

Connecting Metadata, Data, and Software Repositories in a Generic Data Management Lifecycle

Oliver Knodel¹, Maik Fiedler², Thomas Gruber¹, Lokamani¹, Stefan E. Müller¹, Martin Voigt¹, David Pape¹ and Guido Juckeland¹

¹ Department of Scientific Computing, Helmholtz-Zentrum Dresden-Rossendorf | ² Library, Helmholtz-Zentrum Dresden-Rossendorf

The connection between metadata, data, and software and the integration in an overall lifecycle is crucial for effective data management in research. The generic data management lifecycle, developed at HZDR, bridges these critical components, ensuring seamless data discovery, accessibility, and reproducibility.

The approach emphasises the planning of experiments, the role of metadata, data storage, as well as software versioning, and the final publication of digital research artefacts, which enables comprehensive traceability from data creation to long-term archiving.

By aligning these elements in a unified procedure, we recommend a uniform lifecycle that can be adapted to different research areas, with a particular focus on services from our various communities, such as SciCat, that improves data integrity and promote collaborative research.

Share, Reuse and Search

Scientific Publications based on data and software should cite them or include a statement indicating where and under what conditions the data, supplementary metadata, documentation, software, are accessible and reusable.

Data, digital artefacts available for discovery and access may be reused by other researchers, either to corroborate the findings of the original research or to gain new insights through further interrogation and analysis. At this stage, the data can become raw material collected for a new research cycle.

Data/Software Publication

The publication of data, software and documentation includes the conversion of file formats, the creation of descriptive metadata records with the assignment of digital object identifiers (DOIs) and the licensing of data records for reuse.

Preserve and Archive

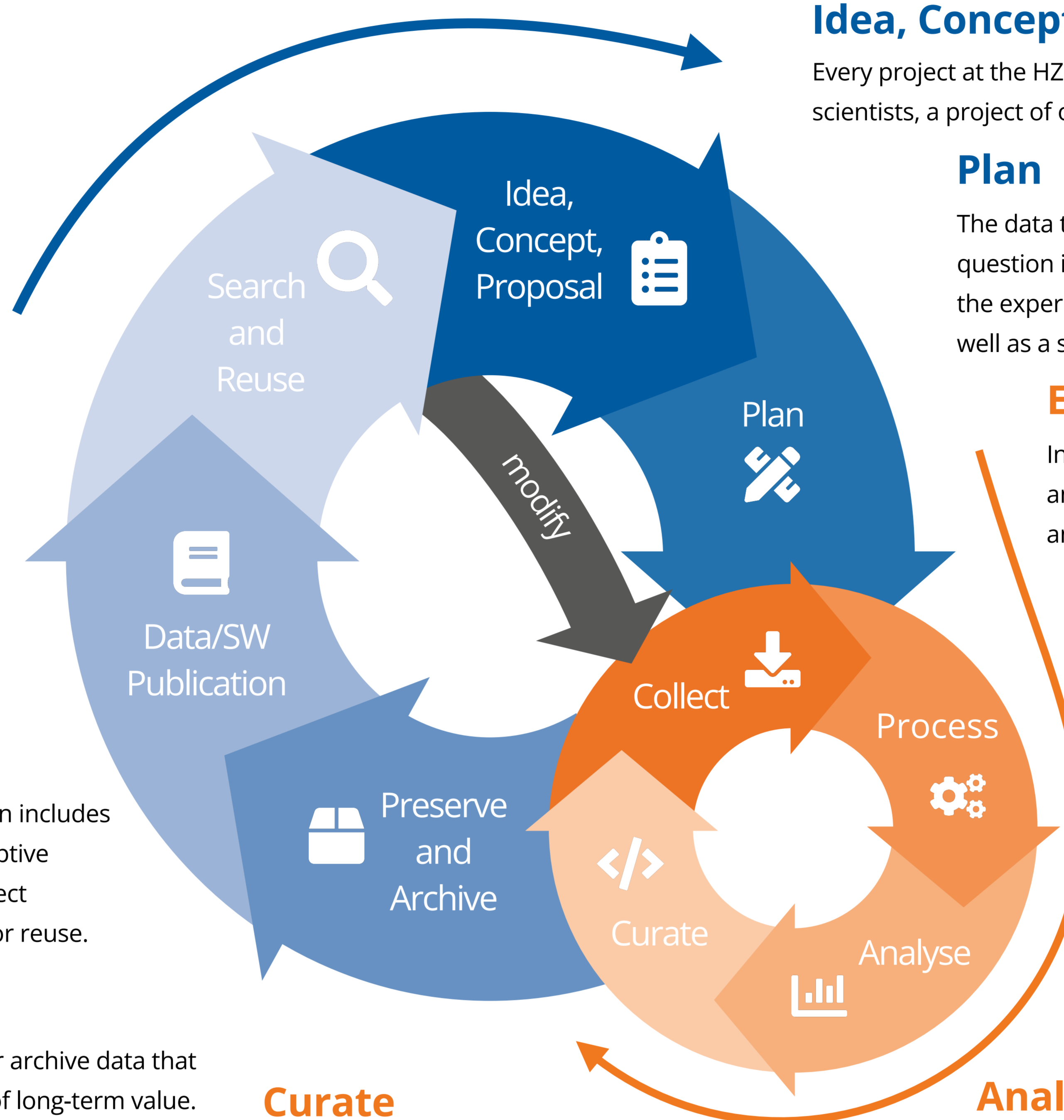
Preservation activities may include to store or archive data that substantiate your research findings and are of long-term value. A data storage/data centre is suitable for digital data. For metadata and software, suitable measures such as metadata catalogues or software repositories should be selected. The cross-linking of all data products is important in this context.

