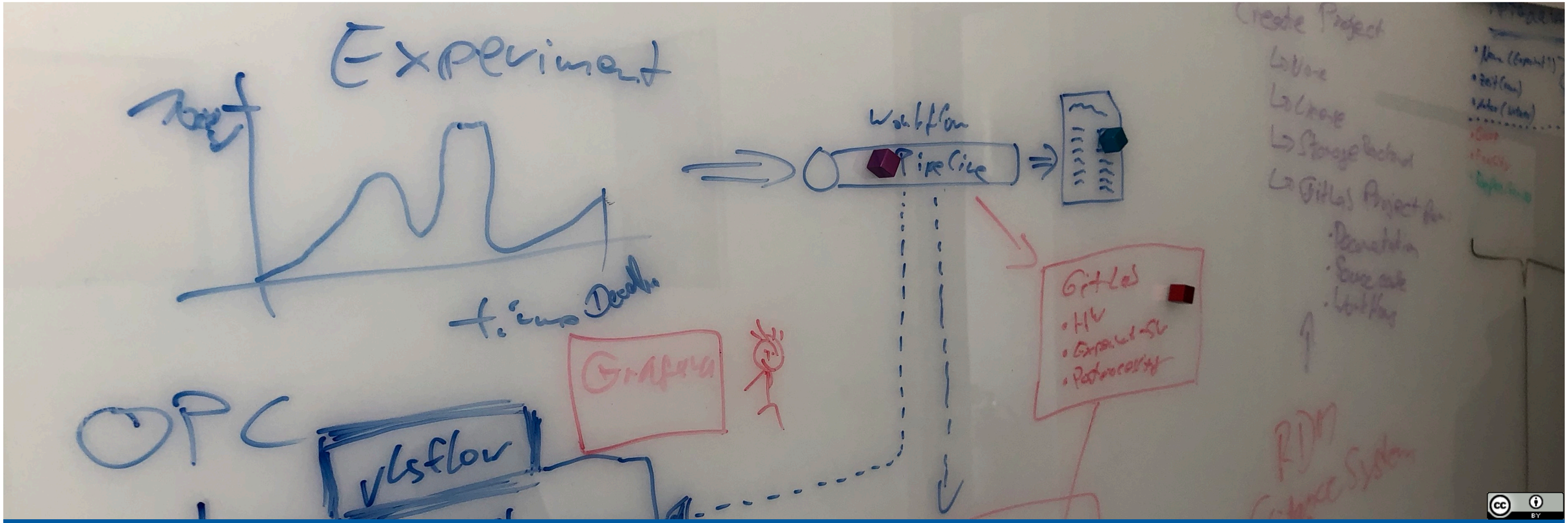
Scientific
Data Management

HELMHOLTZ ZENTRUM
DRESDEN ROSSENDORF



Strategies for the Integration of Metadata Catalogues, Data Publications and Archives

FWK Seminar, March, 2025

Oliver Knodel // contact: o.knodel@hzdr.de

DRESDEN concept

Development of the HZDR

Open Data Ecosystem

Scientific Computing Department at HZDR — 2017

Establishment of the Scientific Computing division within the IT as a bridge to the science

Data Policy — 2018

Legal framework for the institutional data management at HZDR.

Data Repository RODARE — 2018

Provision of the Rosendorf Data Repository for data and software publications.

RODARE
ROSSENDORF DATA REPOSITORY

Scientific Data Management (and HPC) — since 2019

Analysis of data/service landscape at HZDR and development of an uniform data lifecycle.

Scientific Data Management

HZDR Metadata Catalogue — 2024

Metadata catalogue for additional experiment-specific metadata as extension of RODARE.

ROMECA
Rossendorf Metadata Catalogue

HZDR RDM Strategy — 2025+

Identification of the need to develop a data management strategy together with our scientists.

CC BY

DRESDEN concept

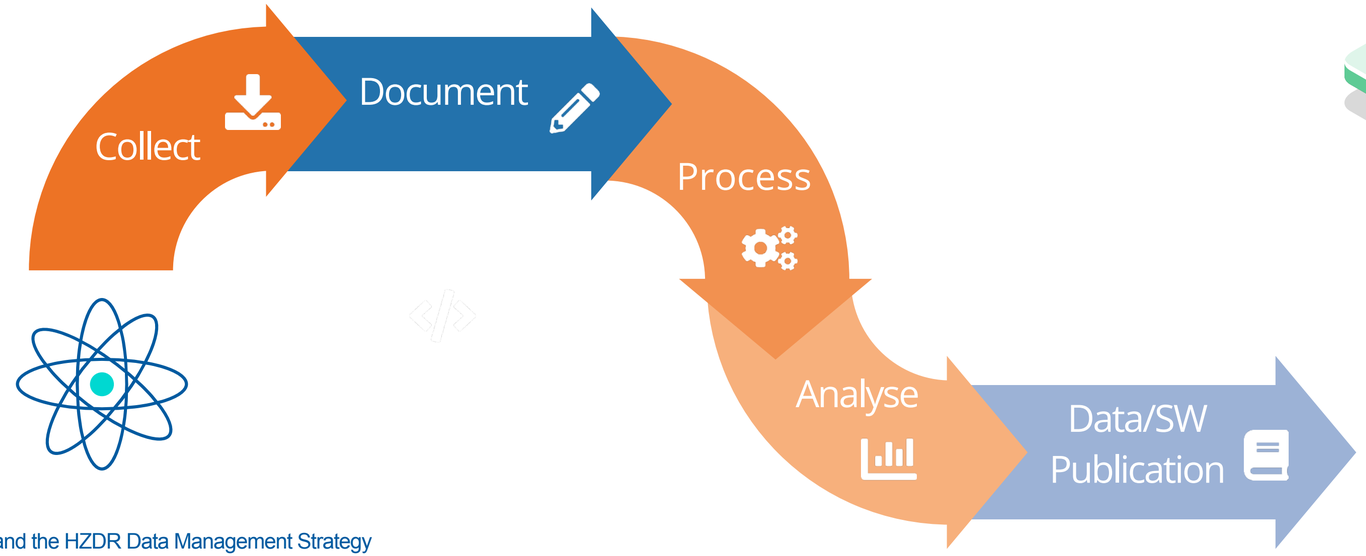
HZDR

The Mission of the Scientific Data Management Group

— Supporting our scientists and research experiment with tools:
  - electronic lab books,
  - interactive analysis,
  - publication repositories for data and software,
  - scientific workflow management,
  - PID (handle) generation and management.

— Providing support in connecting experiments, detectors and diagnostics to our infrastructure.

— Establishment of analysis workflows and HPC applications.



