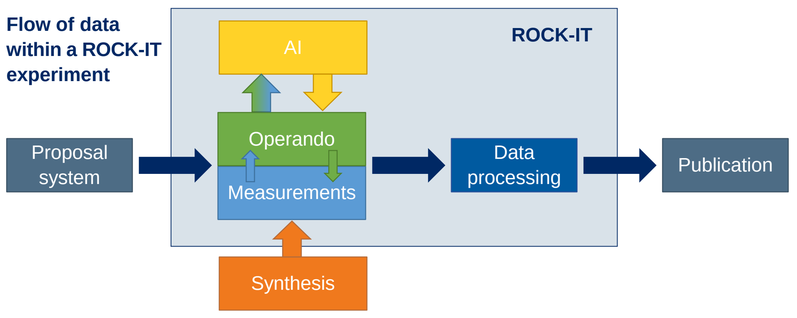
In other cases, new tools fill a hole in the established systems. The centres have not historically stored details about samples. In order for data to be useful beyond the scope of the original experiments, this information is vital. To this end, HZB (alongside HMC) is developing SEPIA, a new database to store sample data and mint IGSN identifiers that will be linked directly to the corresponding data. HZDR is beginning to develop a similar system that integrates with their infrastructure, and the two centres are sharing knowledge to both reduce repetition of work and provide the same functionality.

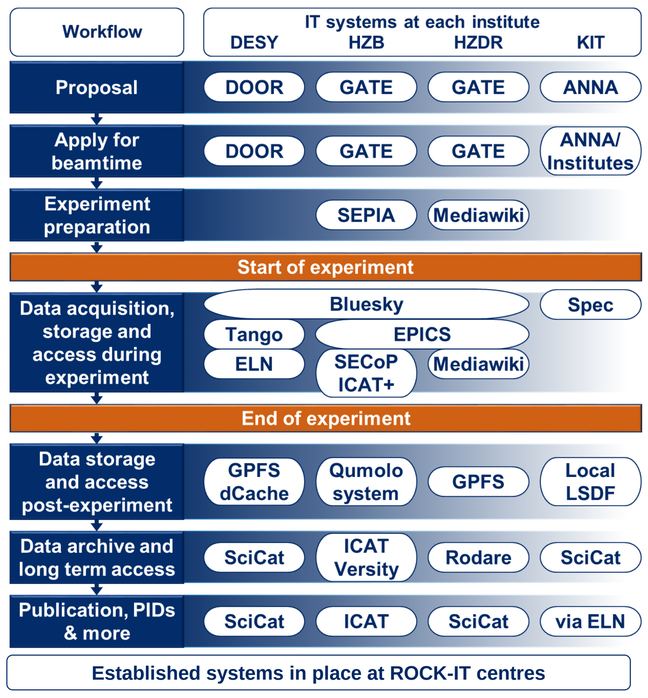
Navigating the infrastructure

With so many systems to keep track of across so many centres, ROCK-IT utilises a Data Management Architecture & Plan (DMAP), a living document that clearly lays out each required function and which systems fulfil that function. The DMAP is updated as plans change or new ideas prove useful, and serves as a guide to the entire ROCK-IT (meta)data infrastructure.

Working with existing systems

The four ROCK-IT centres have all been operational for many years, and so have a variety of systems already in place. ROCK-IT uses the existing infrastructure wherever possible. Many systems cannot be changed; for example the archiving and cataloguing of data is achieved using SciCat at some centres and ICAT at others, and this will not change in the foreseeable future.





To bring all these systems together, we first determined the flow of data through the experiments. We then identified the systems in place at each centre and investigated how similar or different they are. We agreed on common formats that all four centres could use within their existing setup.

We also identified gaps in the current setups, either universally or for individual centres. Some processes could not be automated, some information was not stored by the centre (which prevented the data being 'FAIR at the point it leaves the facility'). In particular, top-level systems were introduced that could interface with the established setup at each centre and then provide a uniform experience across the project. These have allowed us to standardise the data collection for ROCK-IT without replacing functional systems.

SUMMARY

- ROCK-IT requires both global and centre-specific solutions to provide a uniform experience despite centres using different underlying frameworks.
- Existing systems can be aligned to work similarly, enabling a uniform approach at higher levels.
- New tools are required to complement the existing systems, either to help them conform to global requirements or to supply additional functionality.
- Metadata requirements can only be established with input from a wide array of stakeholders. Crucially, we are working with catalysis scientists to identify metadata needs
- Metadata terms are not currently unified, and although community projects are making progress in this area, we need to identify equivalent terms relevant to ROCK-IT.
- Documentation is vital to navigate the complexities of the data infrastructures, but must be flexible to accommodate new ideas.

To Learn More, Contact:
Katherine Rial
+49 (30) 8062 – 12540
katherine.rial@helmholtz-berlin.de
ORCiD: 0009-0007-7266-485X

DAPHNE4NFDI Annual Meeting 2025 Helmholtz-Zentrum Berlin 24-26 March 2025

We gratefully acknowledge funding from the Helmholtz Association HGF for the ROCK-IT project.

rock-it-project.de