Curate

In the curation phase, the data is provided with additional metadata and a decision is made as to whether the quality of the data and the scientific results are sufficient or whether changes to the structure are necessary in order to collect additional data.

Analyse

The analysis interrogates the raw material of the research to gain the insights that make up the research findings that are written up and published in the research results. The methods used for the analysis should be documented; the code written for the analysis and visualisation may need to be retained and made available.



Idea, Concept, Proposal

Every project at the HZDR usually starts with a proposal from external scientists, a project of our associates or simply with an idea.

Plan

The data that is to be collected or used to answer the research question is determined and data management is planned throughout the experiments life cycle. The phase in which data management-, as well as a software management plans, are created.

Experiment Cycle

In this phase, the actual data is collected, pre-processed, analysed, but also enriched with corresponding metadata and finally curated before further iteration takes place:

Collect

This is the phase in which experiments are conducted, observations made, surveys carried out, secondary material obtained, etc. This includes the documentation of the setup and the data generated data along with the associated metadata.

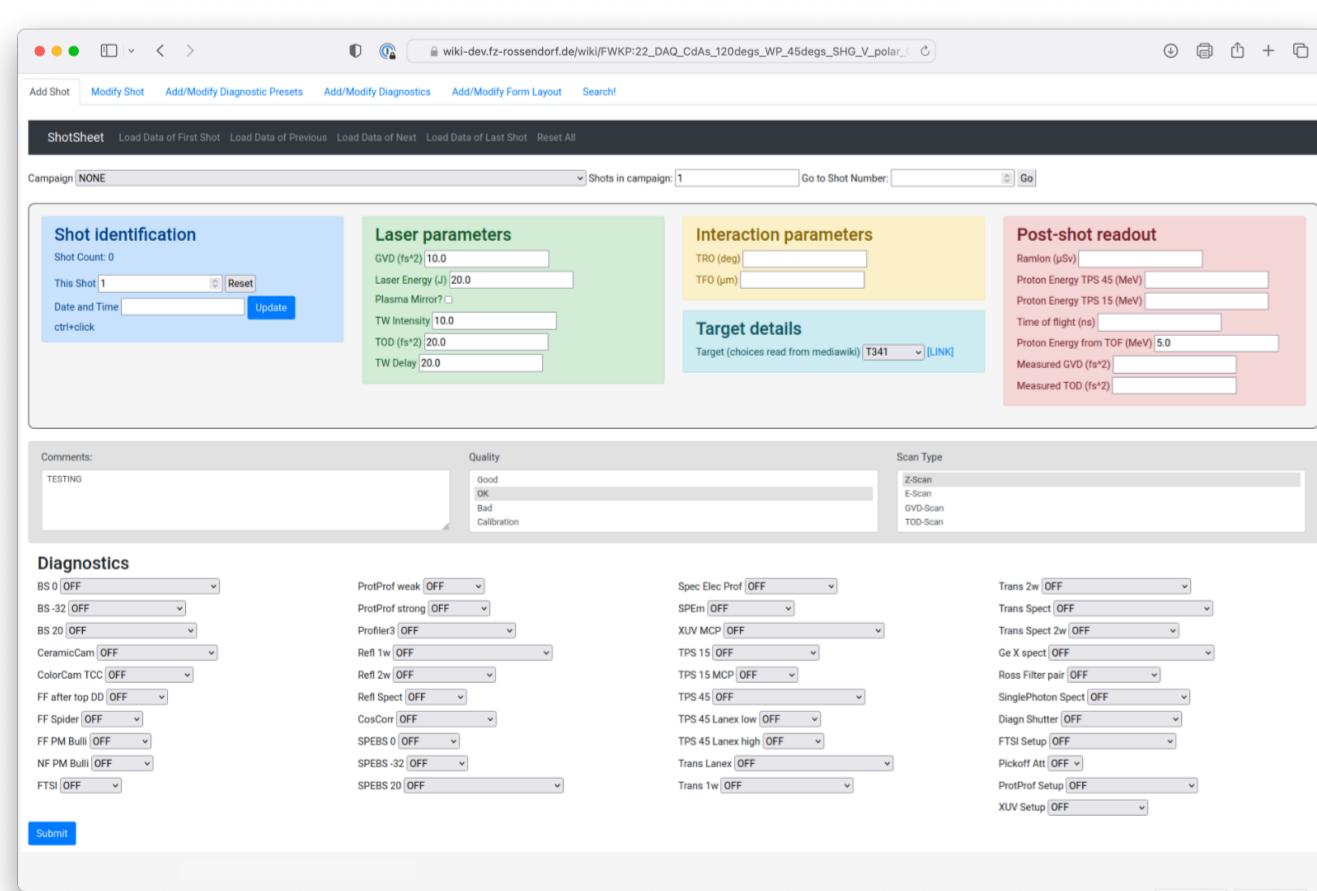
Process

Once data has been collected, it must be processed (e.g. cleaned, combined, converted, validated) to make it usable. All data processing must be documented so that the final result can be reproduced using the raw data.