CC BY
DRESDEN concept
HZDR

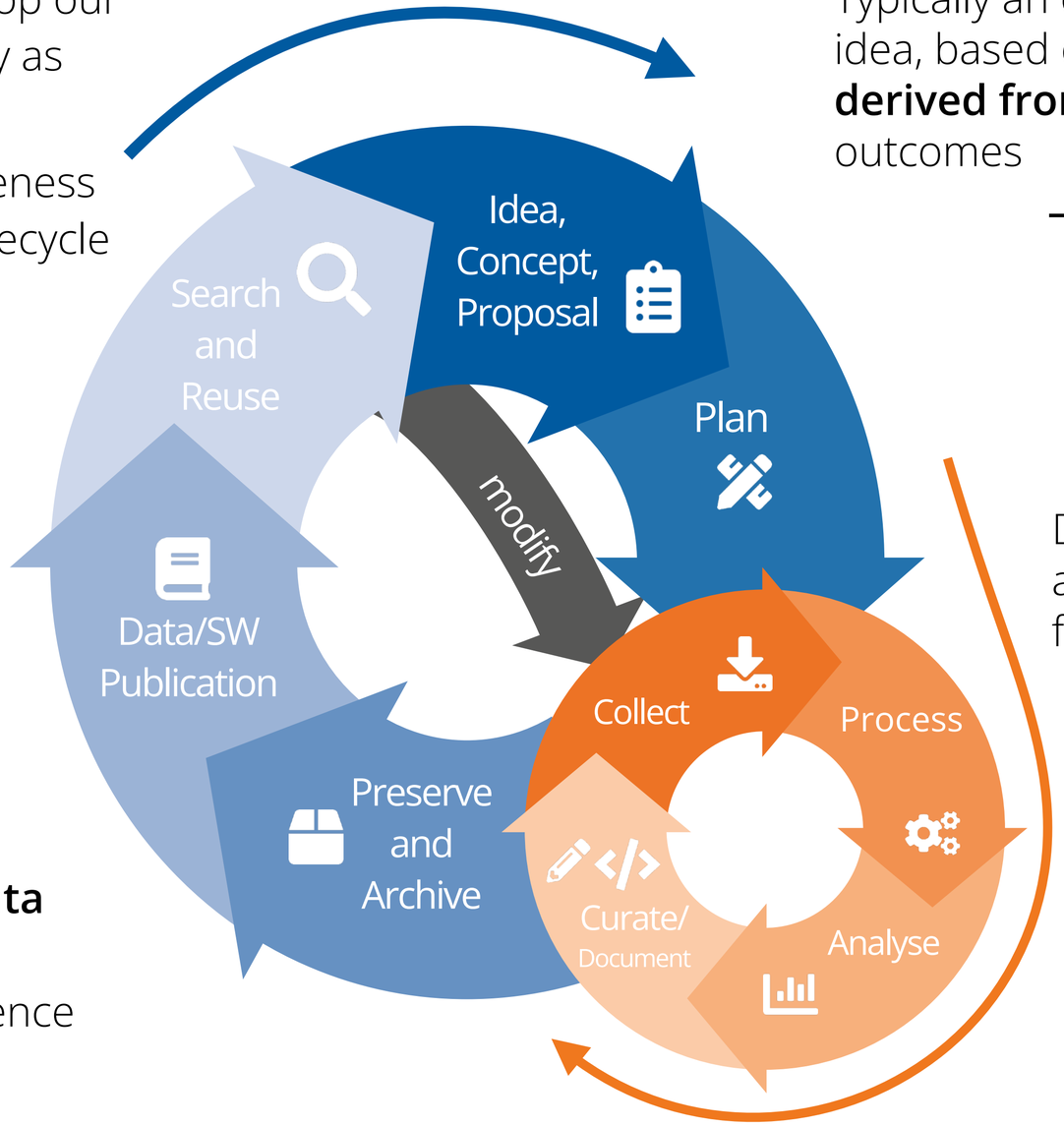
Data Management Lifecycle

— Necessary in order to develop our services as comprehensively as possible.

— Should also create an awareness for all components of the lifecycle

Publication of **data** and **metadata** is at the end of a typical research lifecycle and can be the starting point for new iteration cycles…

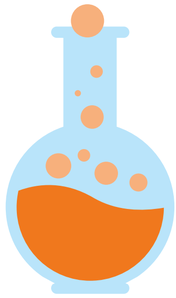
Archive or (additional) **data publication** to provide sustainable and open science



Typically an experiment starts with an idea, based on a research question, **derived from available research** outcomes

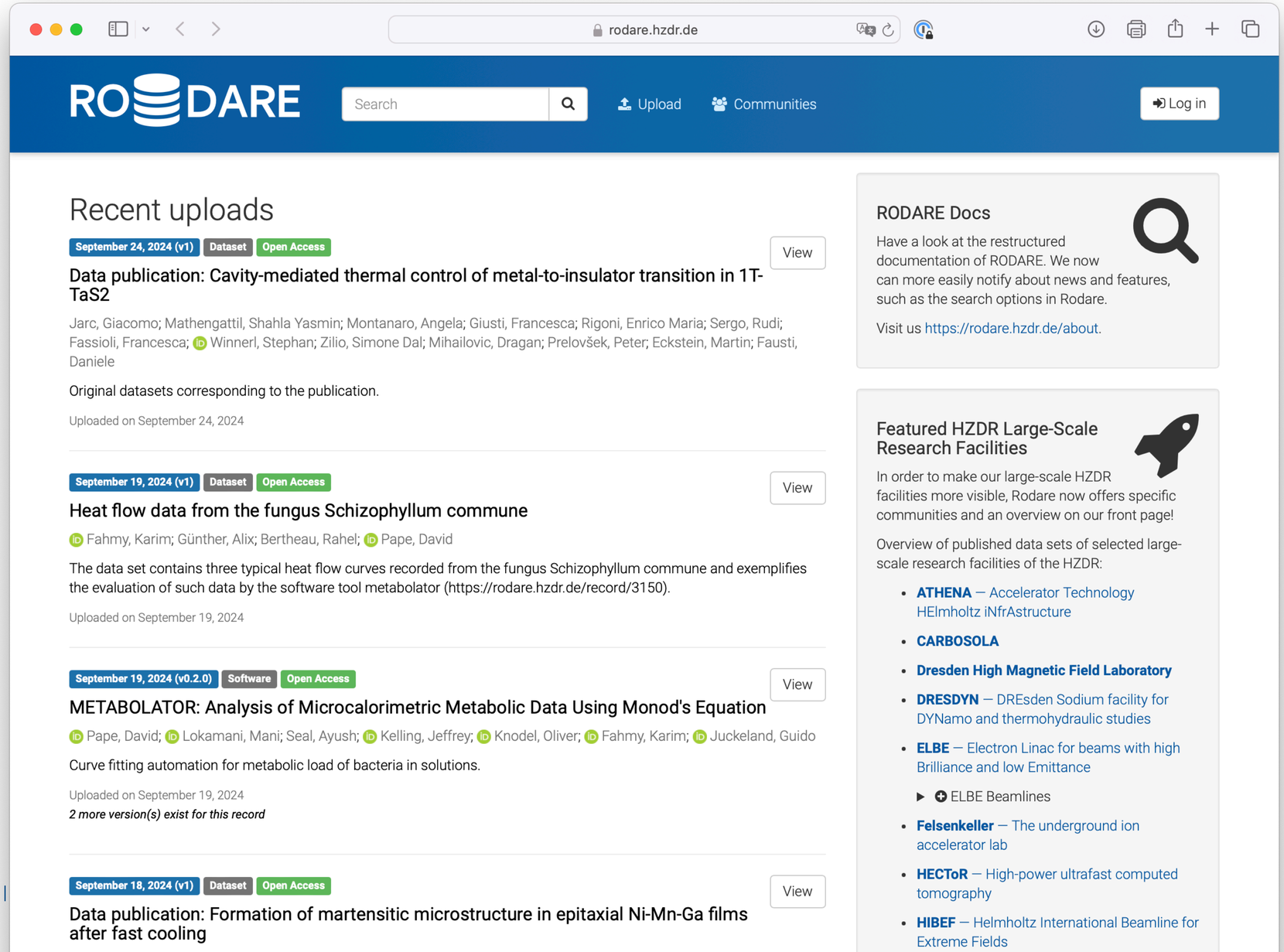
→ Valuable metadata in our proposal systems

During experiment **documentation** and **metadata** curation are essential for a **comprehensible** experiment



CC BY

The Data and Software Publication Platform RODARE



Powered by:

INVENIO

Harvested via OAI-PMH by:

B2FIND EUDAT

Registered in:

