Pioneering Digital Research Landscapes: Innovations at HZDR

Towards Open Digital Research Ecosystems – Interconnecting Infrastructures, Open Science Forum, February 14, 2024

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

Our Research Facility and our Large Scale Research Infrastructures

The Helmholtz-Zentrum Dresden - Rossendorf

— Employees approx. 1,470. Thereof 670 scientists

— **HELMHOLTZ**
RESEARCH FOR GRAND CHALLENGES

Research Fields

— Energy, Health and Matter

ELBE – Center for High-Power Radiation Sources

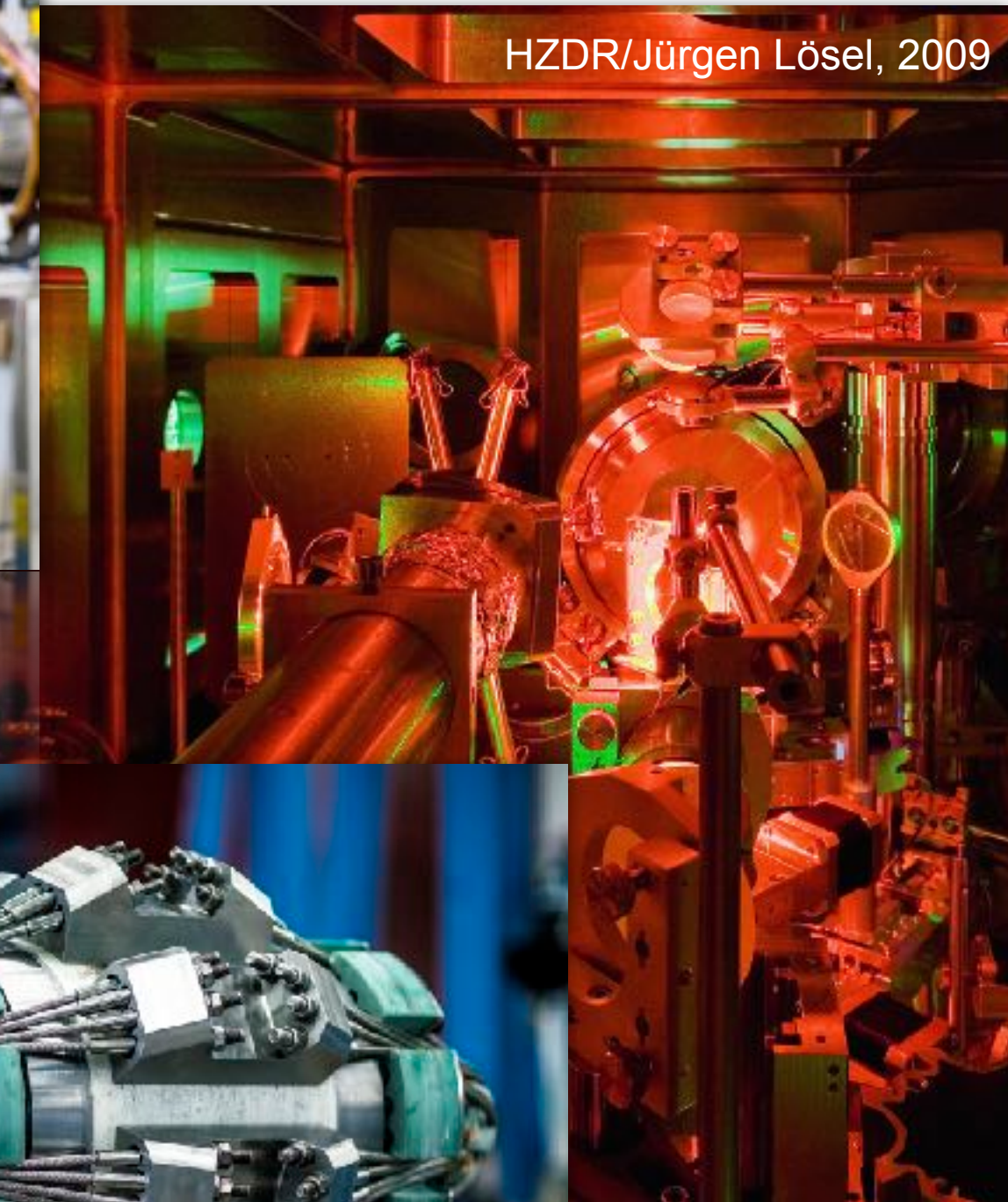— Electron accelerator, free-electron lasers & THz source.

— Positrons, protons, neutrons as well as X-ray and gamma radiation

Dresden High Magnetic Field Laboratory (HLD)
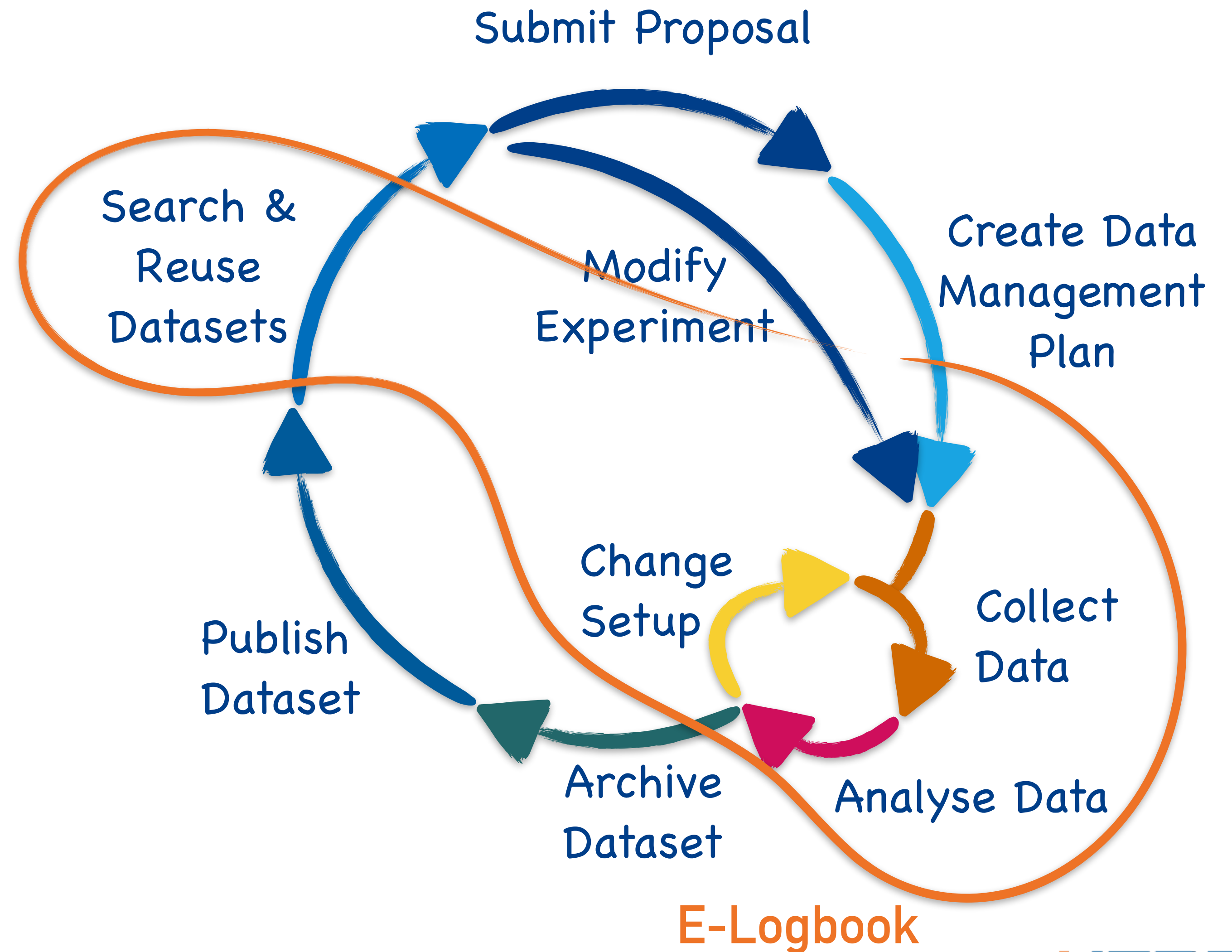
— Europe's highest pulsed magnetic fields

Ion Beam Center (IBC)

— Nanoscale surface analysis and modification

HZDR/Jürgen Lösel, 2009

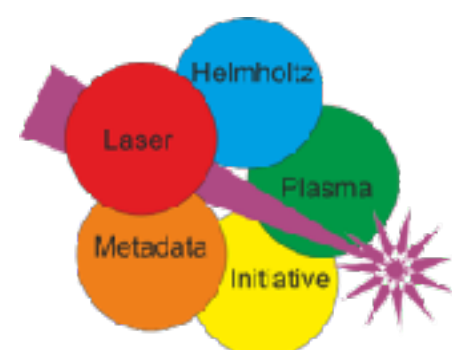HZDR/André Wirsing, 2017

HZDR/Oliver Killig, 2013

HZDR

Our Challenge: An End-to-End Digital Data Lifecycle

— We support many steps of our different research experiment (matter, energy and health) with a wide range of tools:

- Electronic lab notebook (**E-Logbook**),
- Interactive analysis,
- **FAIR Publication** of datasets,
- Scientific **workflow** management,
- **Handle** (PID) generation and management.

— A uniform and smooth access to and **between** all services and systems in our ecosystem is necessary.

— The documentation of all these linked resources is essential to create a **comprehensible** and **FAIR** data lifecycle.
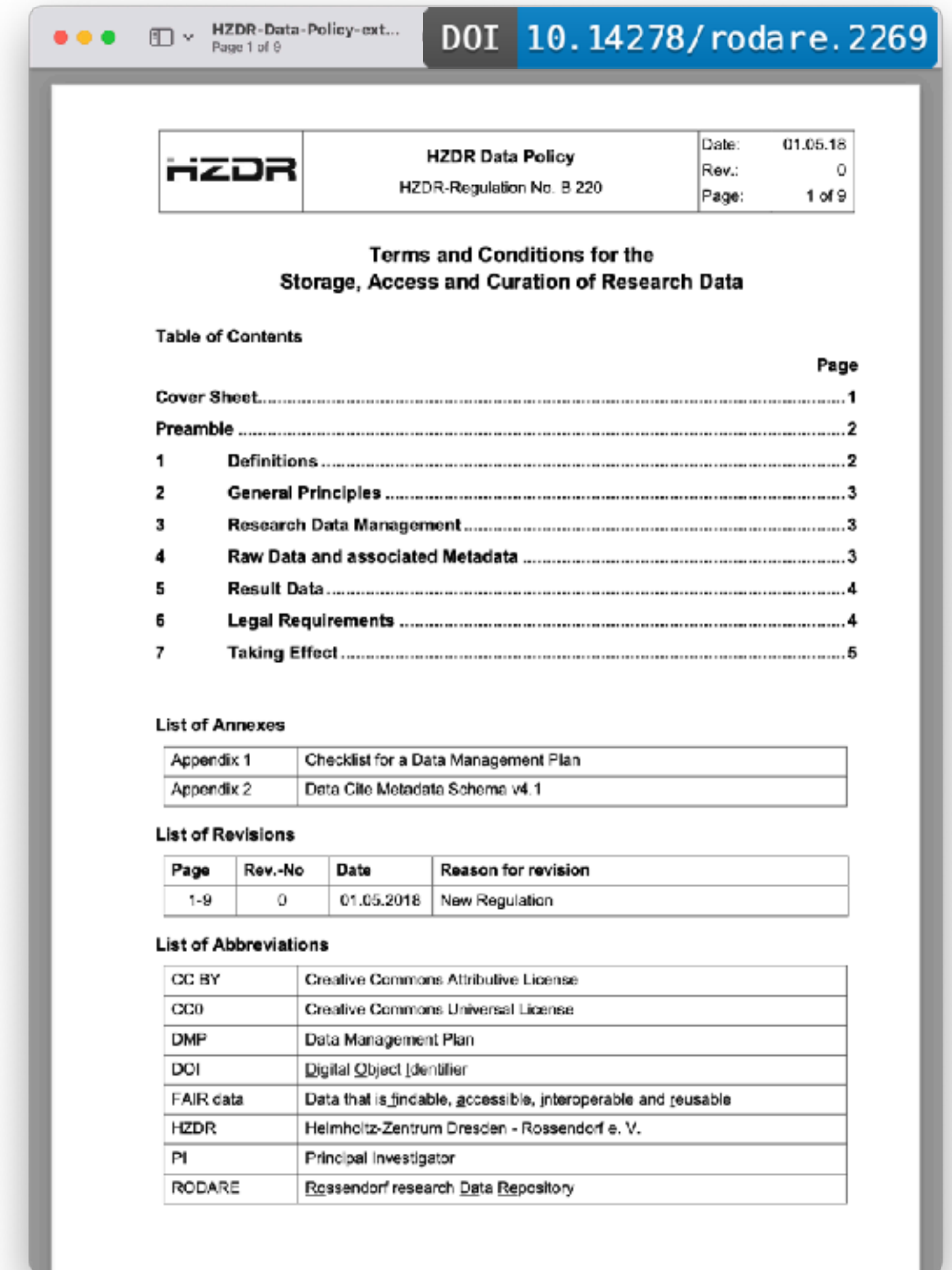
Submit Proposal