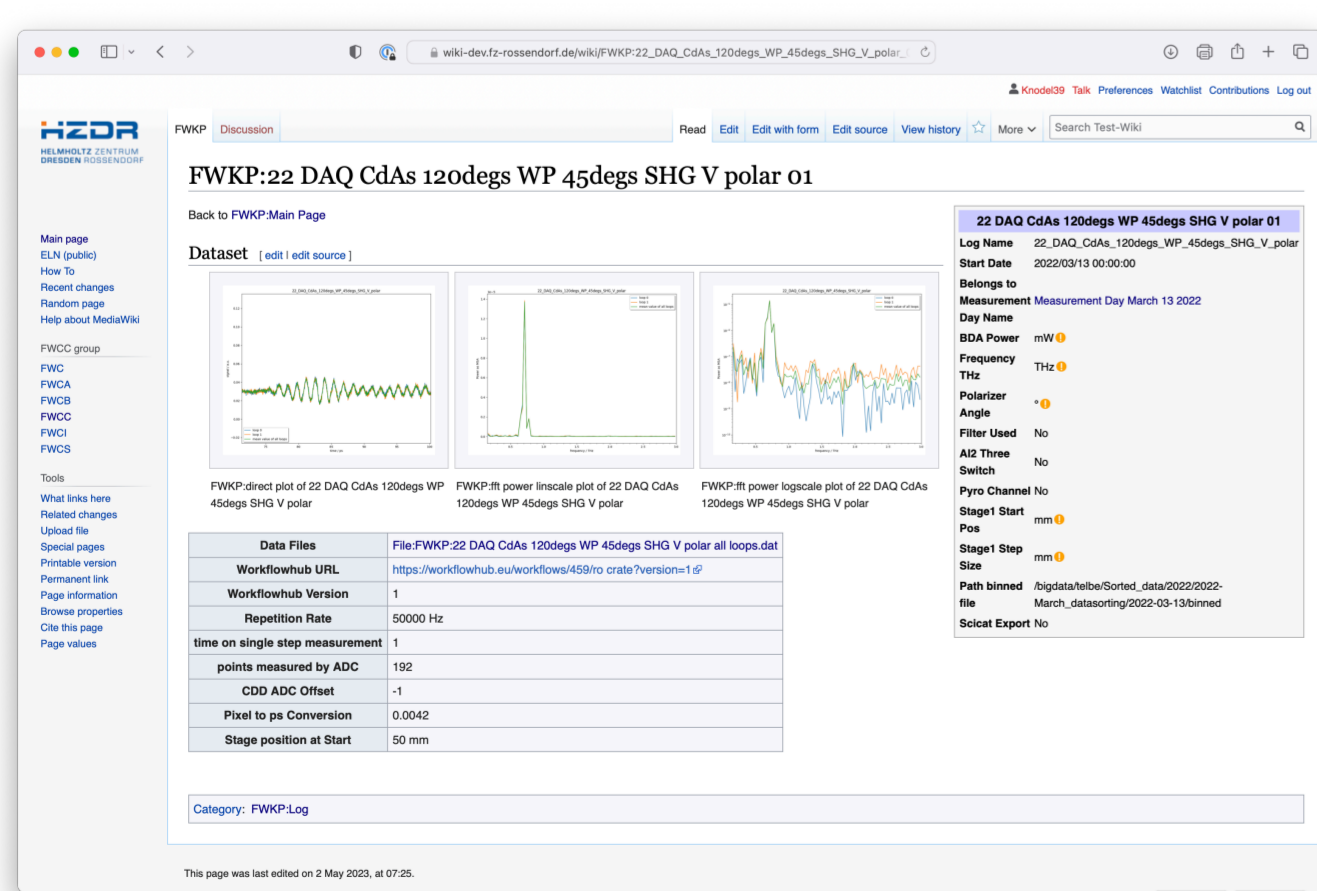
From the Raw Metadata in the Experiment Through Linked Metadata to the Final FAIR Data Publication

Curated Metadata Sources

ExperimentLogging app (ExL)