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

http://doi.org/10.17616/R3BR40

CC BY

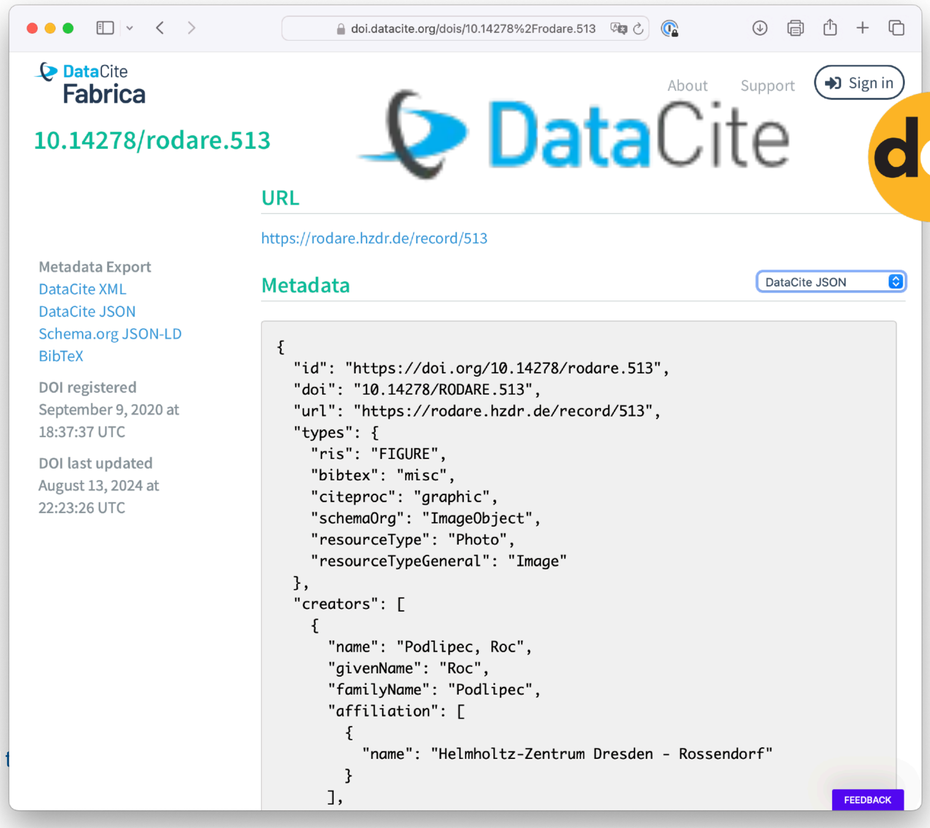
DRESDEN concept HZDR

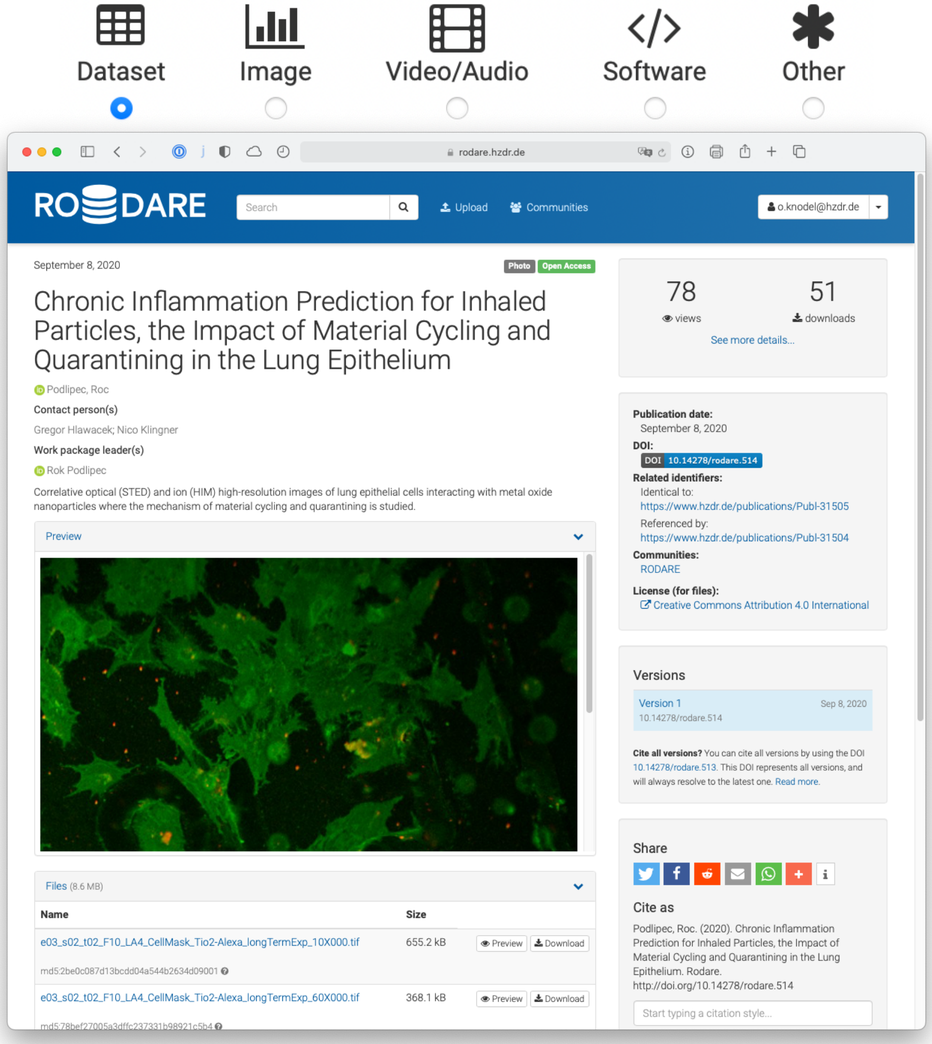
Metadata and Data

— In our data publication system the datasets are described via **DataCite** metadata to be FAIR.

— The DataCite metadata is attached to the DOI and harvested via portals, such as **B2Find**.

— In Theory: max 50 GB per file, max 100 GB per dataset (largest dataset: 7,2 TByte).





File Access:
Rest API

Filesystem

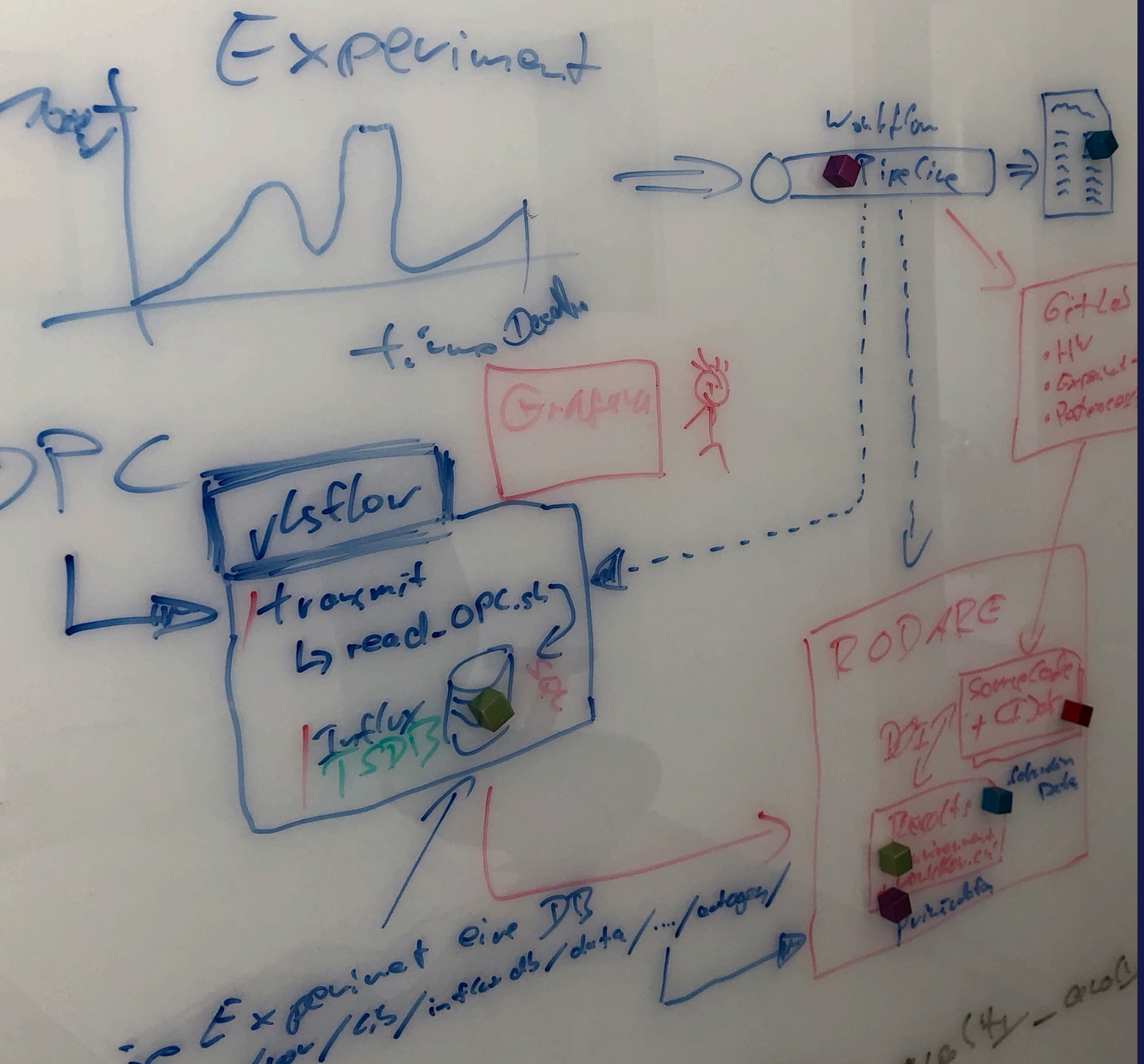
Metadata API:
Rest API, OAI-PMH

B2FIND
EUDAT

CC BY

DRESDEN concept

HZDR



The NAPMIX Project

OSCARS
Open Science Clusters' Action for Research & Society

Funded Project

NAPMIX - Nuclear, Astroparticle, and Particle Physics Metadata Integration for eXperiments

