

# 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.

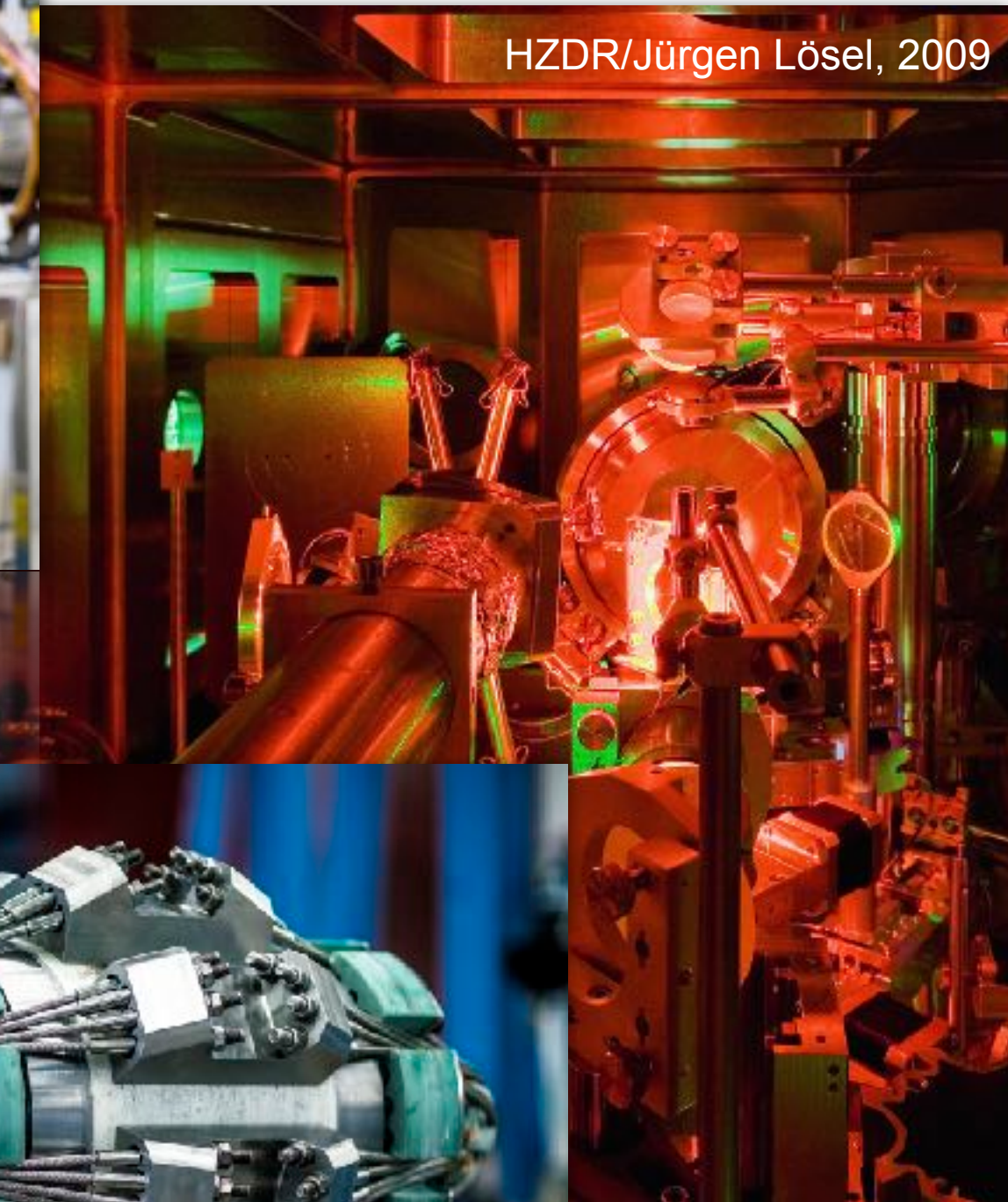
— Positrons, protons, neutrons as well as X-ray and gamma radiation