E-Logbook



Fully Automated Process for DRACO

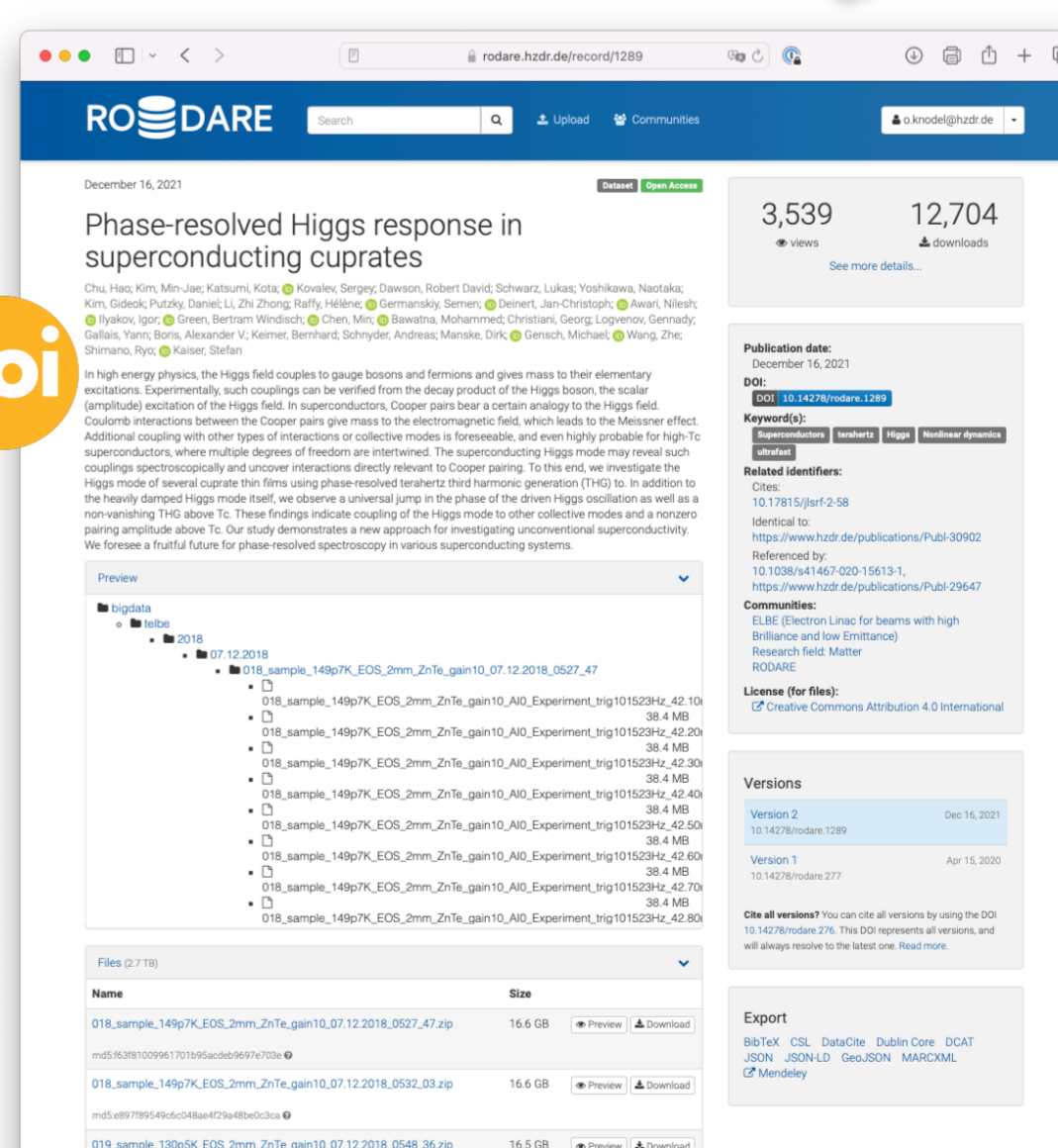
Public Metadata Catalogue

SciCat