NAPMIX

Implemented by:

HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF
FAIR
HUN REN
atomki
KIT Karlsruher Institut für Technologie
GSI
DESY.
UAM Universidad Autónoma de Madrid
CMAM
IST TÉCNICO LISBOA
GANIL
TECHNISCHE UNIVERSITÄT DARMSTADT
CNA Centro Nacional de Aceleradores
HMC HELMHOLTZ Metadata Collaboration
CLEAR

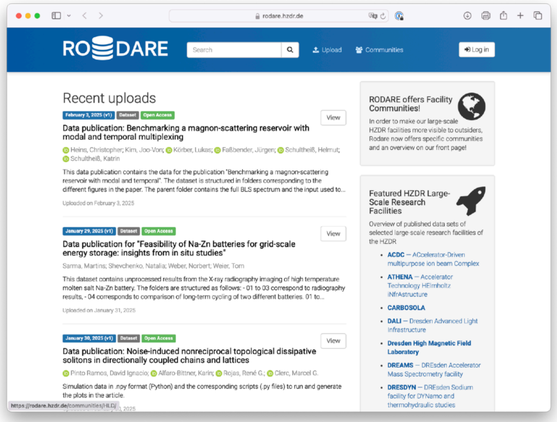
Funded by the European Union

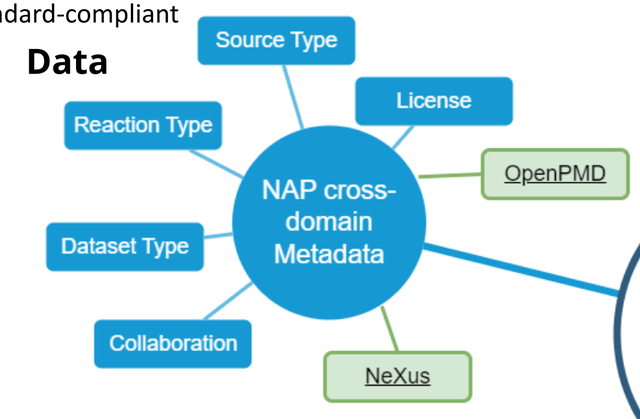
NAPMIX at the Felsenkeller

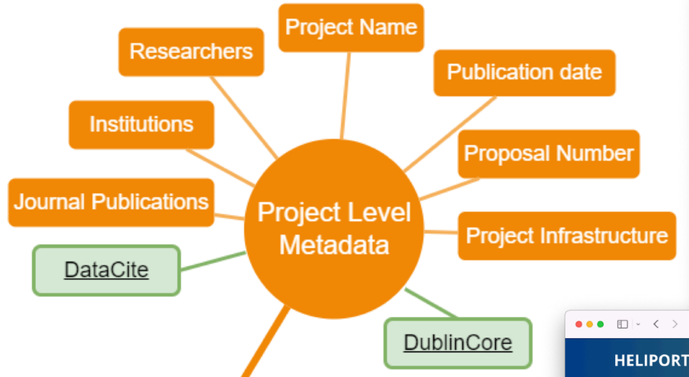
RODARE
ROSSENDORF DATA REPOSITORY

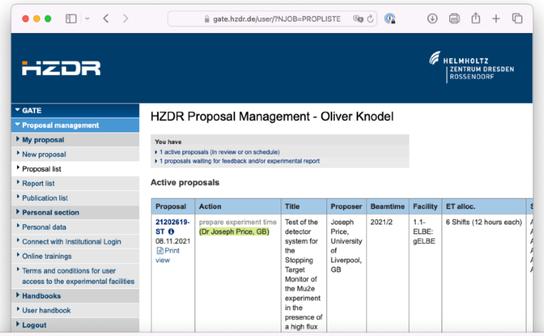
standard-compliant

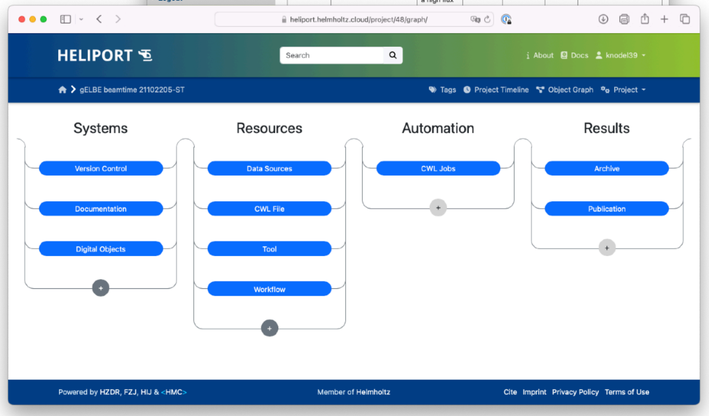
Data

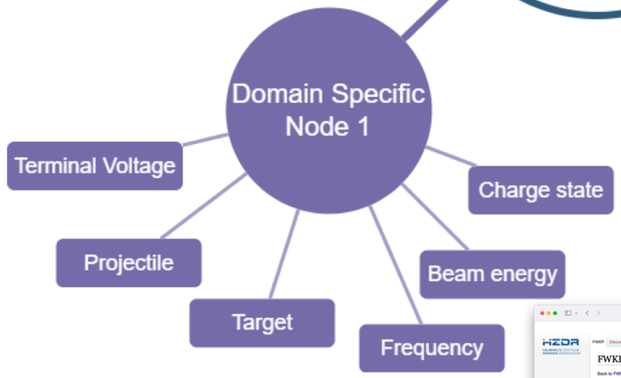


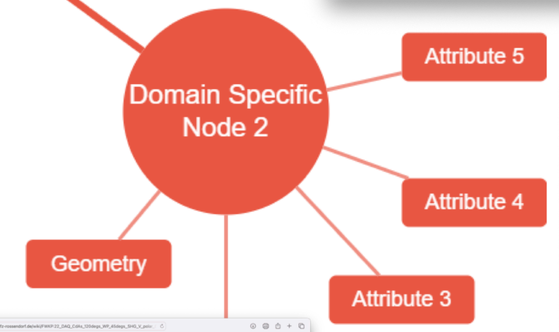




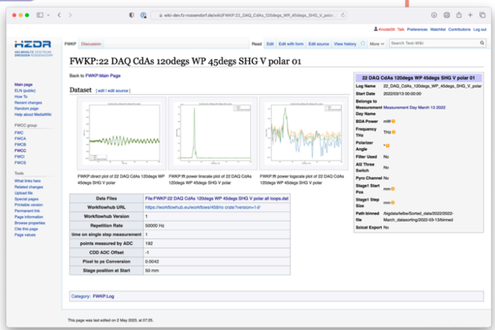


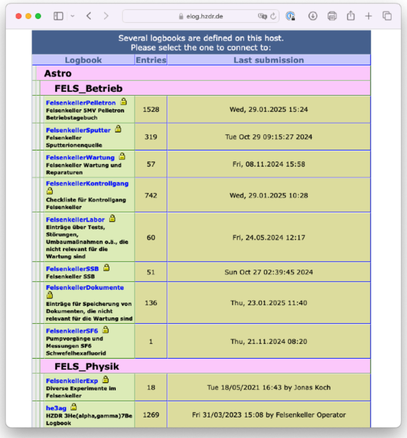






?