Search & Reuse Datasets

Modify Experiment

Create Data Management Plan

Change Setup

Collect Data

Publish Dataset

Archive Dataset

Analyse Data

E-Logbook

HZDR

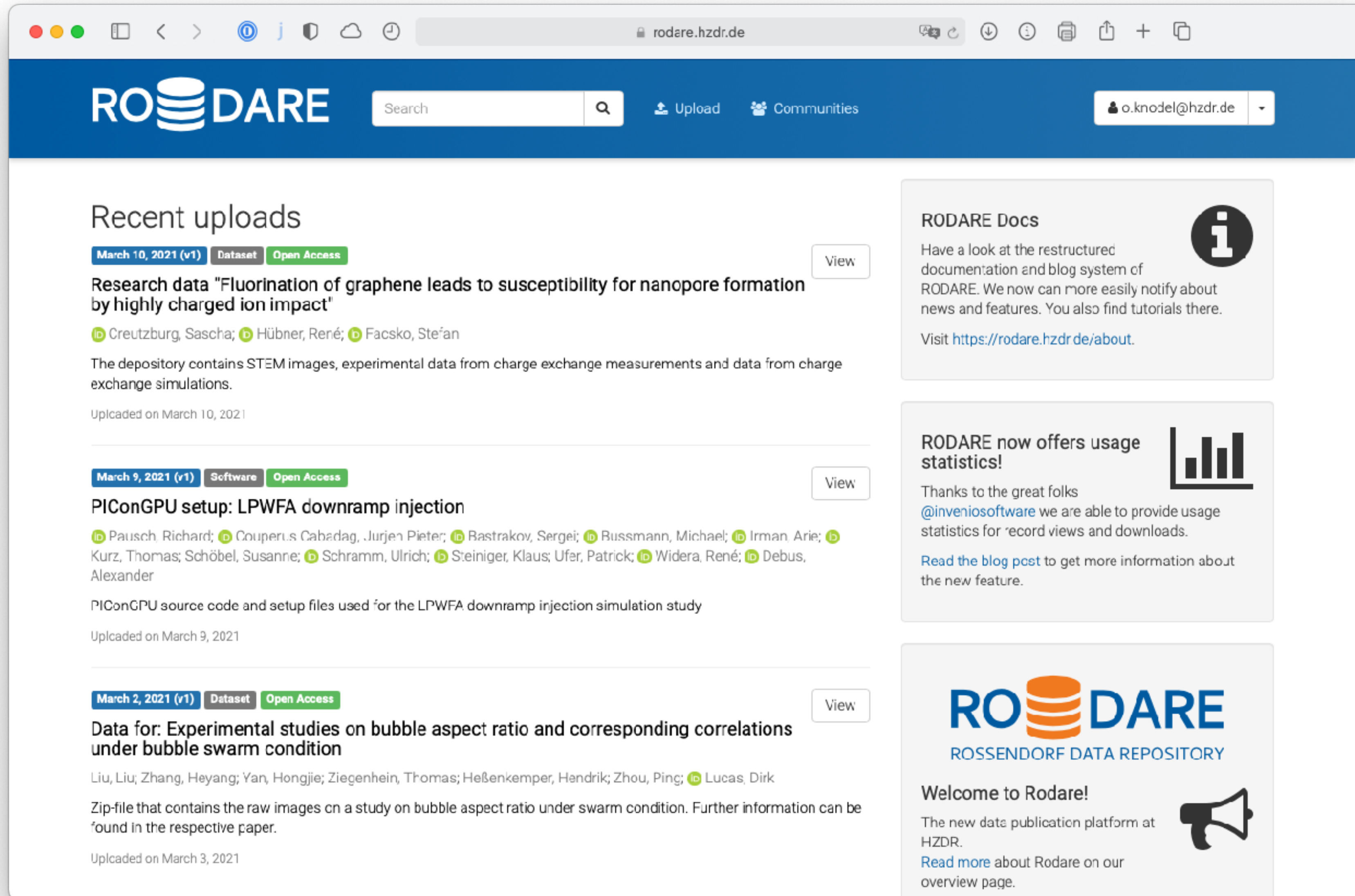# The Foundation for our Digital Data Lifecycle: HZDR Data Policy

— The HZDR has a data policy since **May 2018**

— Reasons for the development:

- Legal framework for data management and publication,

- Establishment and legitimisation of coordinated research data management. at HZDR,

— Foundation for the development of tools and services to support our scientists:

© Helmholtz-Gemeinschaft 2023, https://os.helmholtz.de/open-research-data/forschungsdaten-policies/

Starting Point in 2018: Data Publication Platform RODARE 🌐 rodare.hzdr.de

Powered by:

Registered in:
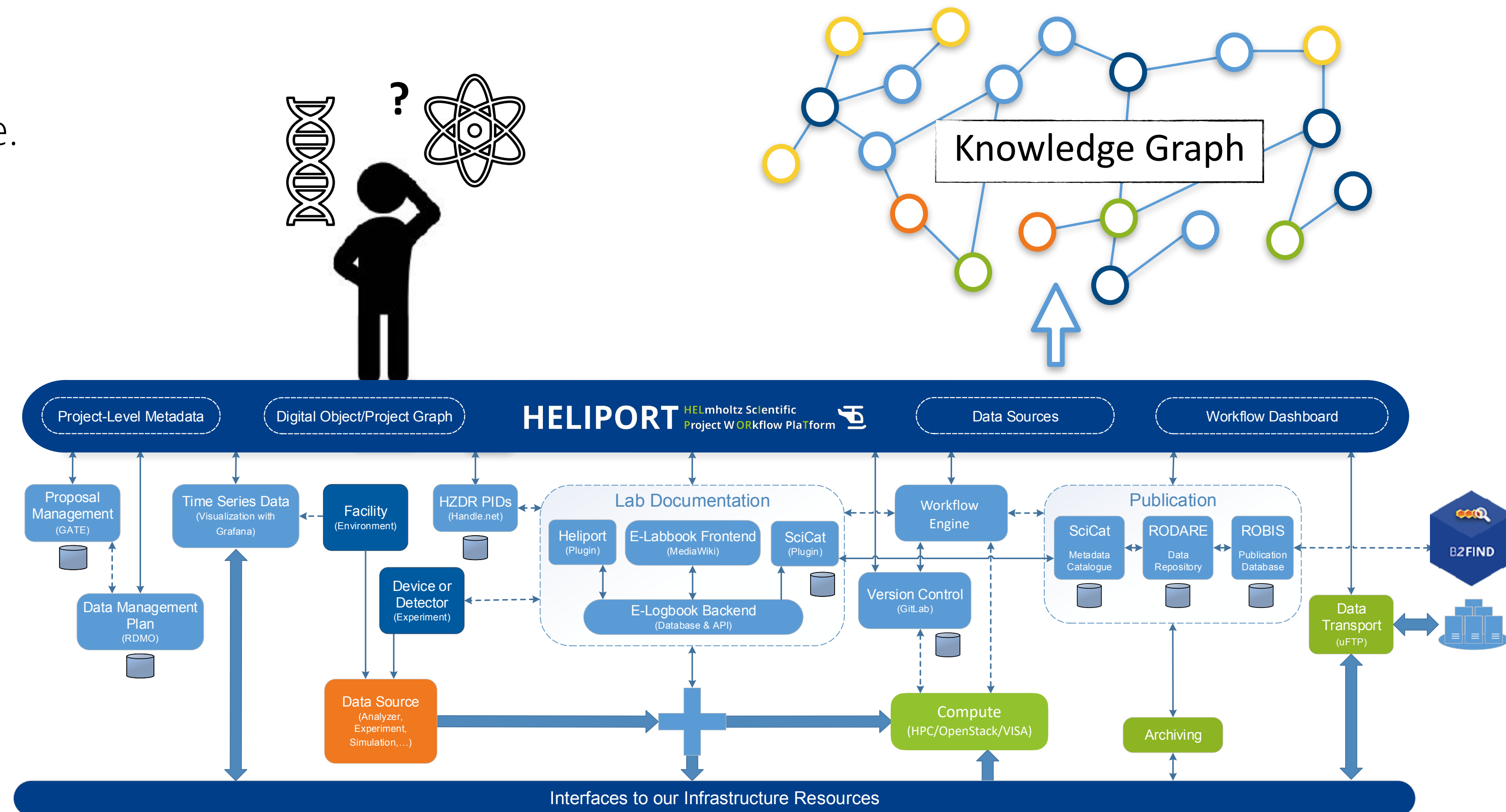
http://doi.org/10.17616/R3BR40

Findable with:

Our Observations Over the Following Years…

— We need to support the entire experiment with reliable **interconnected tools** to enable comprehensible and FAIR science.

— The resulting IT infrastructures are complex.

— Documentation is necessary, but typically time is missing.

— Scientists often don't know which services are available at facilities and how to use them.

— An overarching system **guiding our scientists** (and visitors) through the lifecycle of their research project is essential.

HELIPORT

HELmholtz **ScI**entific **P**roject **WOR**kflow **Pla**Tform

Project Members:

Funded by:

HZDR
HELMHOLTZ ZENTRUM
DRESDEN ROSSENDORF

HI JENA
Helmholtz Institute Jena

JÜLICH
FORSCHUNGSZENTRUM

<HMC> HELMHOLTZ Metadata Collaboration

> The HELIPORT project aims at developing a platform which accommodates the **complete life cycle** of a scientific project and links all corresponding programs, systems and workflows to create a more **FAIR** and comprehensible project description.

HELIPORT

Phase-resolved Higgs response in superconducting cuprates

Search   ℹ About   📄 Docs   👤 knodel39 ▾

🏷 Tags   🕐 Project Timeline   Object Graph   Project ▾

Systems   Resources   Automation   Results

Version Control   Data Source   UNICORE Jobs   Archive

Data Management Plan   SSH Files/Directories   Publication

Documentation   UNICORE Storages

Digital Objects

Powered by HZDR, FZJ, HIJ & <HMC>   Member of Helmholtz   Imprint   Privacy Policy   Terms of Use

ToDo: Metadata crosswalk to **schema.org ResearchProject**

```
{
  "namespaces": {
    "datacite": "http://purl.org/spar/datacite/",
    "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
    "heliport": "https://heliport/schema/",
    "time": "http://www.w3.org/2006/time#",
    "dc": "http://purl.org/dc/terms/"
  },
  "heliport:project_id": 28,
  "datacite:hasIdentifier": "HZDR.FWCC.2021.84769",
  "heliport:uuid": "09779261-200c-48c4-be9c-f298369d6a1c",
  "datacite:handle": "https://hdl.handle.net/None",
  "heliport:project_name": "PaN Research Project",
  "time:hasBeginning": "2021-04-01 09:14:34.296524+00:00",
  "datacite:hasDescription": "",
  "heliport:group": "FWCC",
  "heliport:owner": {
    "datacite:hasIdentifier": "132739",
    "datacite:orcid": null,
    "rdfs:label": "Knodel, Dr. Oliver (FWCC) - 132739"
  },
  "heliport:has_VersionControl": [
    {
      "heliport:version_control_id": 15,
      "datacite:uri": "https://dd",
      "rdfs:label": "
    }
  ],
  "heliport:
  {
    "helip
    "dataci
    "dataci
  }
  ],
  "heliport:ha
  {
    "heliport:documentation_id": 7,
    "datacite:uri": "https://dddd",
    "heliport:documentation_system": "MediaWiki1",
    "datacite:hasDescription": "dddd"
  }
  ],
  "heliport:has_DataSource": [
    {
      "heliport:data_source_id": 11,
      "datacite:uri": "http://ddd",
      "heliport:use_computer": null,
      "rdfs:label": "ddd",
      "datacite:hasDescription": ""
```

Requirements and Expectations

— HELIPORT was intended to provide only the **proposal's metadata**, from internal and external scientists, to allow the assignment of resources.

— Over time, we realised that HELIPORT can also answer our scientists' most important questions, such as:

And how we can support them?!

What are the necessary steps towards a full comprehensible and FAIR research experiment ensuring data provenance?

IT ?

Good Scientific Practice

Datasets

Data Policy

Horizon 2020 FAIR Data

How can we **automate recurring processes** and keep track of status and data products?

Which datasets or software can be **published** (and how)?

doi

Data Publication Repository

Publication Policy

How can we bring **new team members** or **visiting/external scientists** into our project lifecycle and all associated tools?

Where are data, software and how can I gain **access** to both of them?

Software

GitLab

HZDR
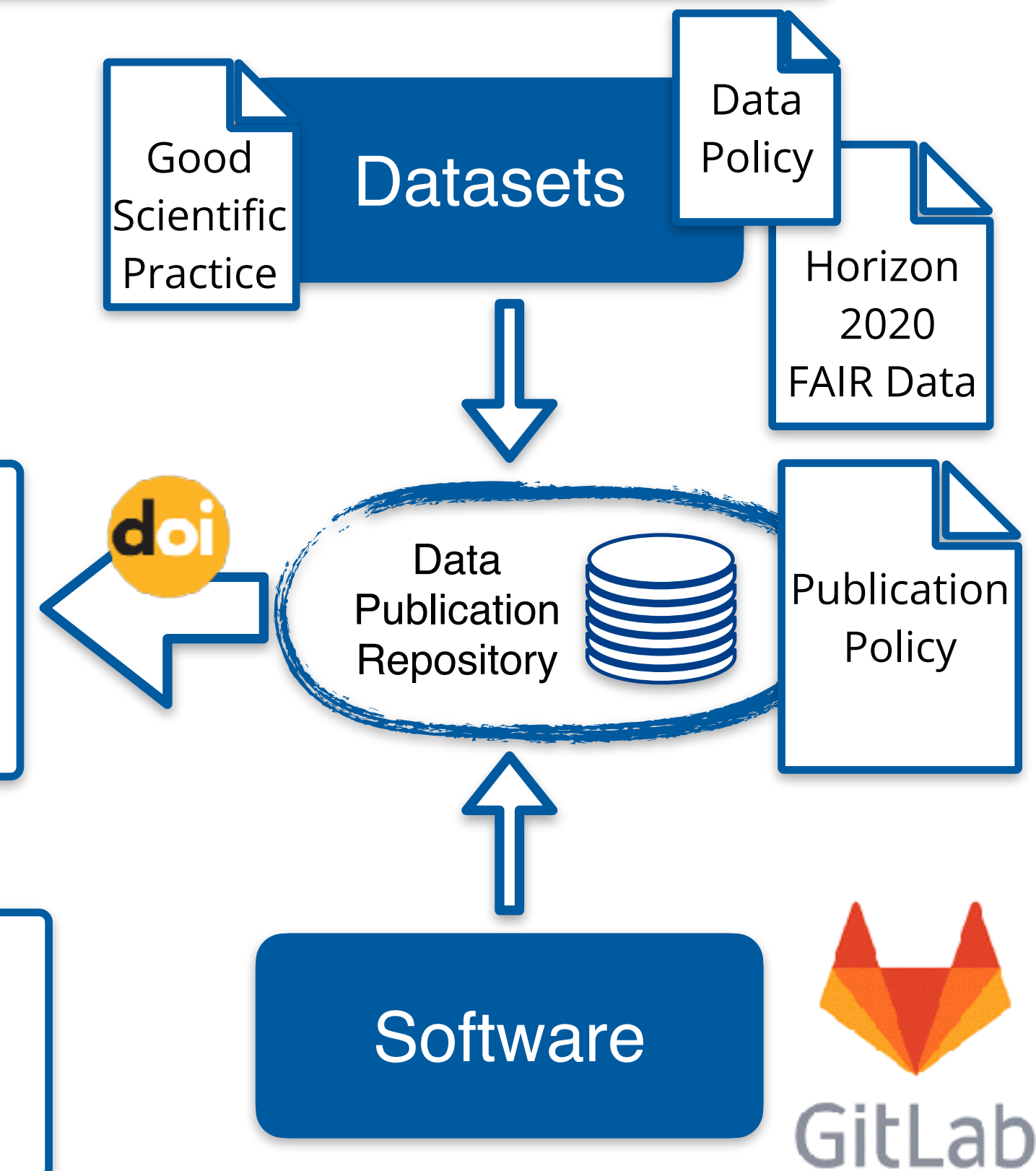
Example: HELIPORT Supported Experiment at The TELBE Beamline