## Dresden High Magnetic Field Laboratory (HLD)

— Europe's highest pulsed magnetic fields

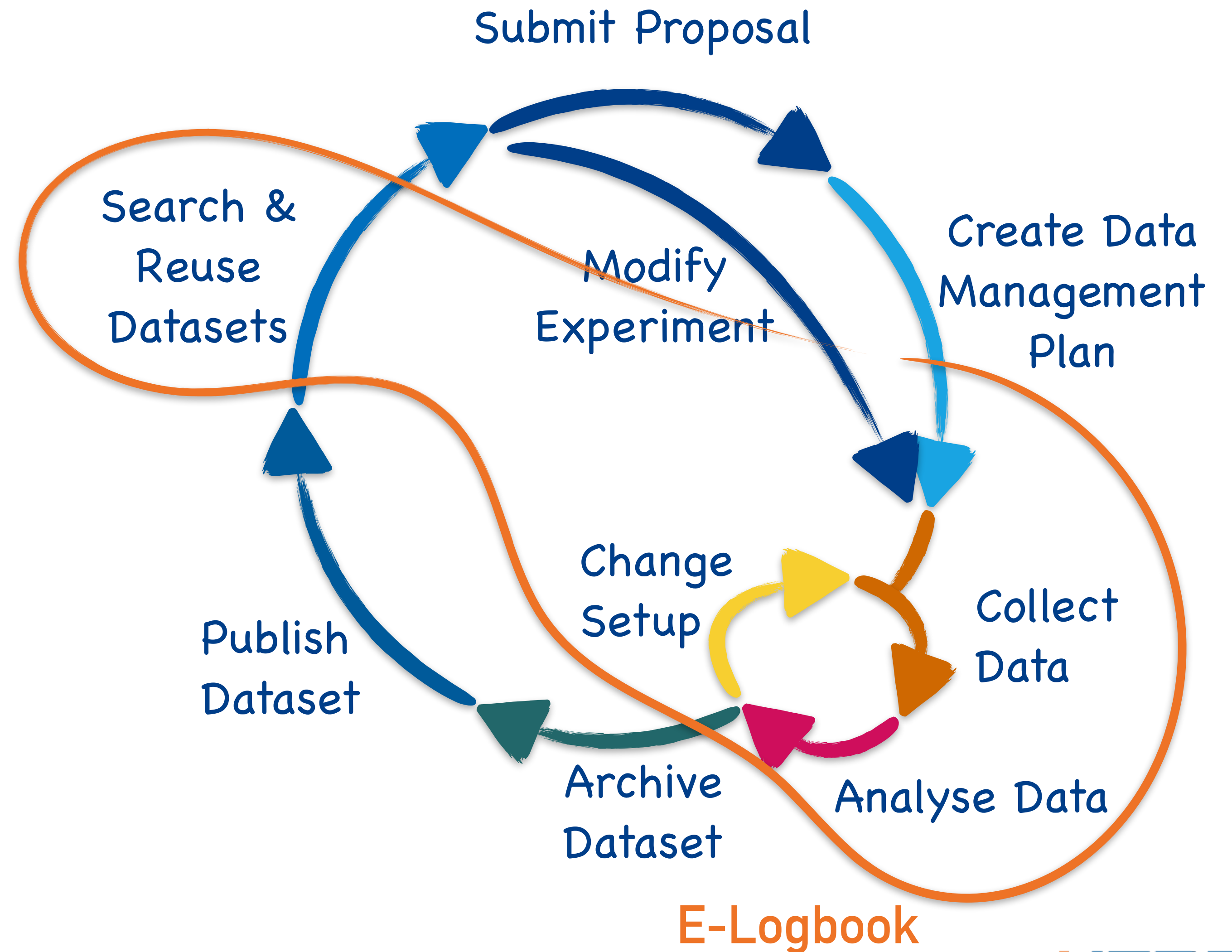
## Ion Beam Center (IBC)

— Nanoscale surface analysis and modification



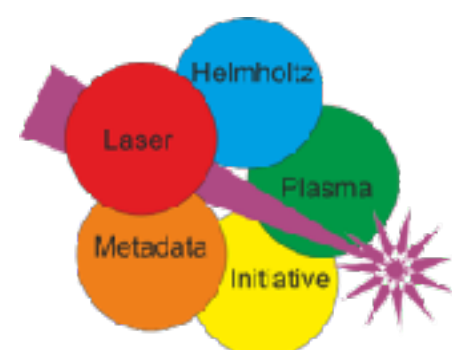
# Our Challenge: An End-to-End Digital Data Lifecycle

- 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.