CC BY

DRESDEN concept

HZDR

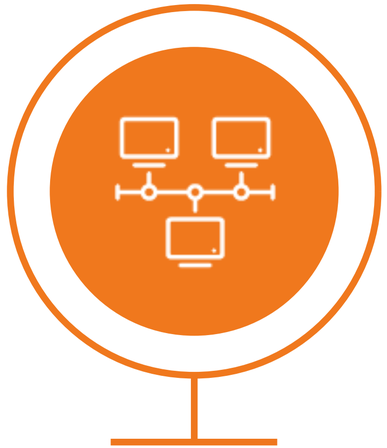
Gap Analysis of our Metadata Ecosystem

GAP

Some approach to provide additional searchable experiment- and discipline-specific metadata

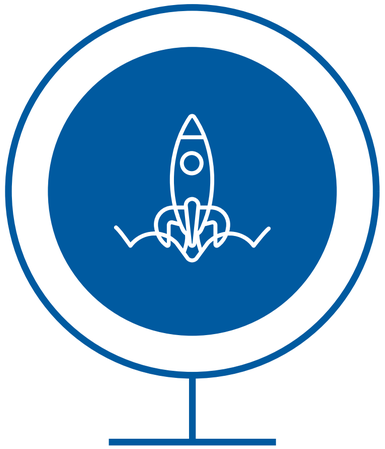
CURRENT STATE

DataCite metadata with only a few domain/experiment-specific fields (e.g. facility) in our data repository RODARE



PLAN

Additional Metadata catalogue between RODARE and the experiments



DESIRED STATE

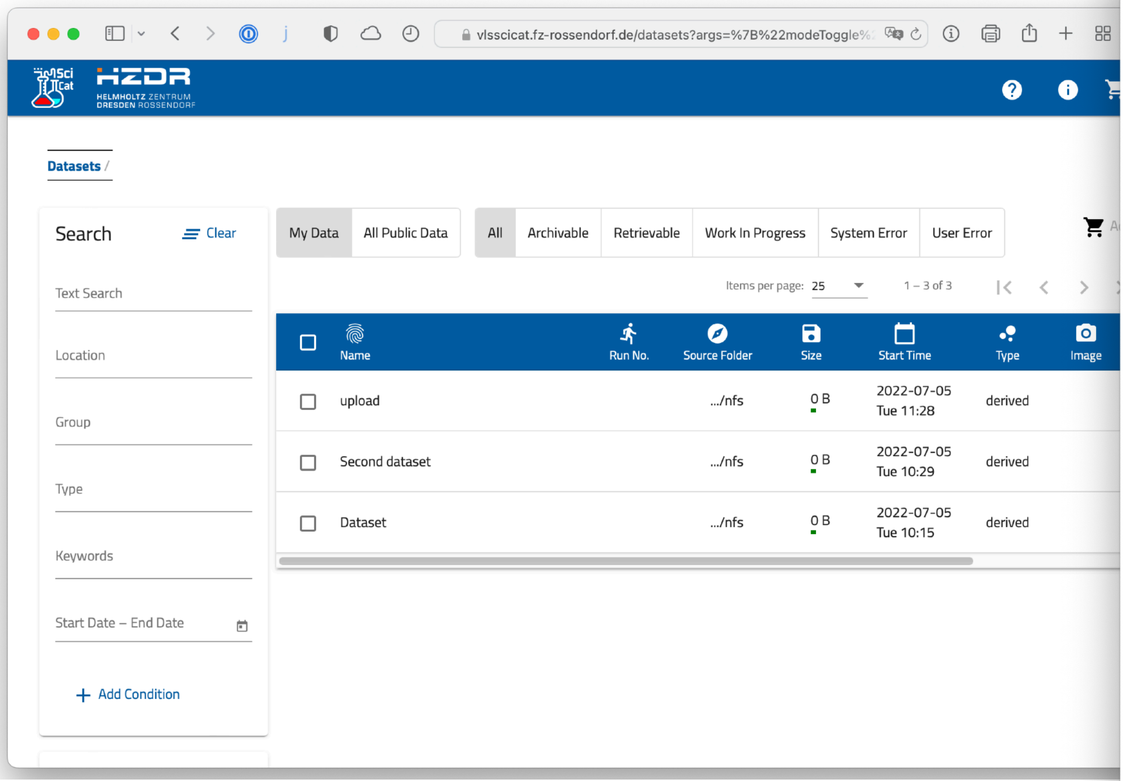
Additional community-specific metadata to search for and reference to the datasets (in RODARE, our archive or filesystem)

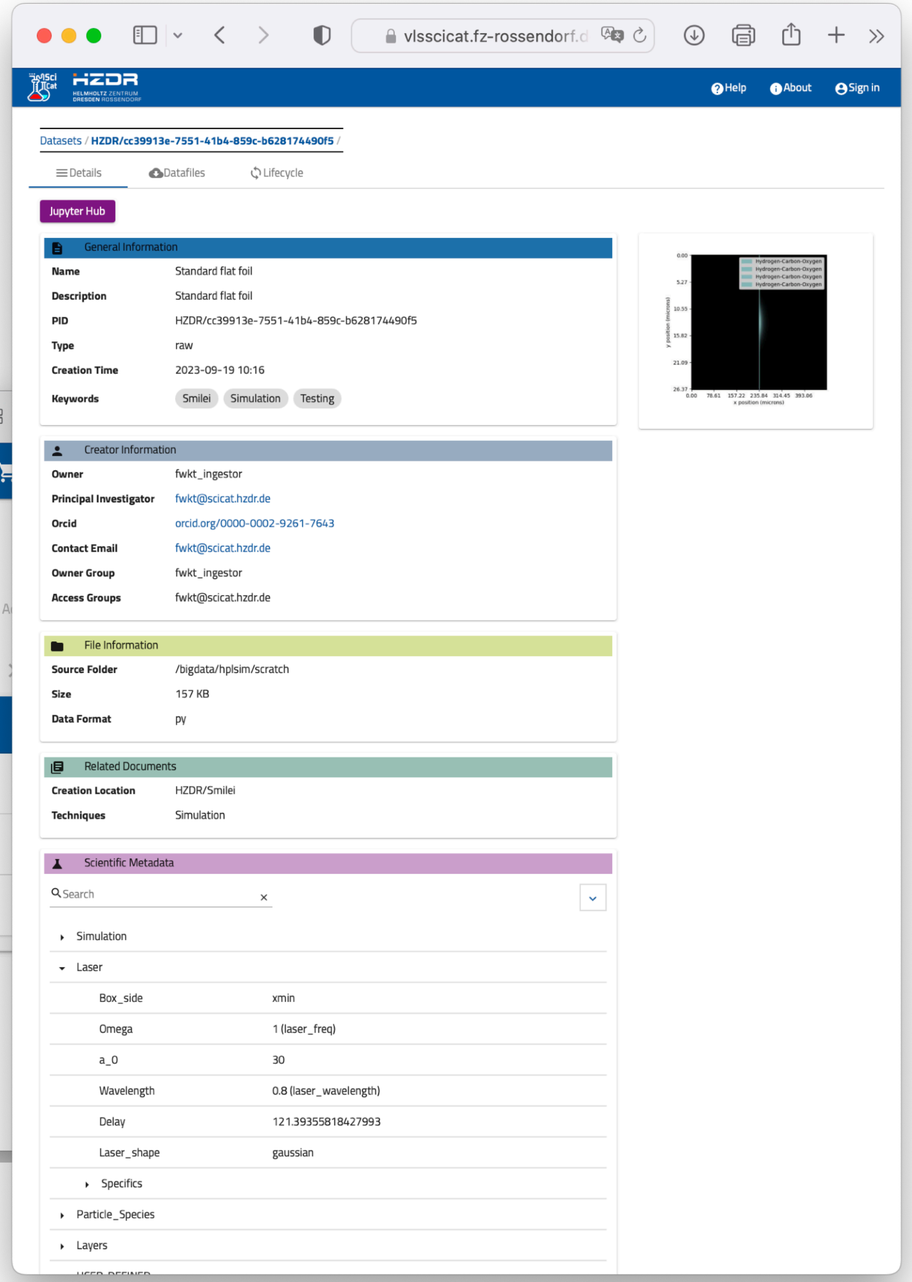
CC BY

Metadata Catalogue SciCat

— SciCat catalogue provides experiment-specific metadata.