— HELIPORT provides access to our services and guides **external/visiting scientists** through the entire experiment.
— This requires the experiment to be mapped to systems in the HZDR infrastructure.
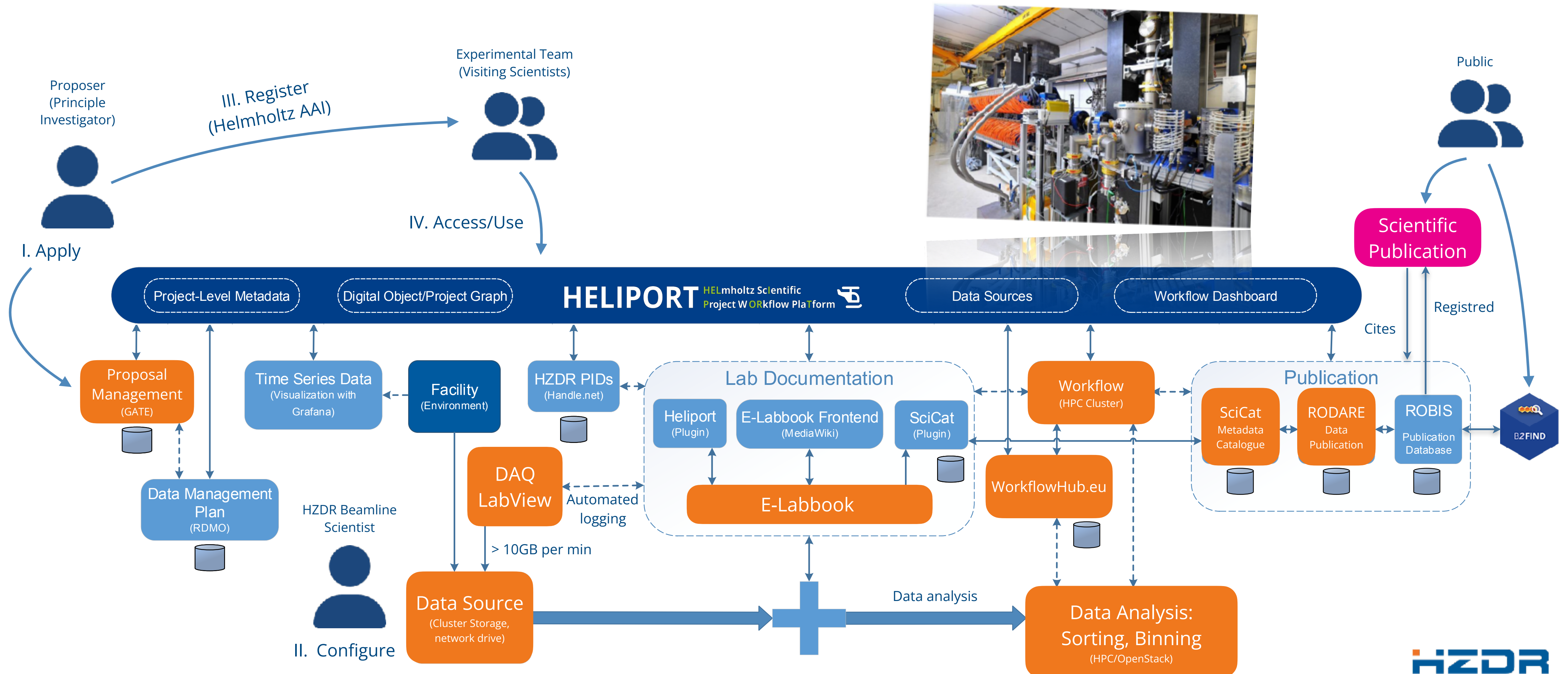
I. Proposal Submission

Automated transfer of project metadata from the proposal system (GATE) into HELIPORT:

— Title, Authors, Description,

— Beamtime schedule,

— Large-scale facility used,

— Scientific method (PaNET)
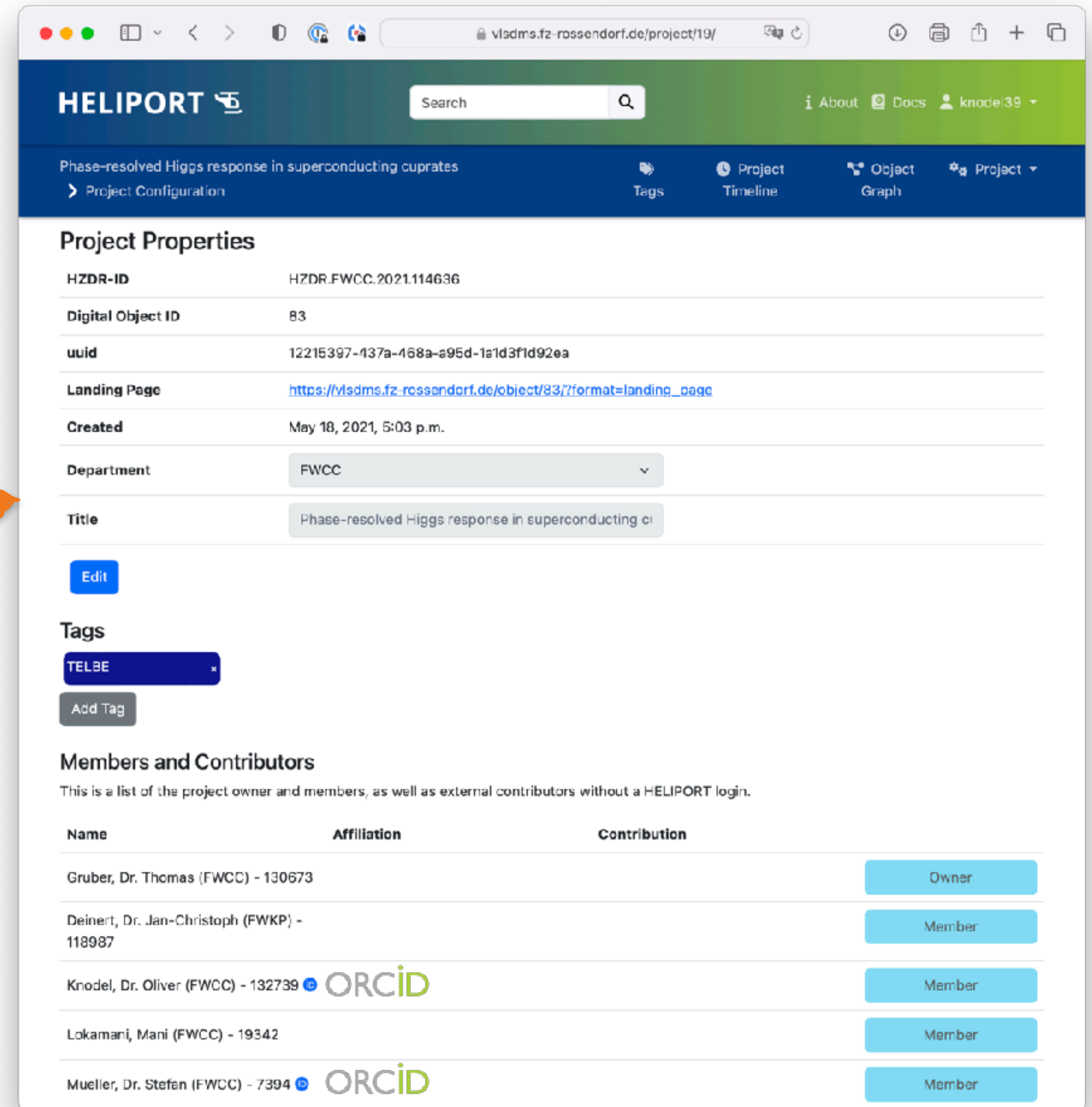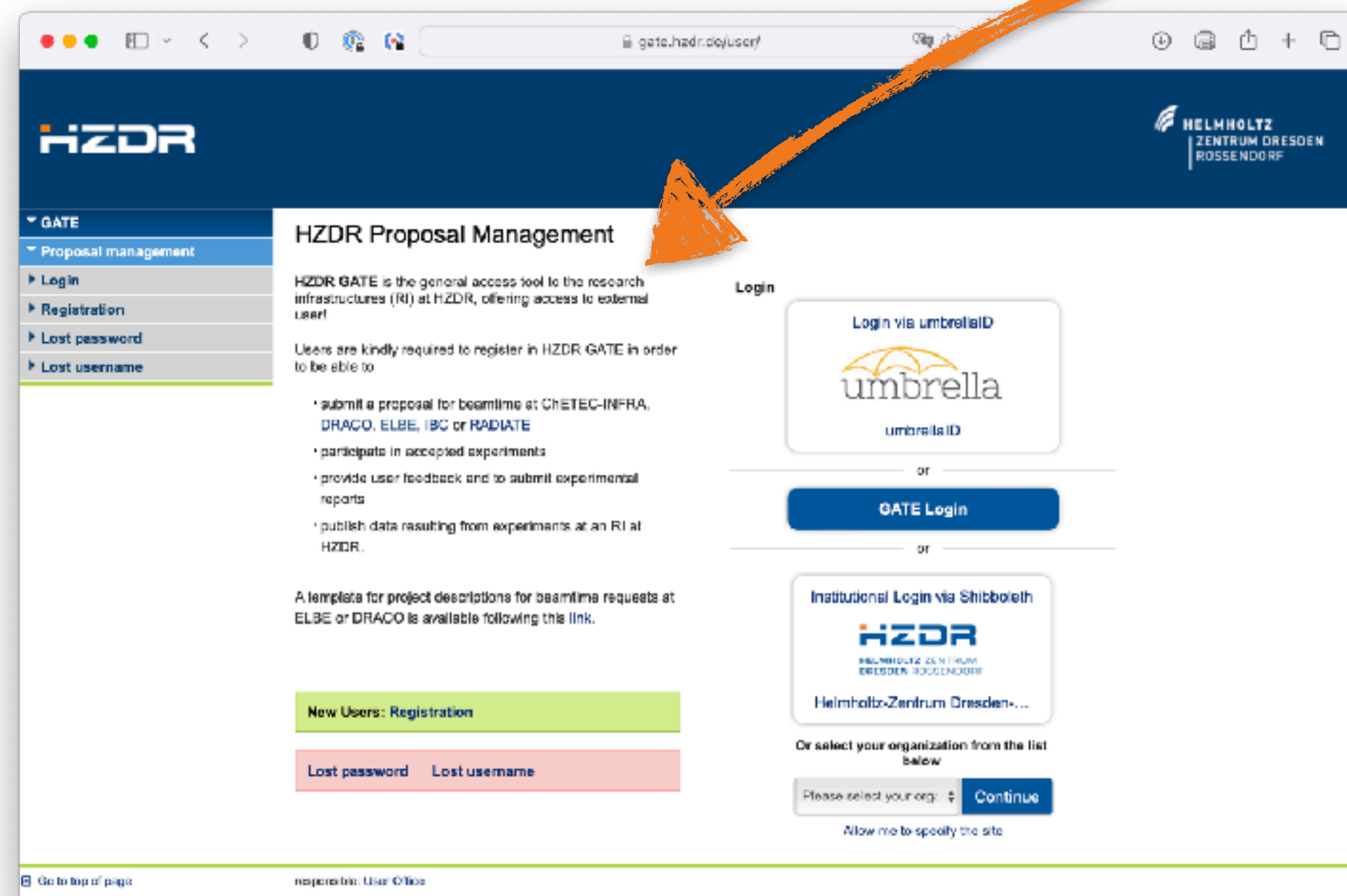
II. Project List and Dashboard

— Typically, a beam line scientist is the owner of a HELIPORT project and the proposer has the role of the manager and can add additional project members.

— Tags and sub-projects including inheritance are possible in the project list.

III. Resources: Documentation and Repositories

The documentation section is typically used to refer to all internal and external systems or services used:

— E-Labbook (Mediawiki),

— GitLab, Github, Workflowhub, …
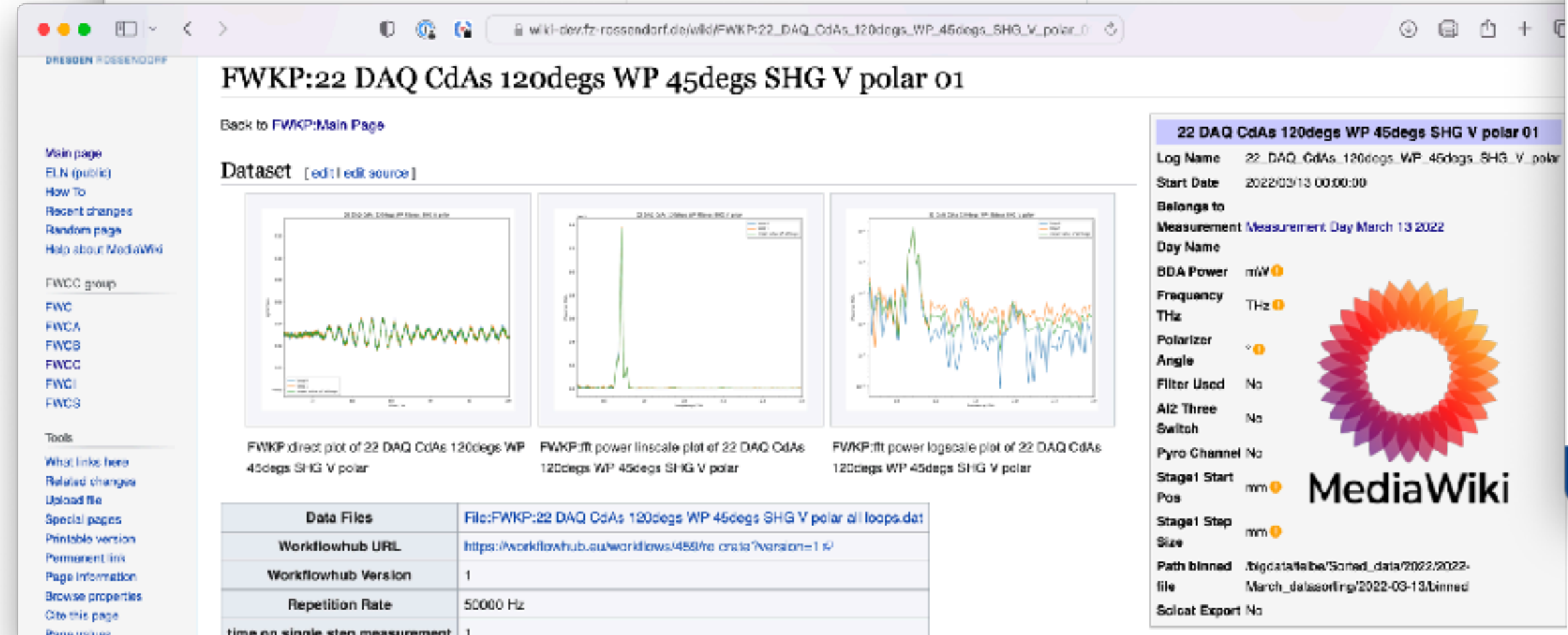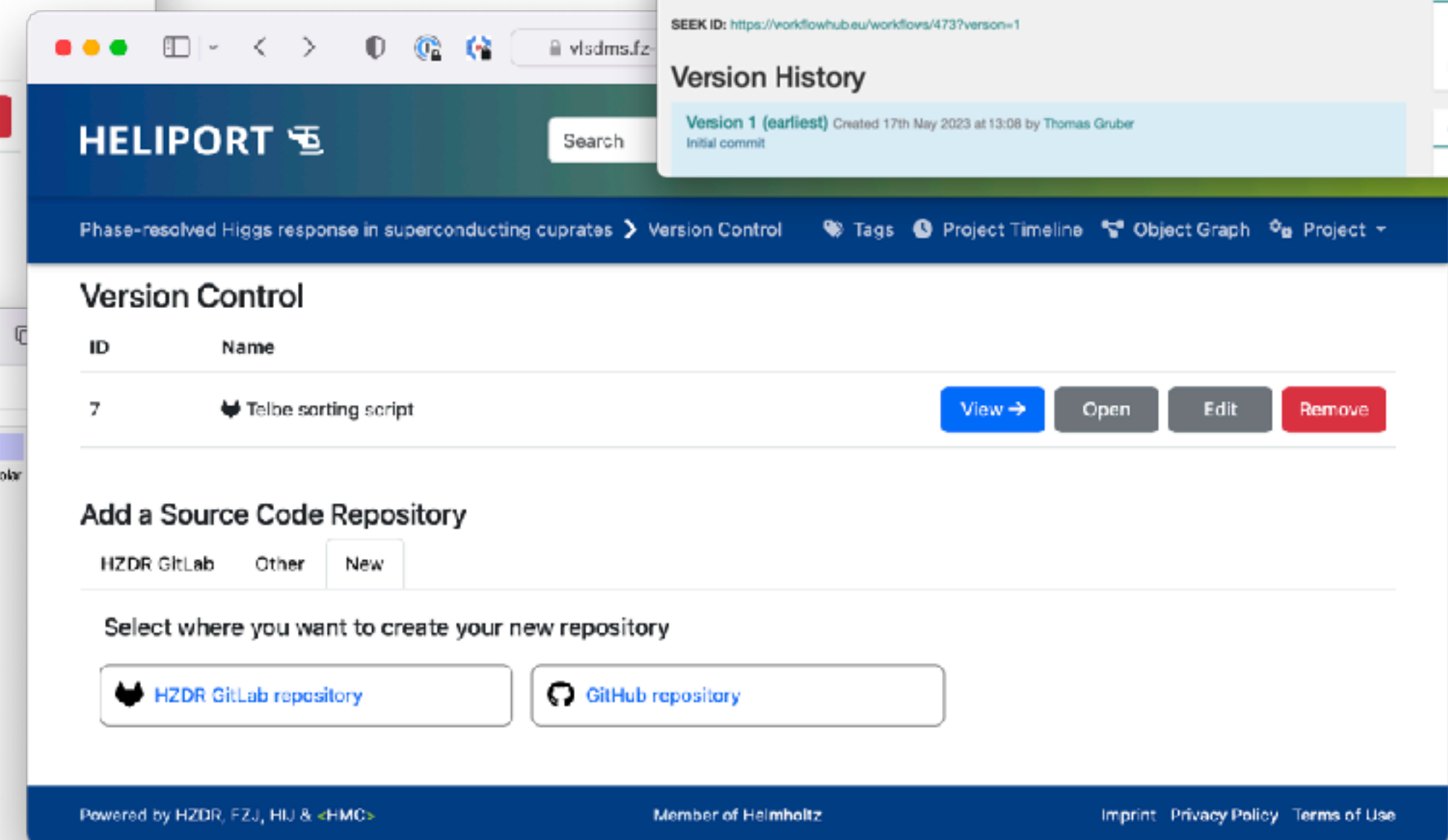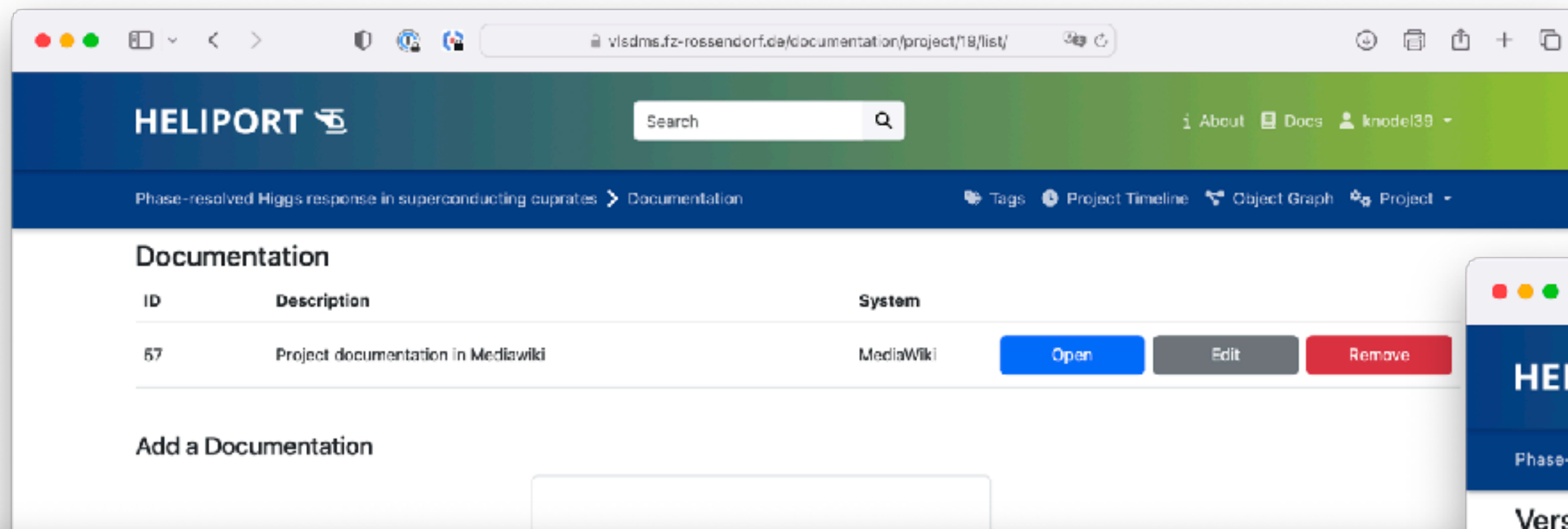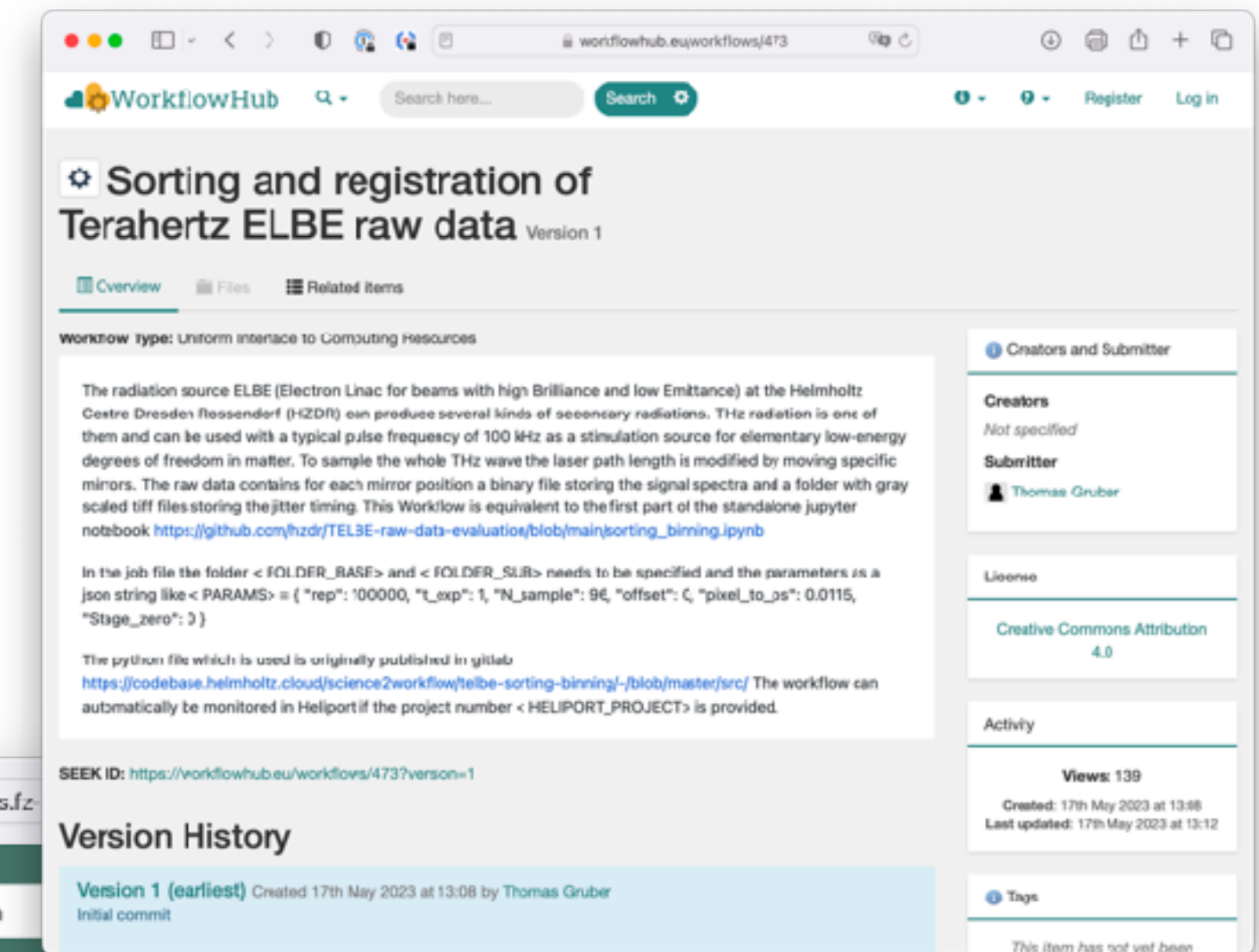
IV. Detector Control and Workflows

— The **HELIPORT REST-API** enables the transfer of metadata between HELIPORT and external systems (e.g. detector control in LabView).

— The integrated job/workflow submission system collects metadata with **provenance information** provided by HELIPORT.