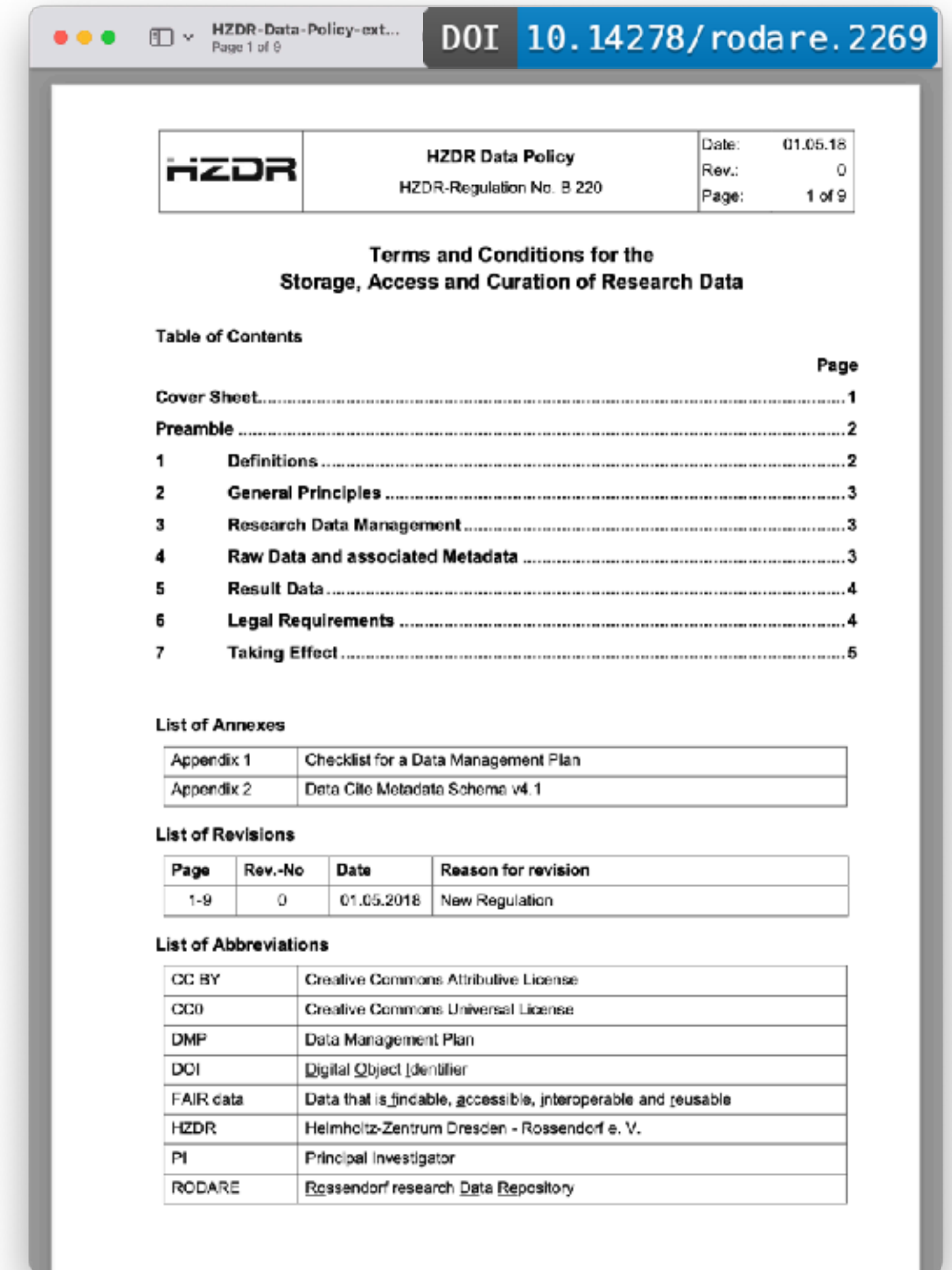


# The Foundation for our Digital Data Lifecycle: HZDR Data Policy

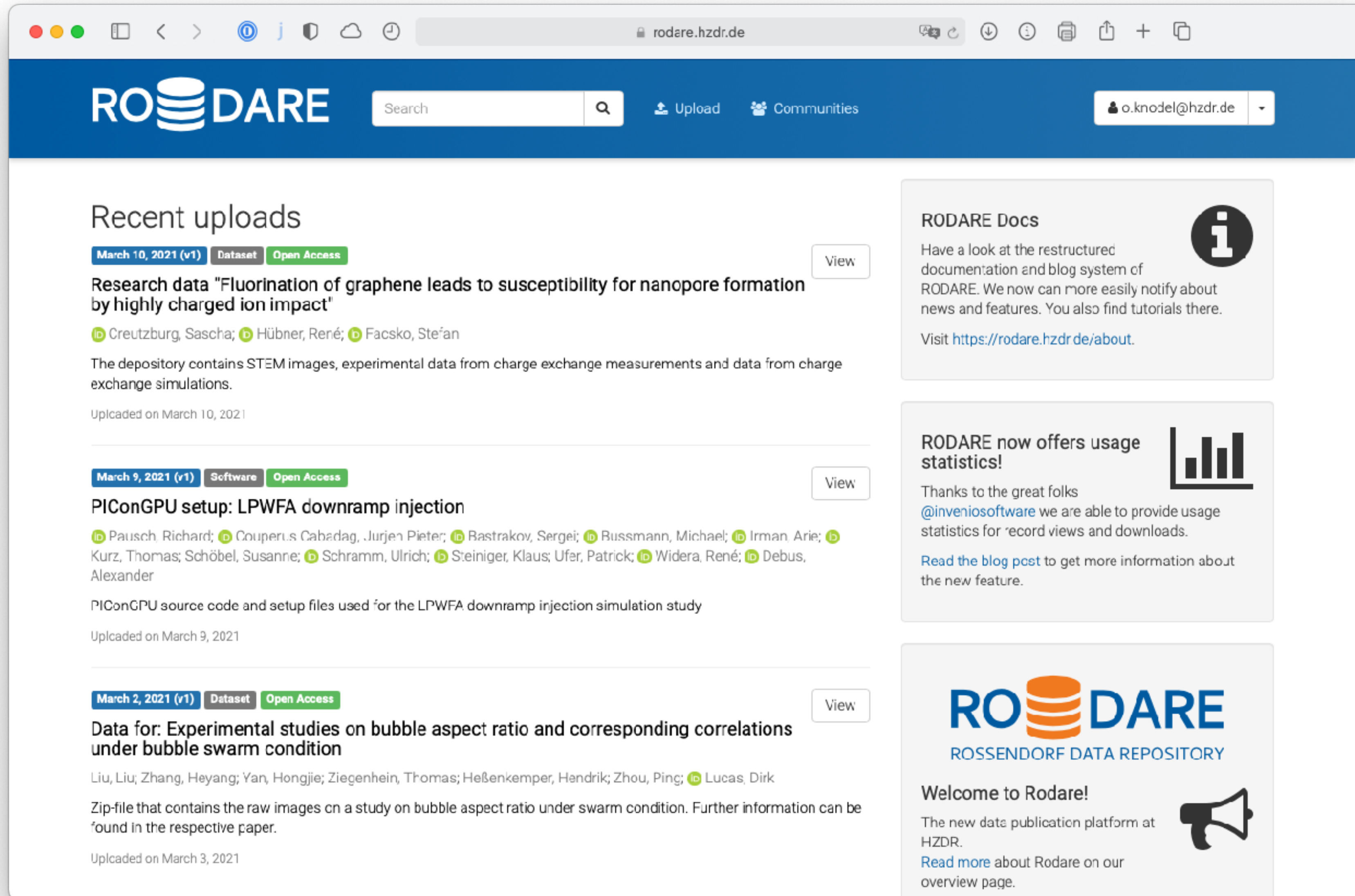
- The HZDR has a data policy since **May 2018**
- Reasons for the development:
  - Legal framework for data management and publication,
  - Establishment and legitimisation of coordinated research data management at HZDR,
- Foundation for the development of tools and services to support our scientists:



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



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



The screenshot shows the RODARE website interface. At the top, there is a navigation bar with the RODARE logo, a search bar, and links for 'Upload' and 'Communities'. The main content area is divided into two columns. The left column features 'Recent uploads' with three entries, each including a date, category, and 'Open Access' status, followed by a title, author list, and a brief description. The right column contains three informational boxes: 'RODARE Docs' with an information icon, 'RODARE now offers usage statistics!' with a bar chart icon, and a 'Welcome to Rodare!' message with a megaphone icon. The bottom right corner of the screenshot shows the RODARE logo and the text 'ROSSENDORF DATA REPOSITORY'.