— Developed in in the projects ExPANDS, PANOSC and DAPHNE.





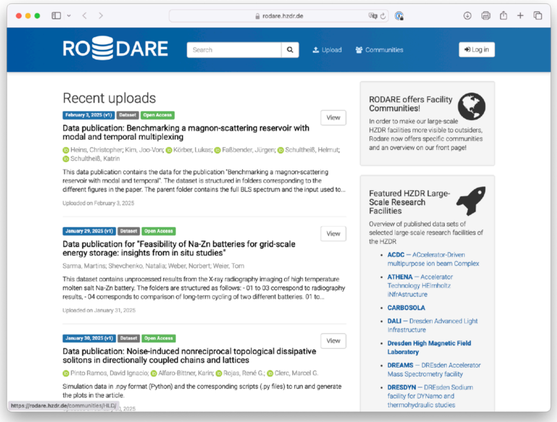
scicat.hzdr.de

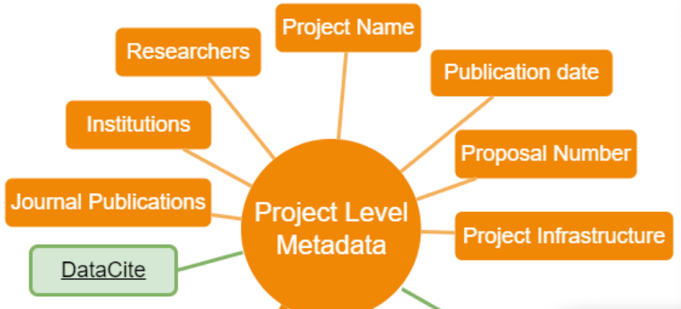
NAPMIX at the Felsenkeller

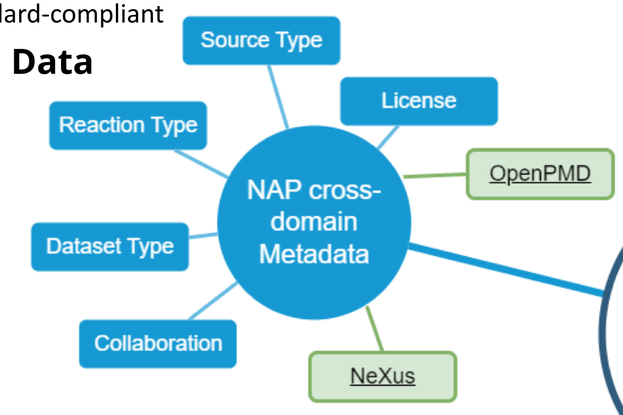
RODARE
ROSSENDORF DATA REPOSITORY

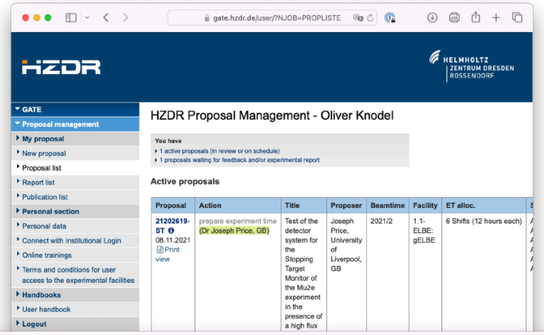
standard-compliant

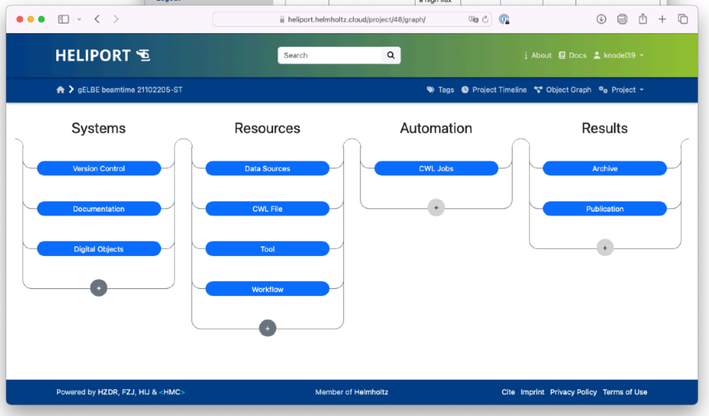
Data

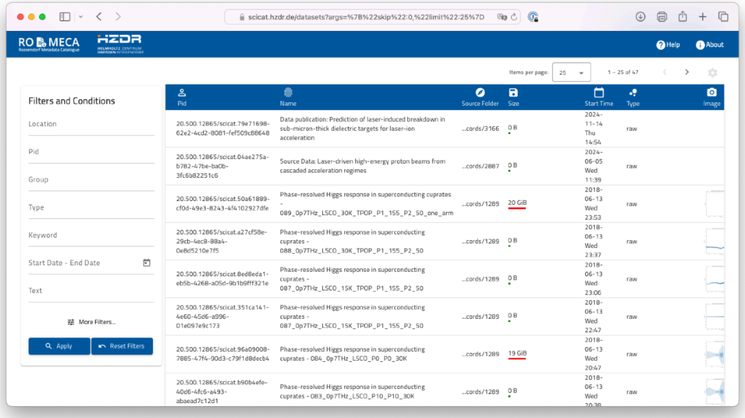


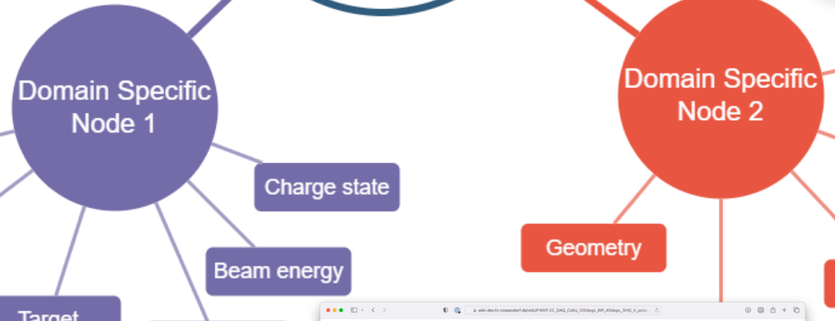


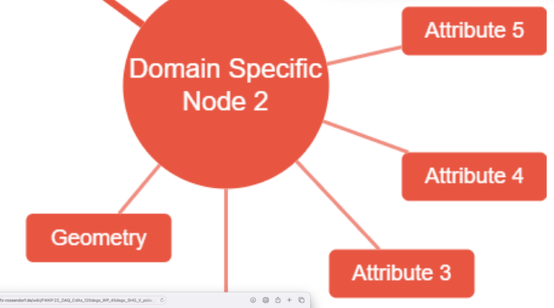


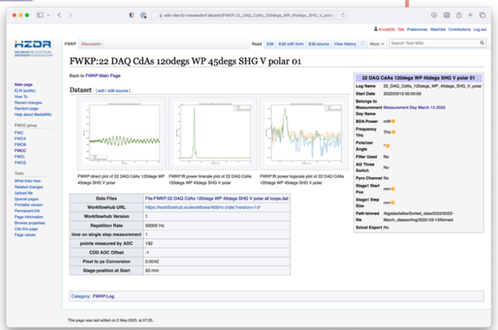


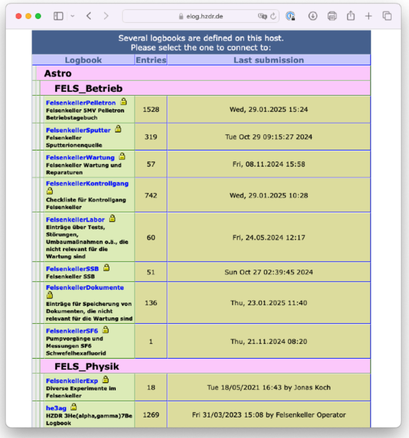












standard-compliant