— Workflows (on our HPC cluster) can be accessed by any project member using the HELIPORT web frontend.
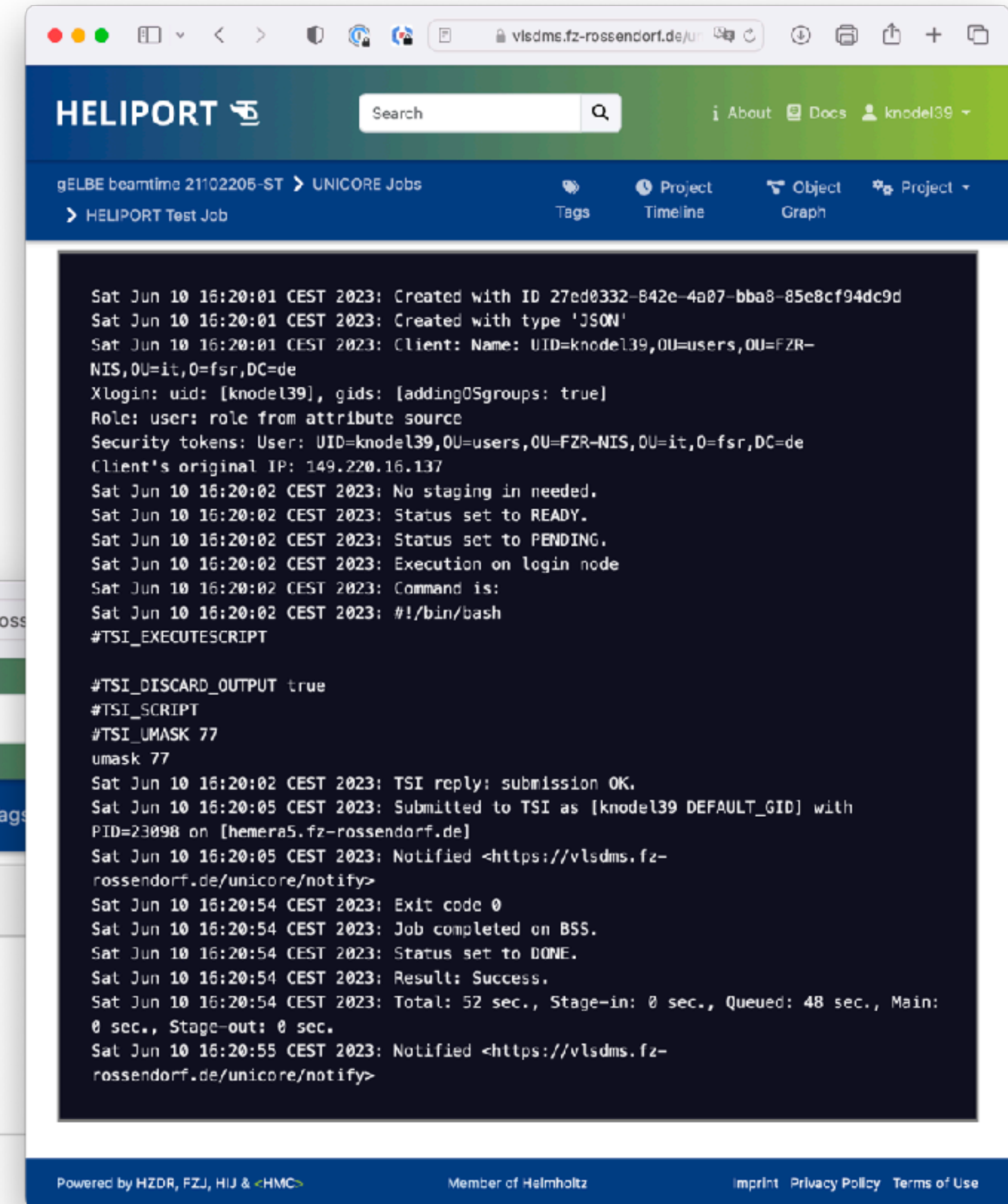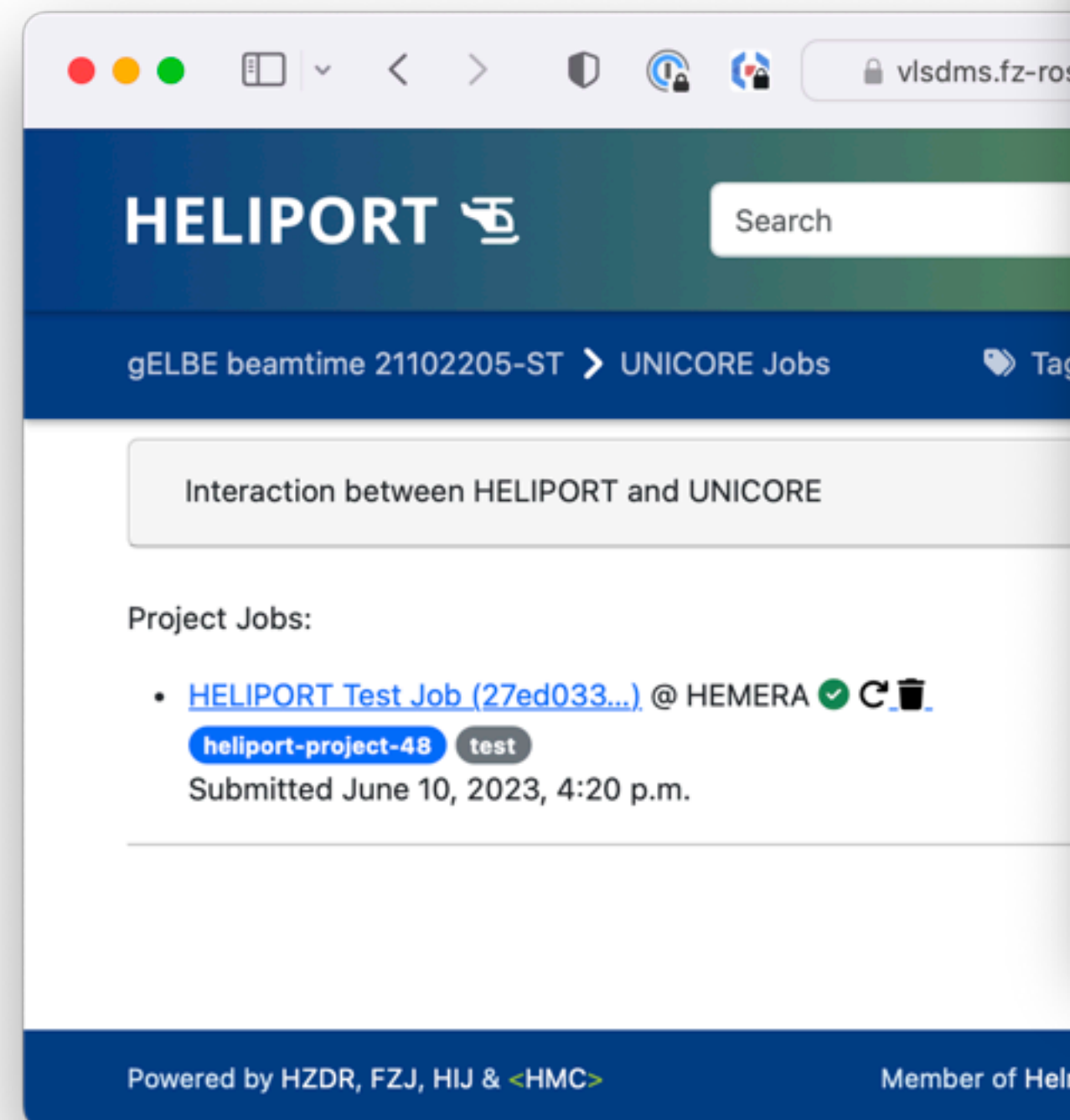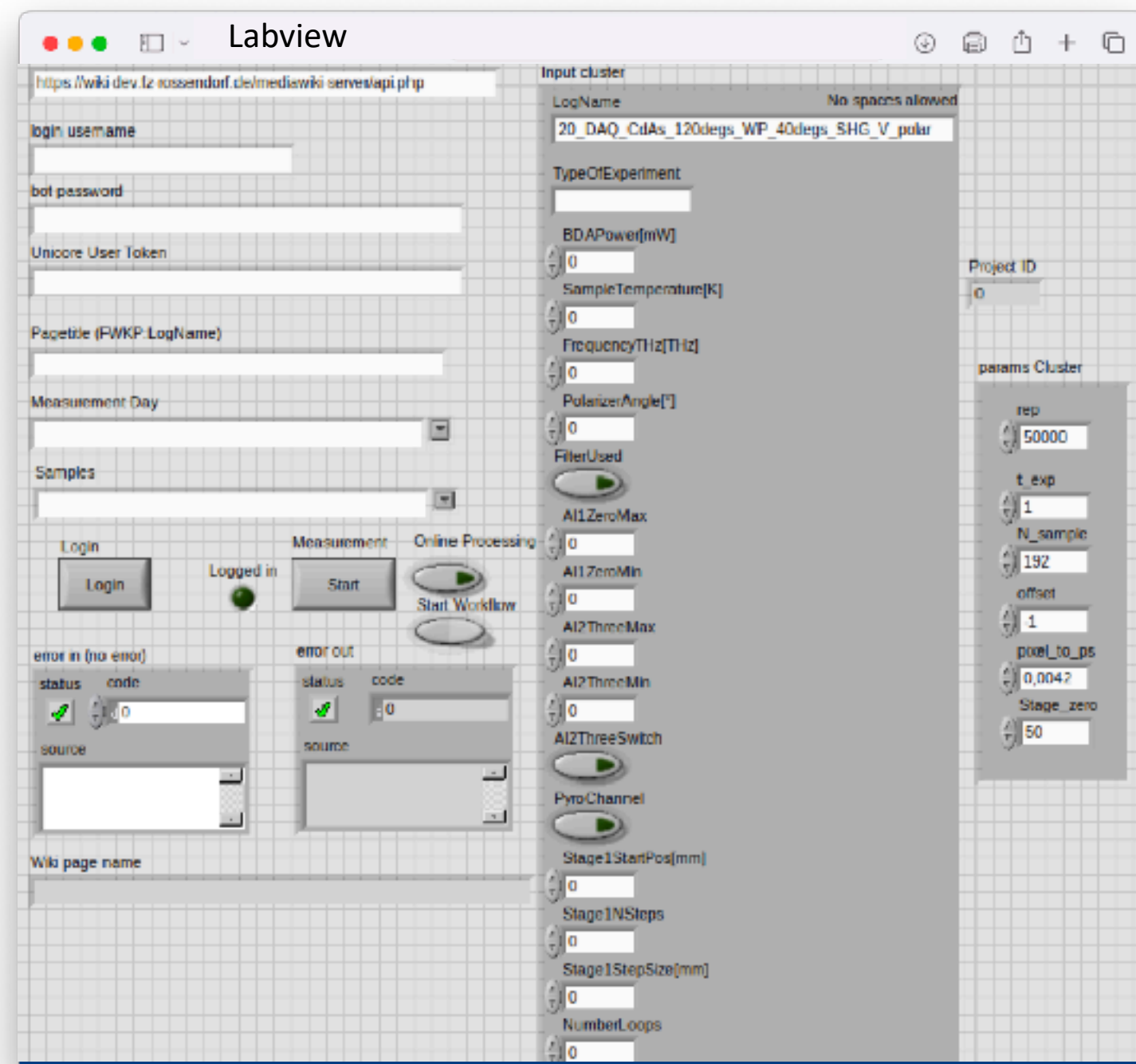
V. Data Sources

— Folders and files in our internal filesystems can be registered in HELIPORT as **data source**.

— Each **project member** has access to the files and folders using HELIPORT.

— The provenance of the data sets generated from an experiment is entirely comprehensible.
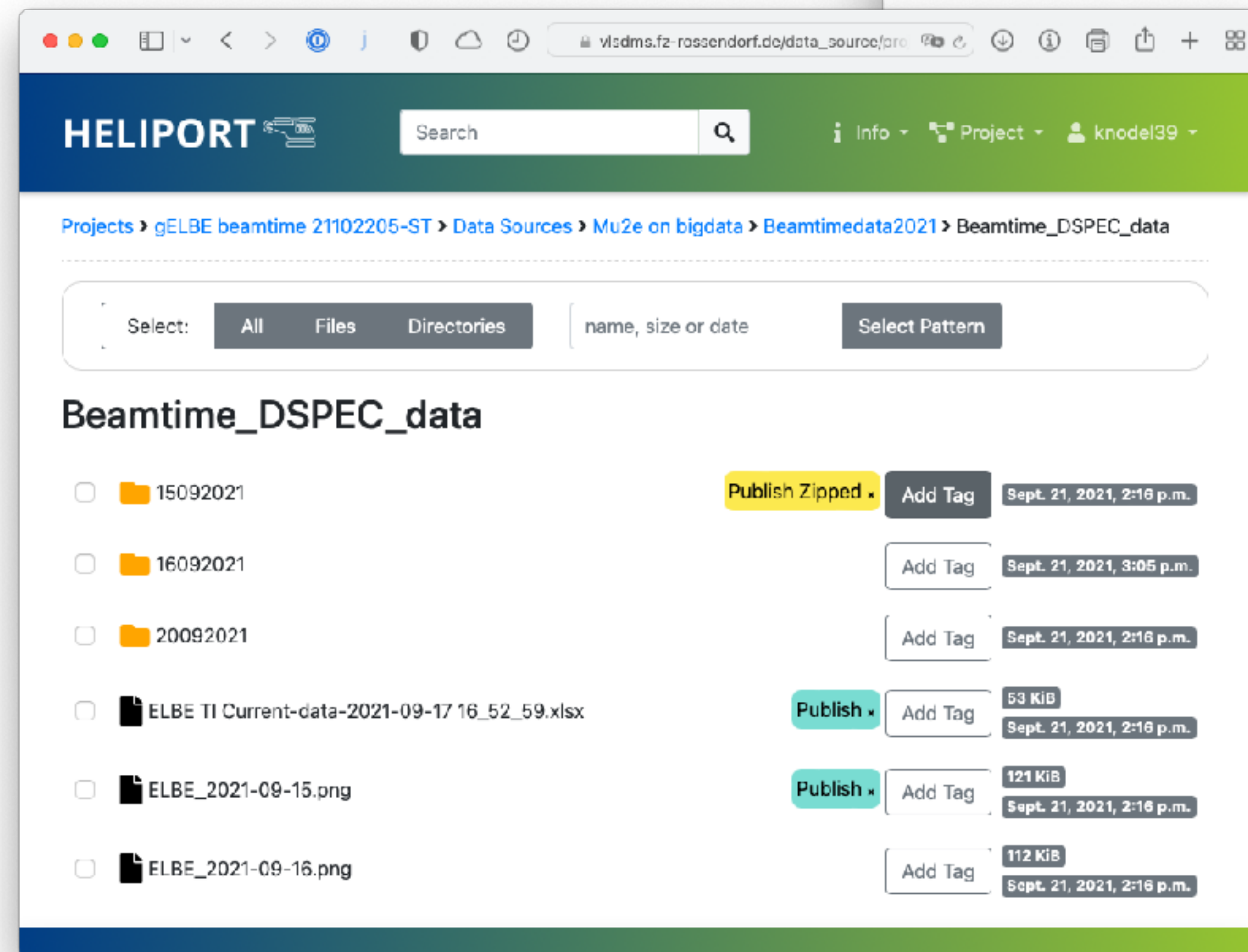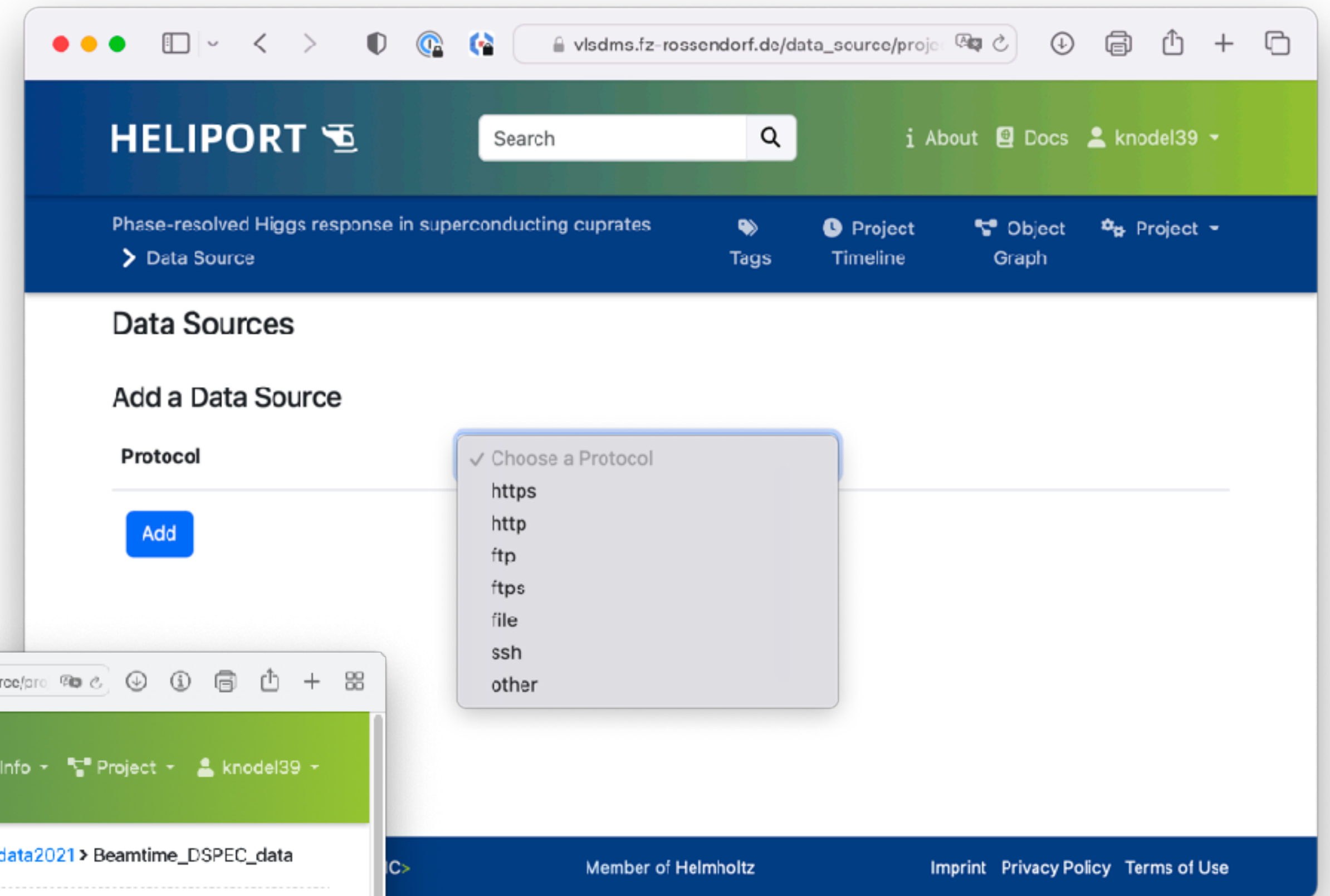
VI. Integration in an Overall Data Publication Workflow

Automated data publication with:

— Metadata from Proposal System,

— Files and folders registered and selected in HELIPORT.