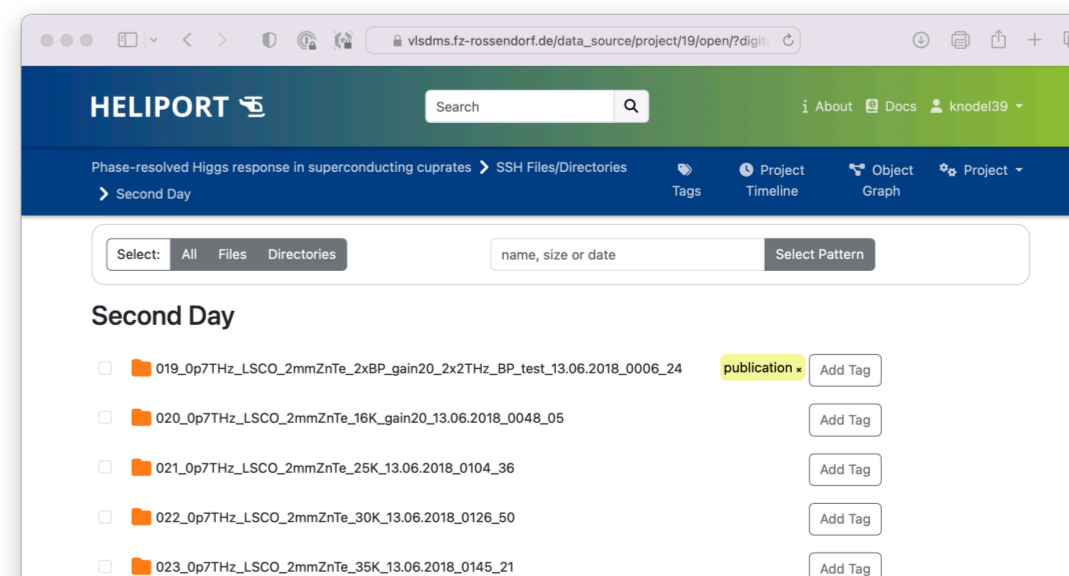


Subsequent Access to Data

RODARE (Invenio)



HELIPORT



Dataset



Tape Archive



Filesystem