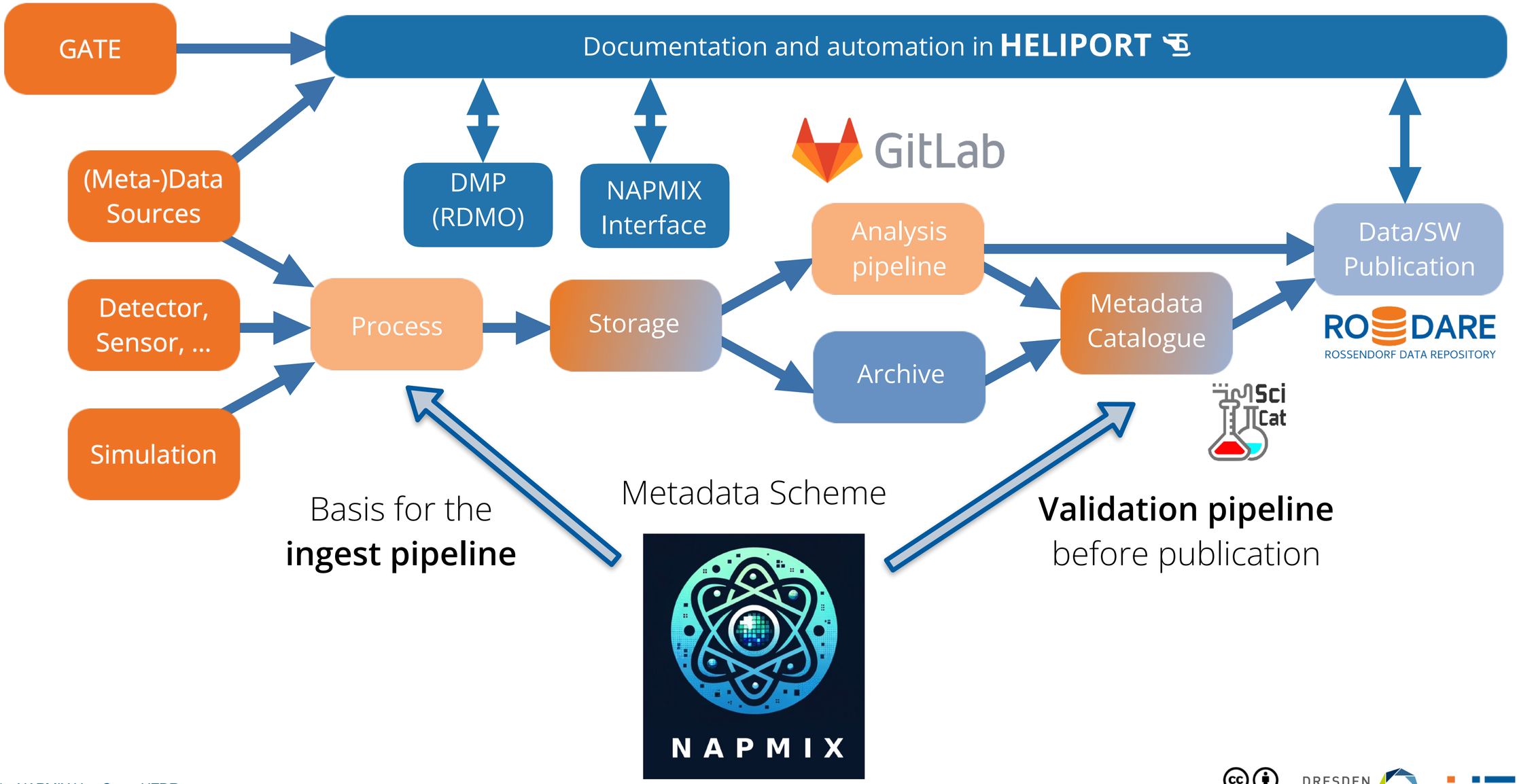
Metadata

CC BY

DRESDEN concept

HZDR

NAPMIX in the Open Science Ecosystem of HZDR



CC BY

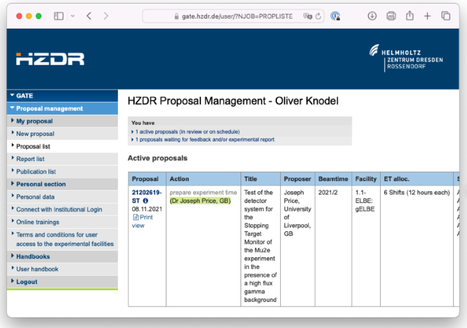
Embedding NAPMIX in the HZDR Data Management Chain

(Curated) Metadata Sources

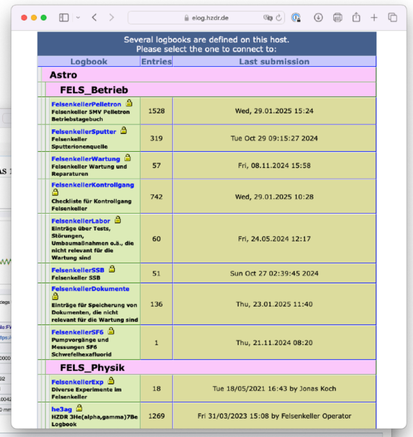
Proposal in GATE

Metadata from Experiment

HZDR GATE



ELOG



HZDR ELN

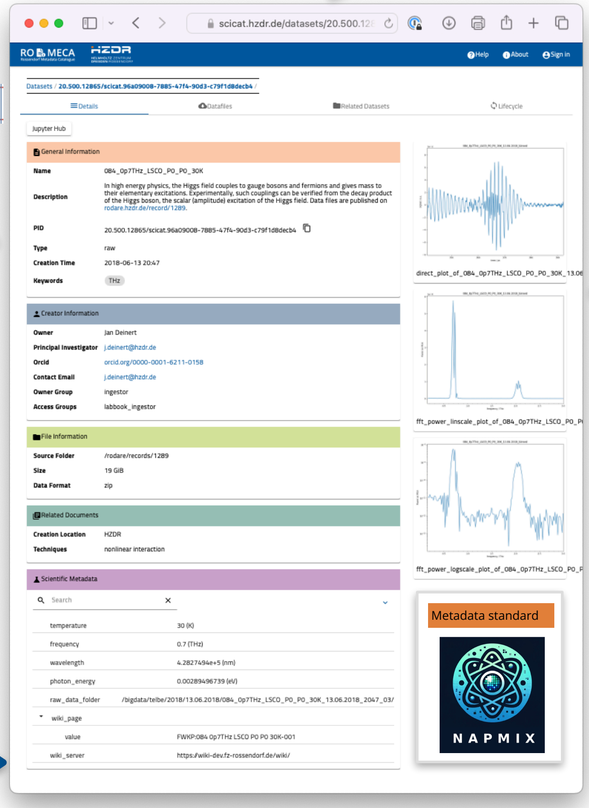
Automated Ingest Pipelines

Public Metadata Catalogue

SciCat

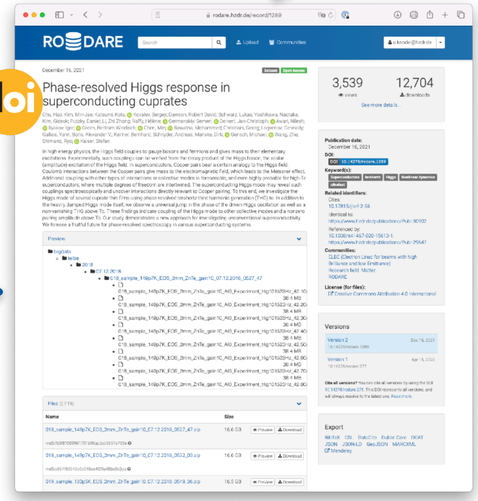
HZDR hdlenabled

panosc data portal



Subsequent Access to Data

RODARE (Invenio)