VII. Search & Reuse: Metadata Catalogue SciCat

— With our data publication repository RODARE we provide a platform for publishing datasets.

— In RODARE, bibliographic metadata is based on **DataCite**.

— For additional **scientific metadata** we use SciCat and reference the specific datasets in RODARE or HELIPORT.

Metadata Catalogue SciCat and Data Repository RODARE (Draft)

Curated Metadata Source

Public Metadata Catalogue

Data Access

ExperimentLogging app (ExL)
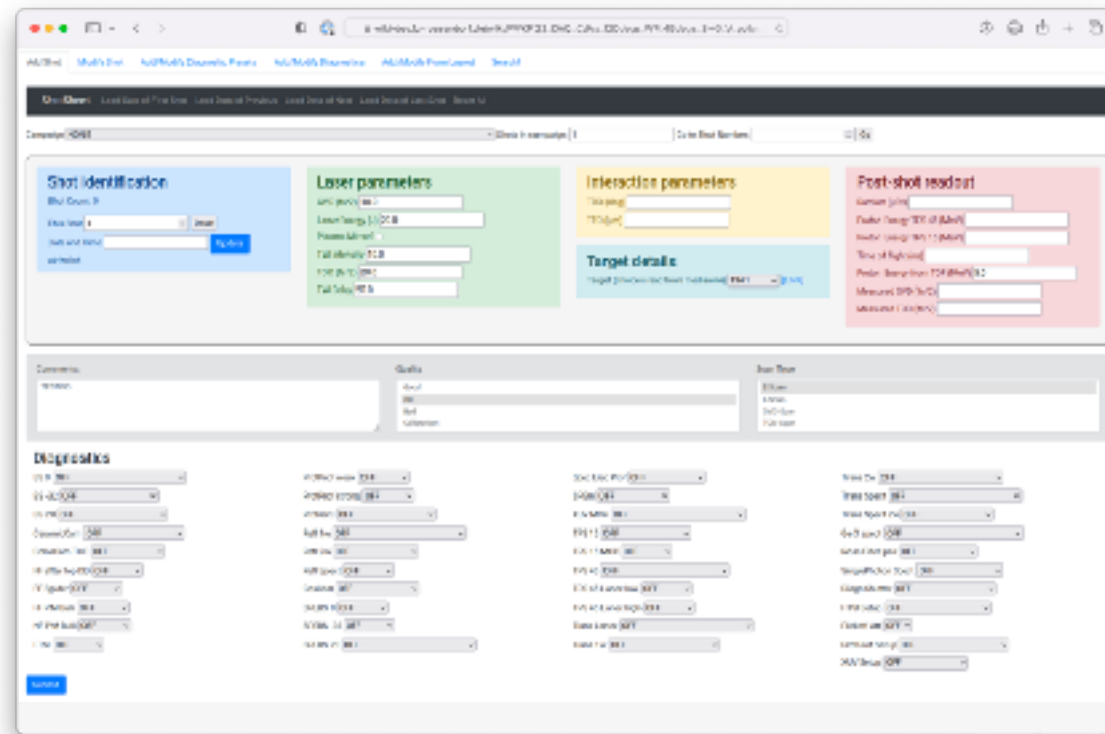
E-Logbook
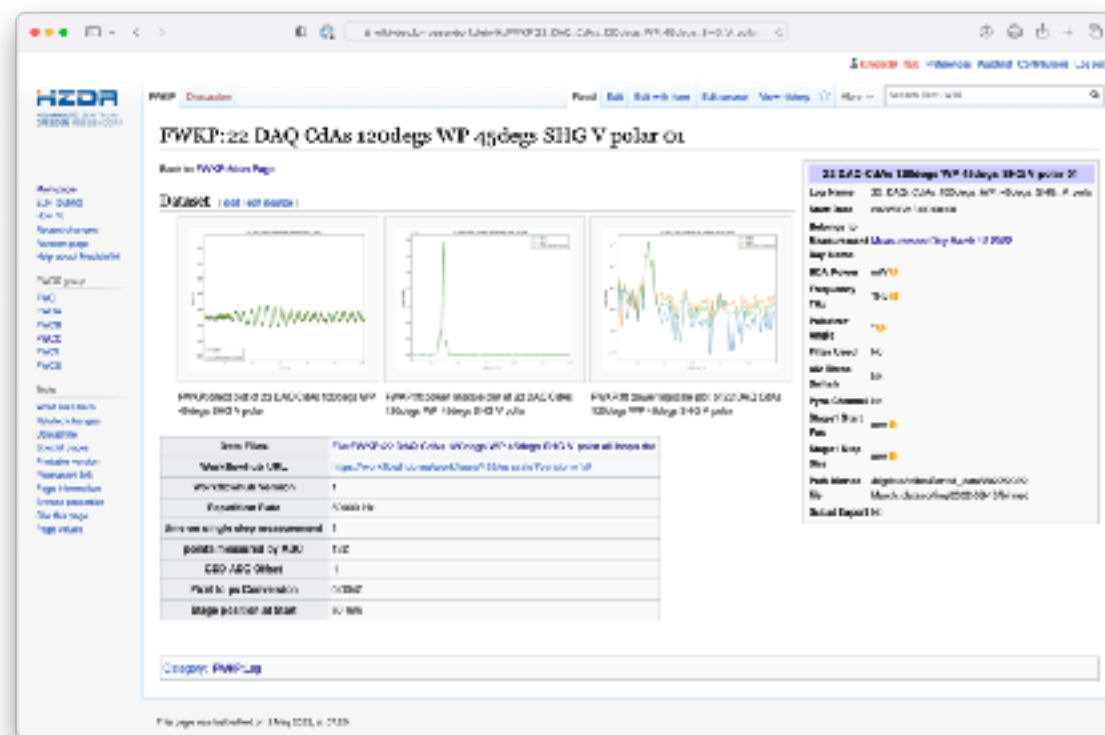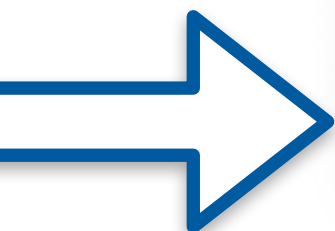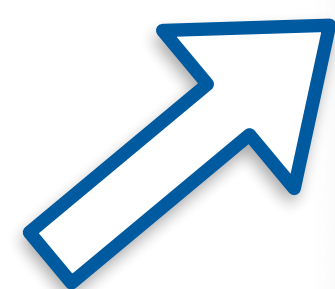
SciCat
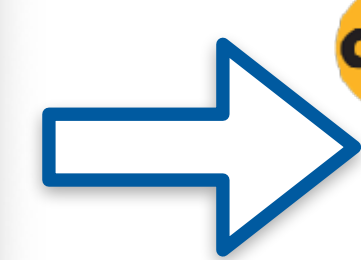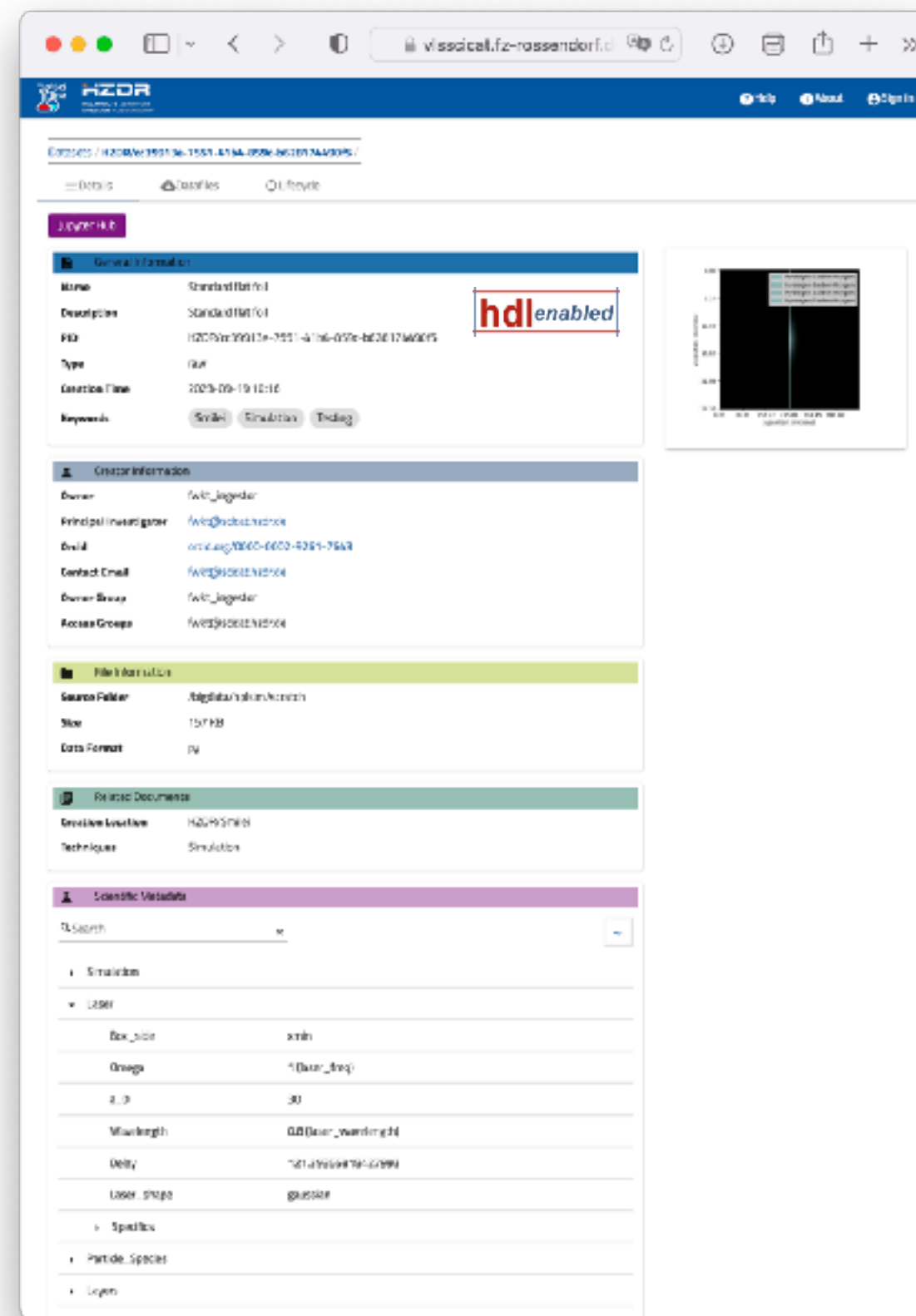
RODARE

HELIPORT
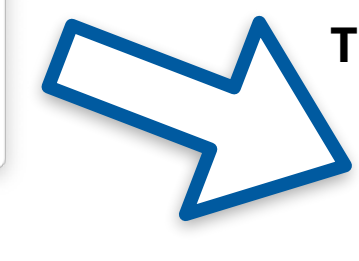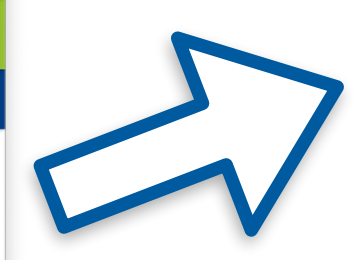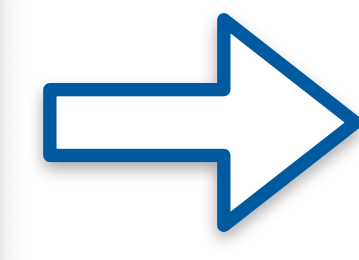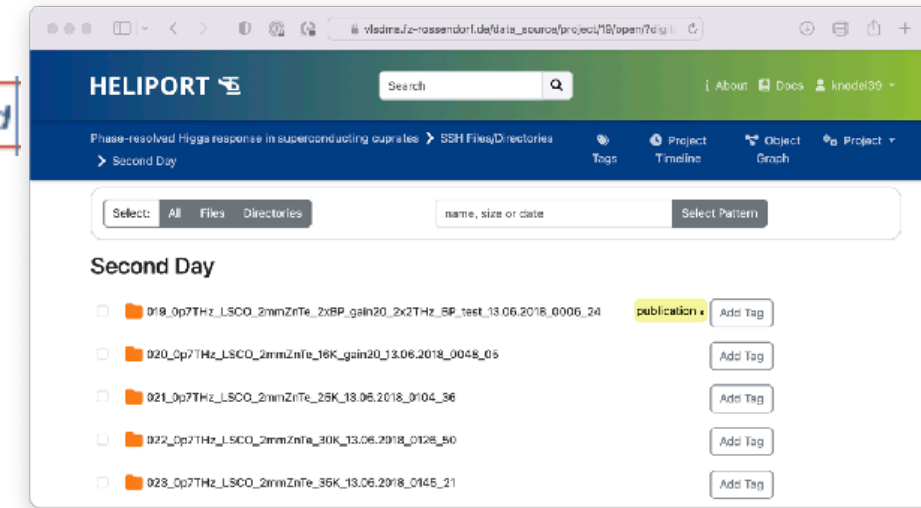
Dataset
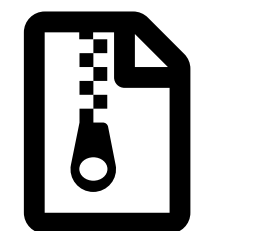
Filesystem

Tape Archive

Metadata from Experiment/Simulation

Direct API Call

HZDR

Conclusions

— Close and automated interaction of services and systems in our digital research landscape is essential to provide **FAIR** and **comprehensible** research projects.