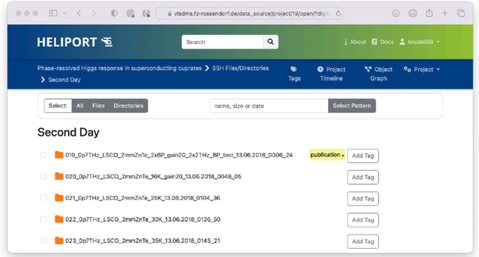


B2FIND EUDAT eosc

Dataset

Tape Archive

HELIPORT



Filesystem

Access: Private/Internal

Access: Public

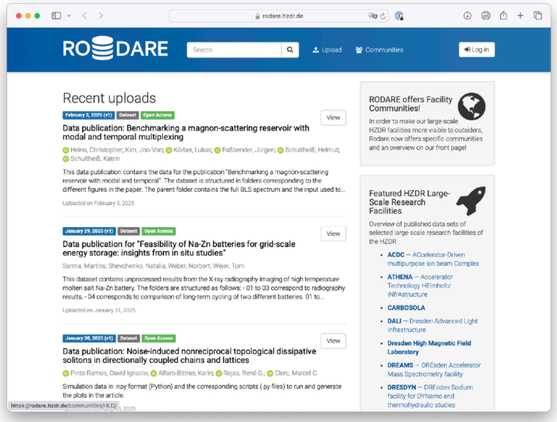
CC BY

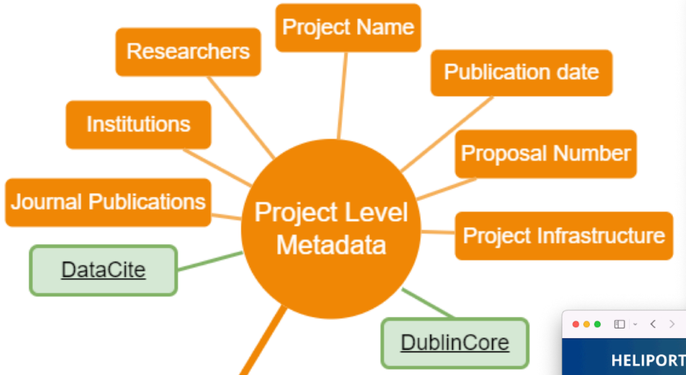
NAPMIX at the Felsenkeller

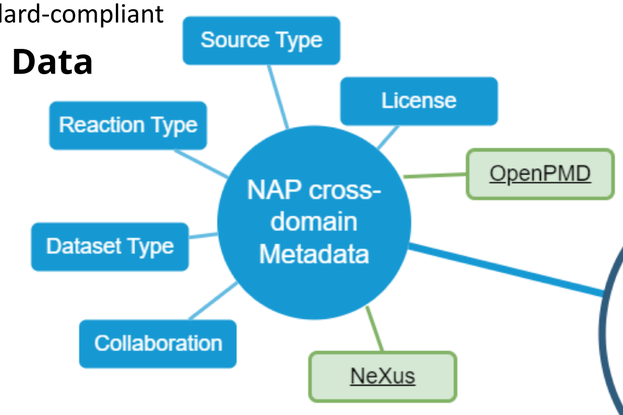
RODARE
ROSSENDORF DATA REPOSITORY

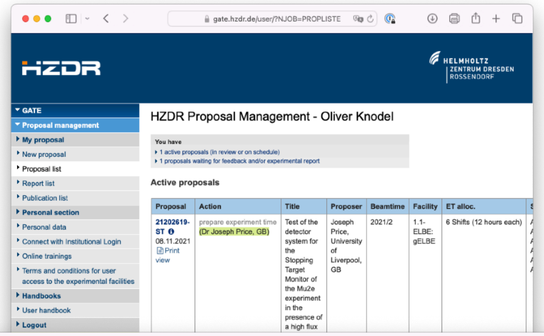
standard-compliant

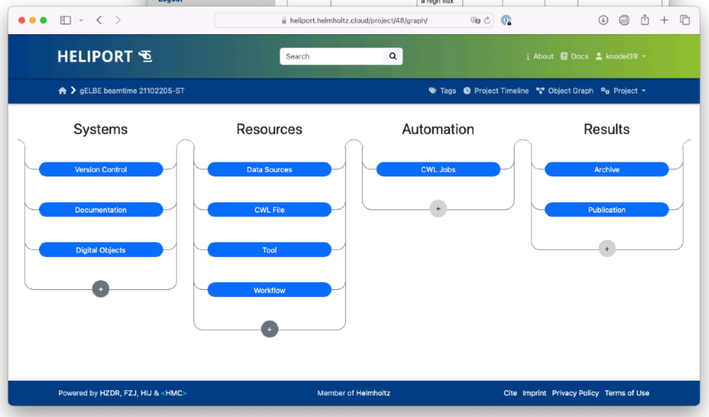
Data

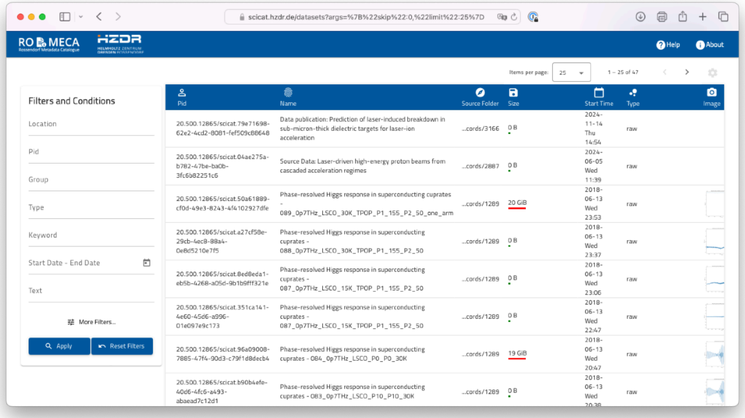


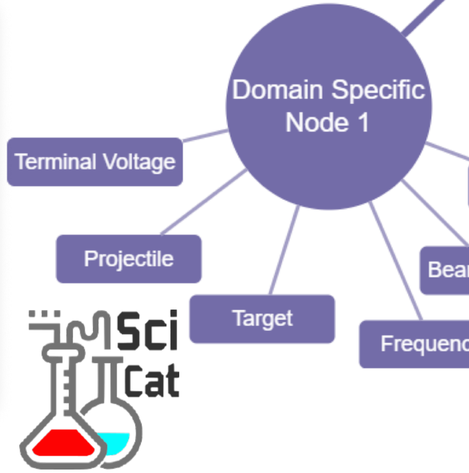


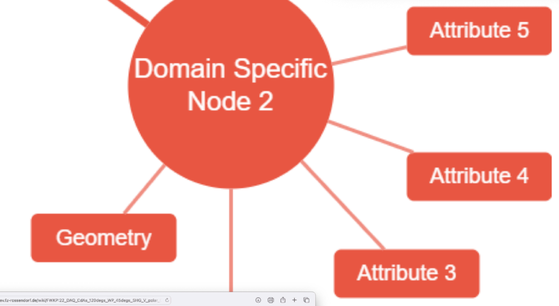


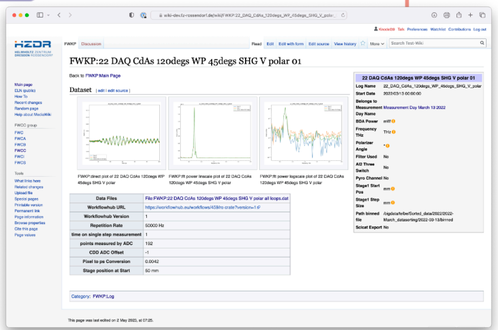


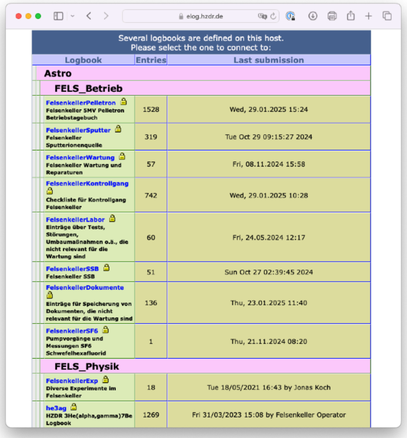












SciCat

standard-compliant

Metadata

CC BY

DRESDEN concept

HZDR