— HELIPORT describes and **collects metadata** from services and systems involved in experiments.

— APIs and workflows are used to transfer metadata between our services and systems.
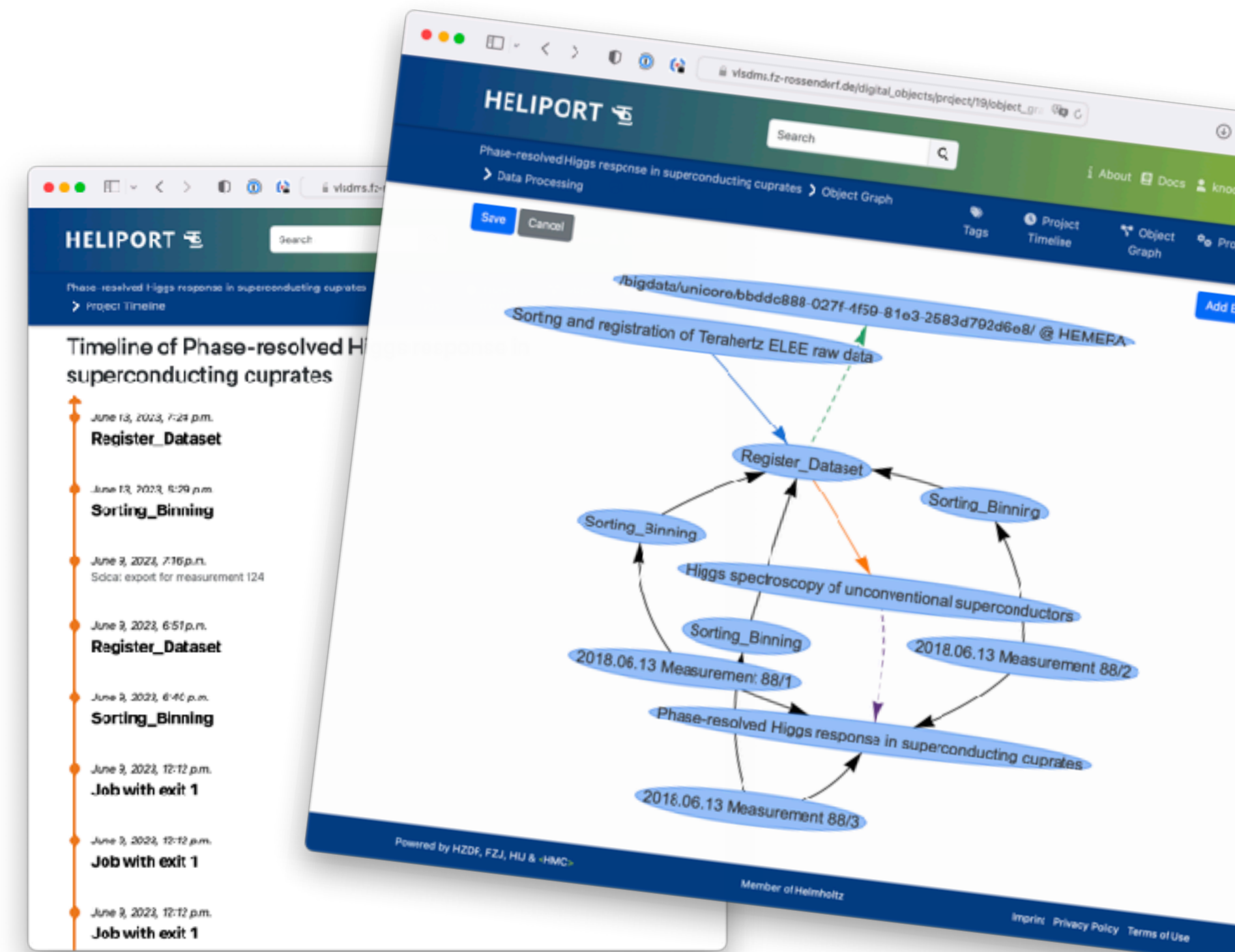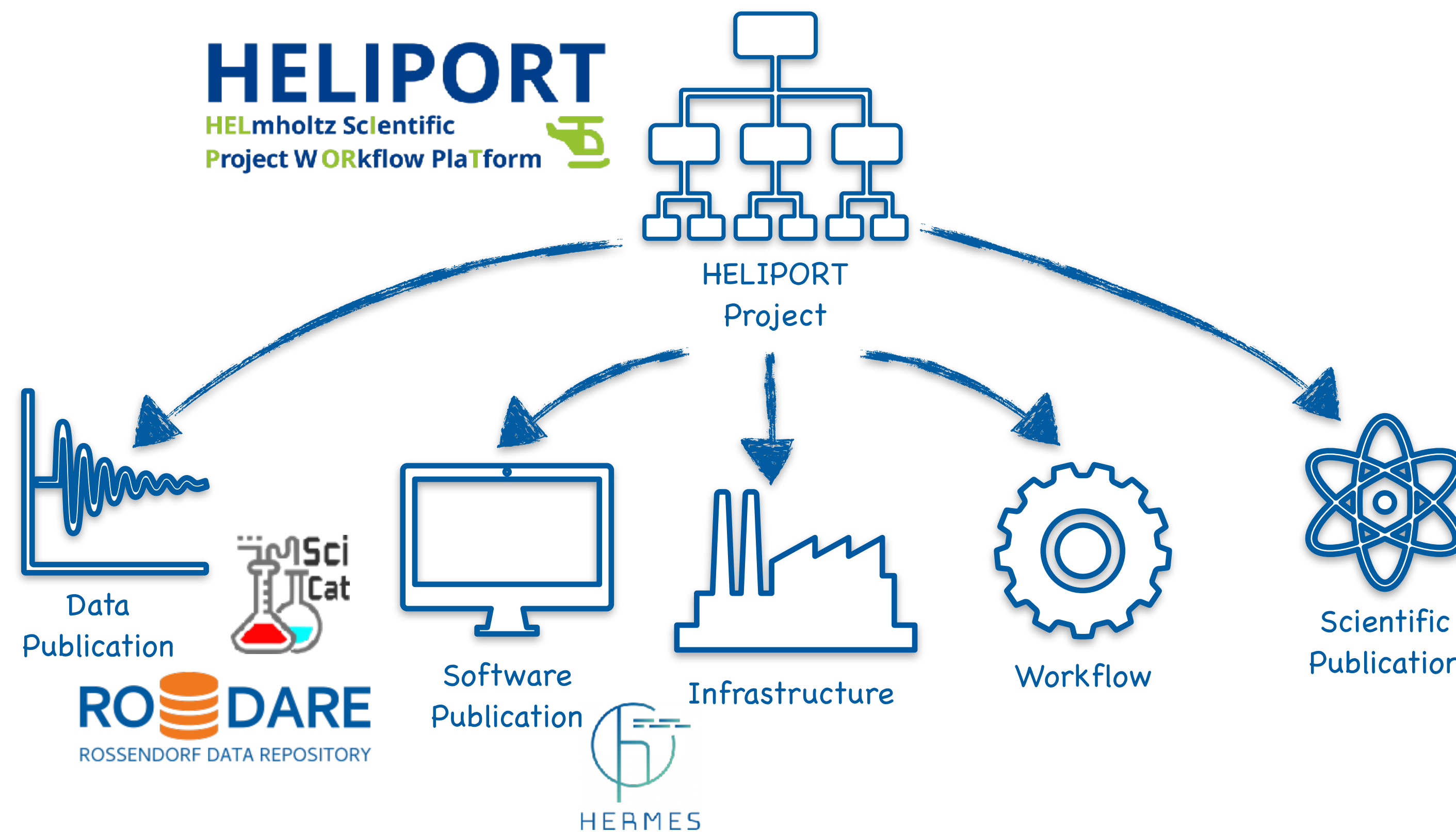
Resources

Website: heliport.hzdr.de

Repository: codebase.helmholtz.cloud/heliport

API Doc: heliport.hzdr.de/redoc/

Appendix

Example: Data Management View of the TELBE Experiment

— Terahertz facility at the ELBE center for High-Power Radiation Sources.

— In the future HELIPORT guides (external) scientists through the complete experiment.

— Submission of data analysis Jobs from LabView to UNICORE with visualisation in HELIPORT.

Deinert, Jan-Christophi. (2021, November 18). TELBE data analysis workflow and the PaN training platform UX — Booklet of presentations from the PaN EOSC Symposium 2021. Zenodo. https://doi.org/10.5281/zenodo.5636331
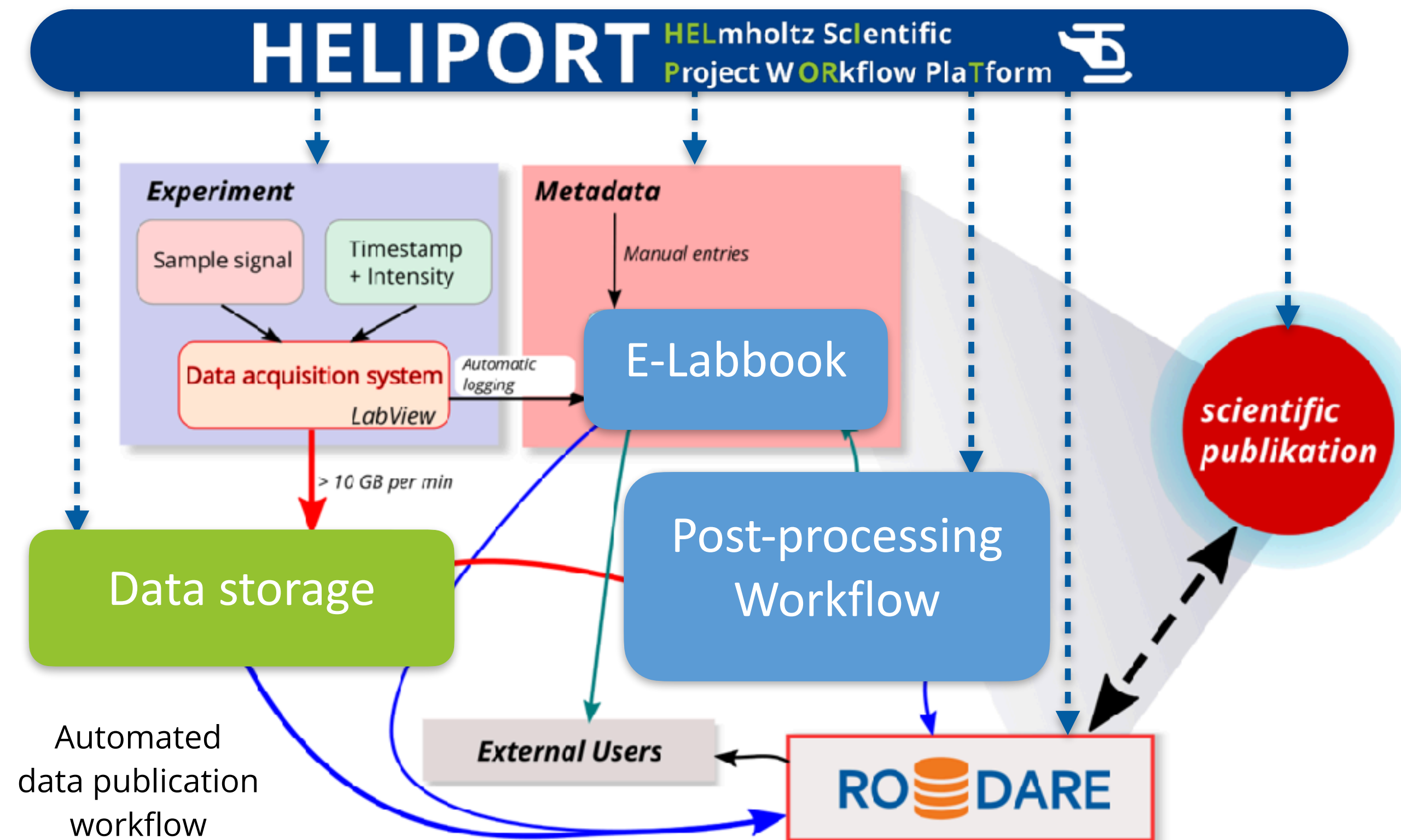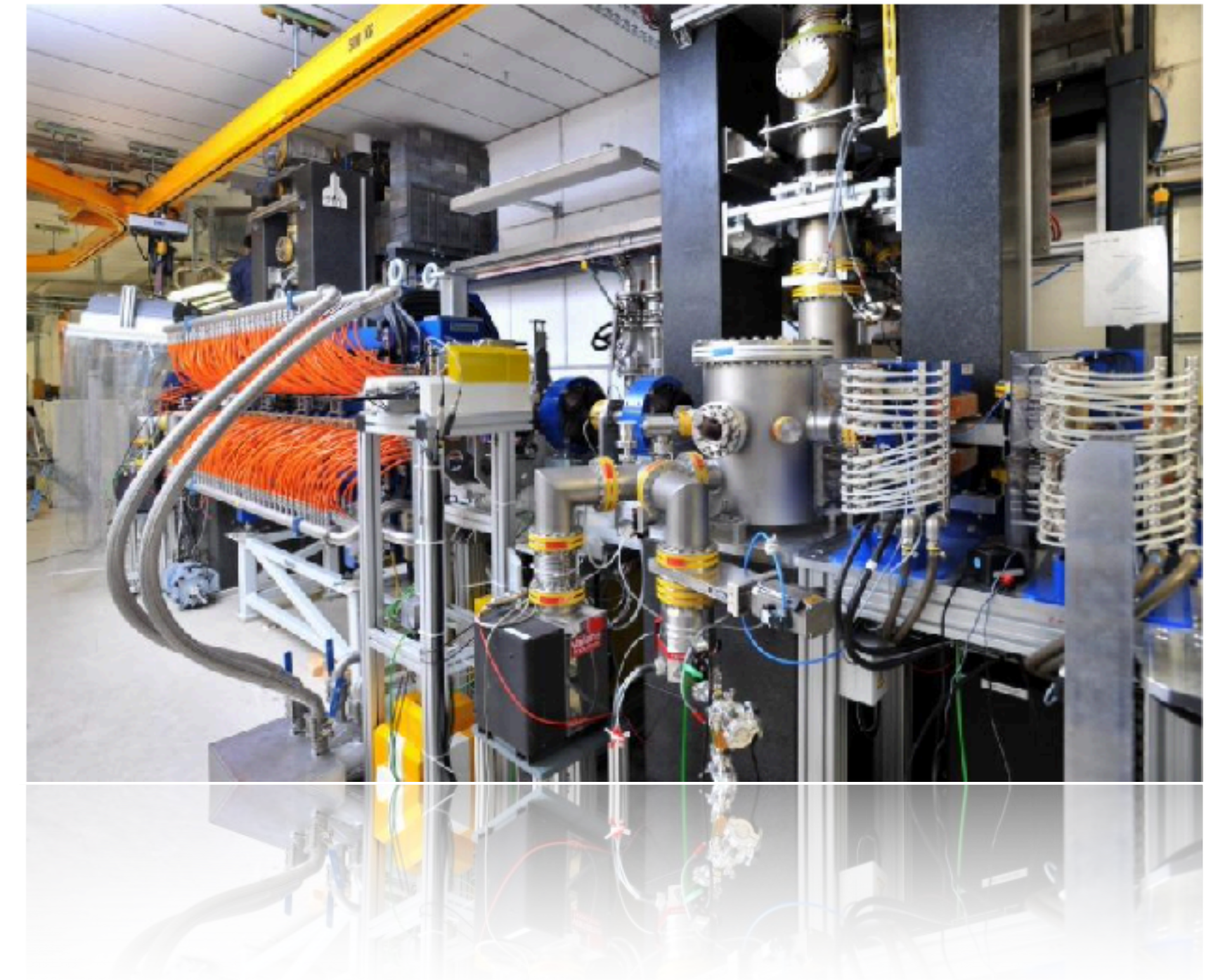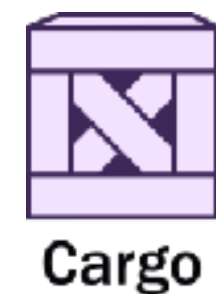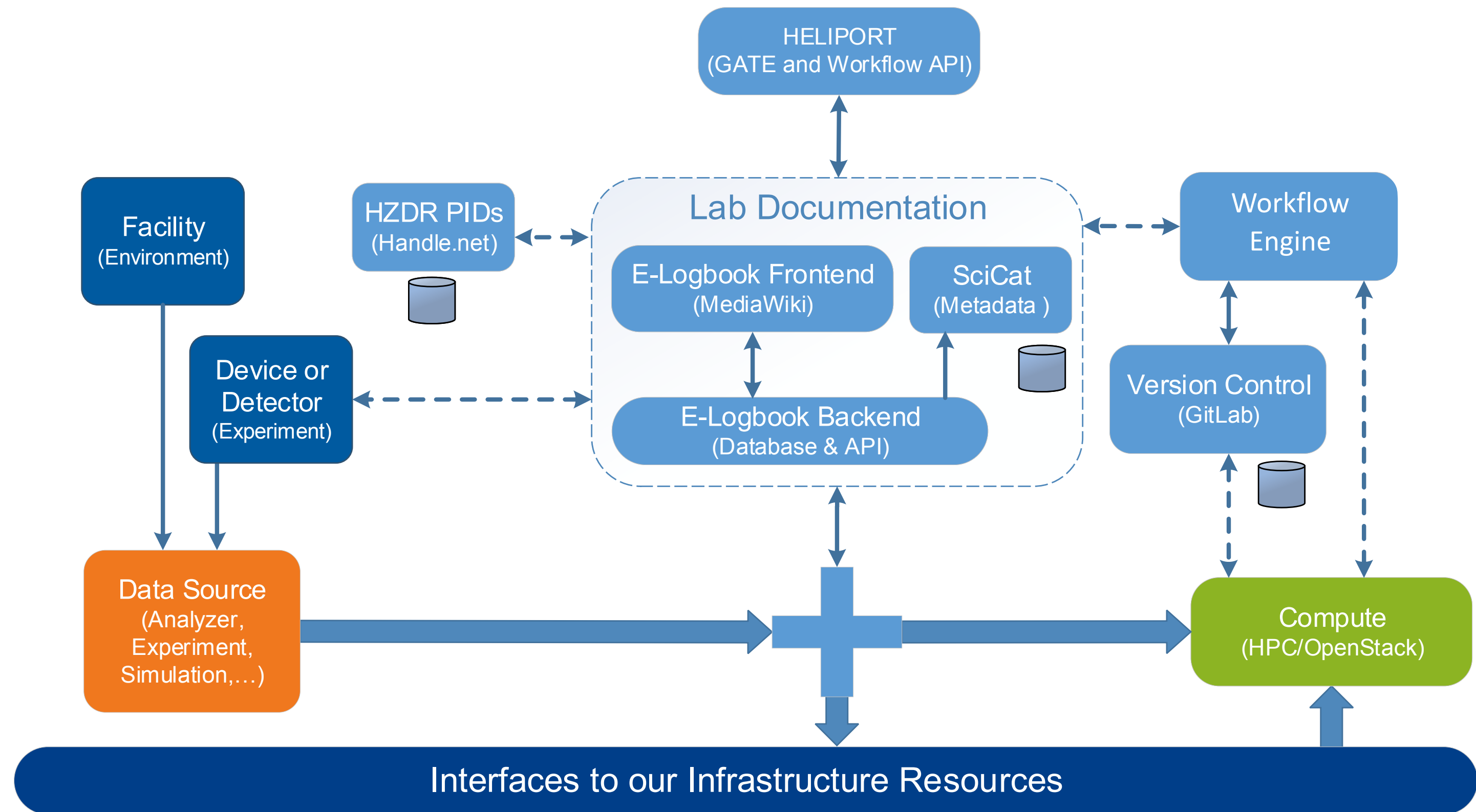
Documentation Ecosystem at HZDR

— Our **Electronic Lab Documentation** is a central database gathering information from various systems:

- Proposal management (GATE),
- Environmental data,
- Devices (e.g. Labview),
- Workflows,
- (Meta)data databases.

— Different frontends are available:

- SciCat (metadata only),
- MediaWiki (structured user-definable views).

— The system is build on-top:

Relations Between Digital Objects and

— Relations between digital objects are visualized to provide a top-level view on the project with dependencies.

— The relationships between simulation (surrogate model) and experiment can also be demonstrated.

— The versioning of an experiment is an essential extension, and first approaches via a timeline are being evaluated.
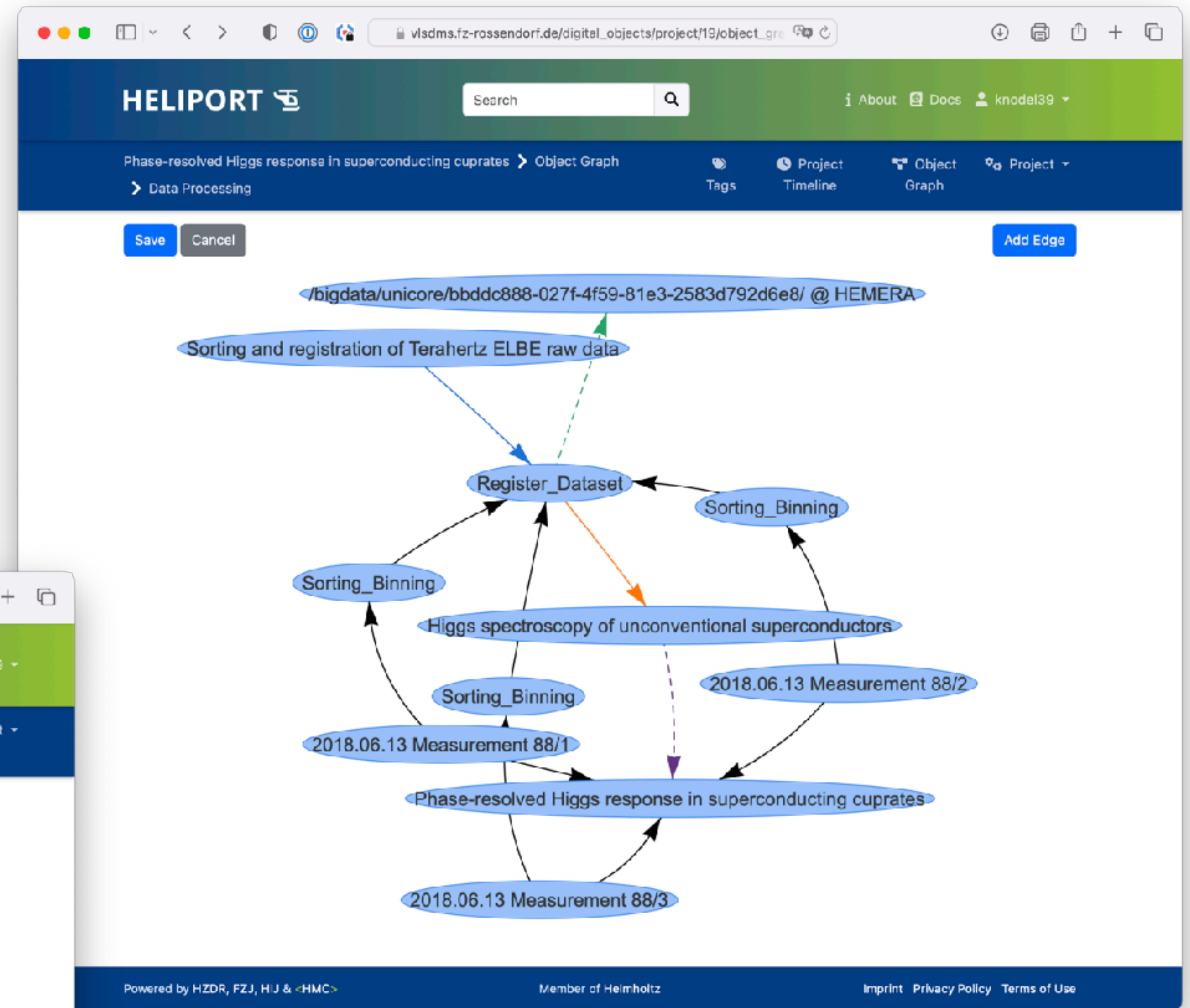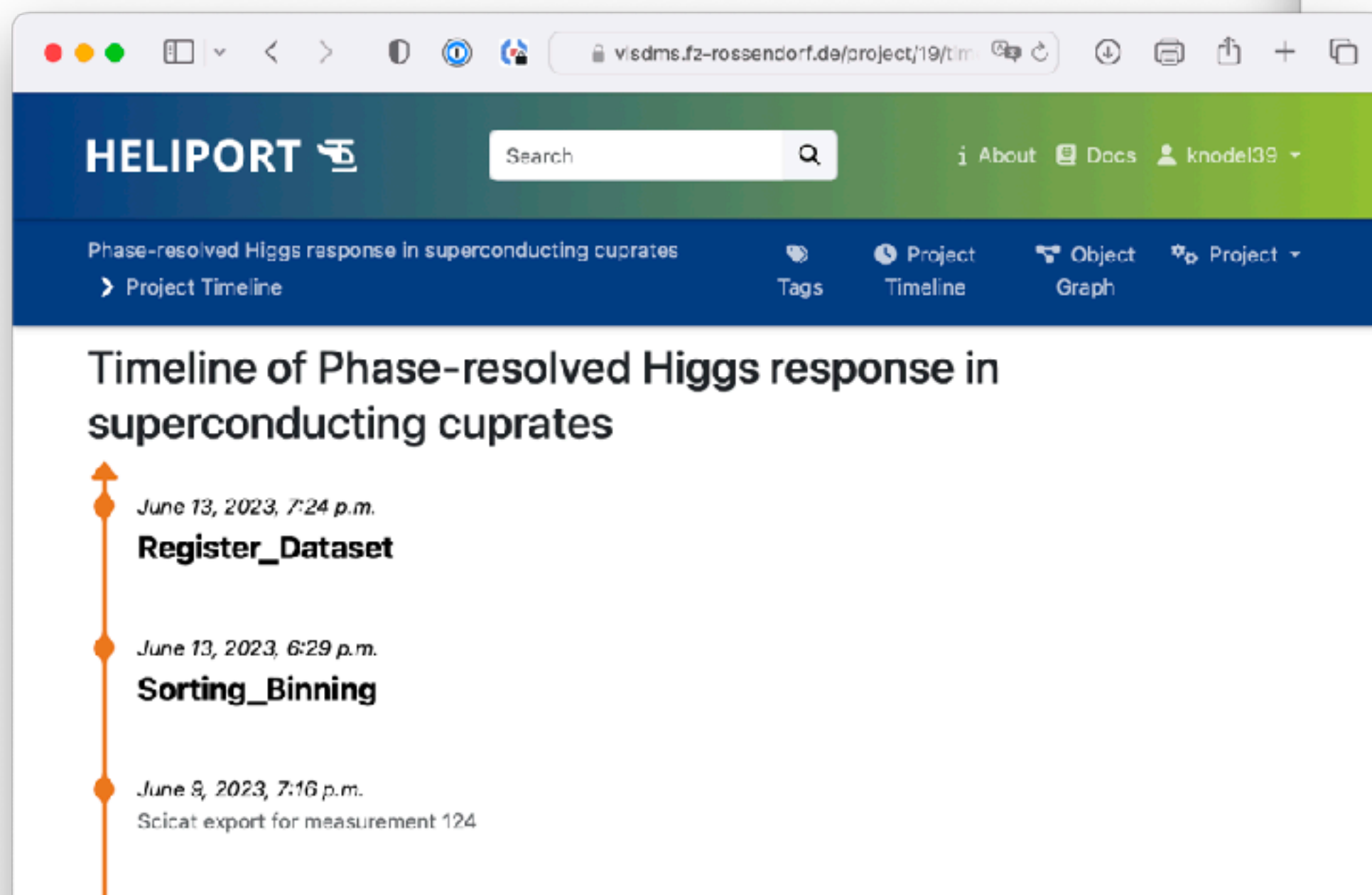
Workflow Architecture (in development)

— HELIPORT offers an infrastructure which permits the integration of various workflow languages and access modes to HPC infrastructures.

— The infrastructure keeps track of and collects the metadata and enables access to all resources involved.

— Next steps:
- Python library sending workflow information directly to HELIPORT,
- Provision of provenance information from Jupyter notebooks,
- Use case: **PIConGPU**