Powered by:



Registered in:



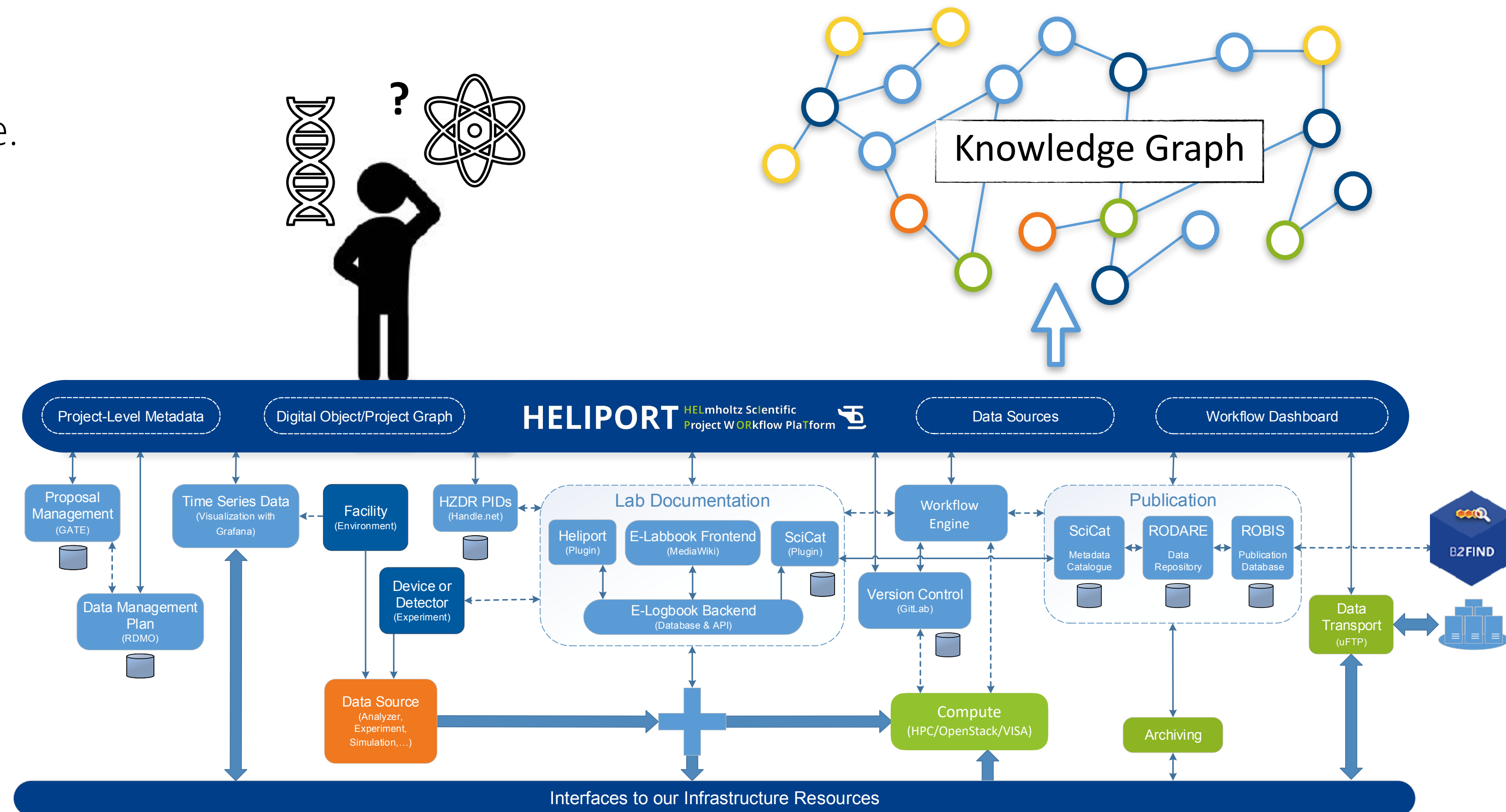
<http://doi.org/10.17616/R3BR40>

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.



“ 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.

Project Members:



Funded by:



```

{
  "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 09:14:34.295524+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://doi.org/10.26434/chemrxiv-2021-04-01-09-14-34.295524+00:00",
      "rdfs:label": "PaN Research Project"
    }
  ],
  "heliport:has_Resources": [
    {
      "heliport:data_source_id": 11,
      "datacite:uri": "http://ddd.org",
      "heliport:use_computer": null,
      "rdfs:label": "ddd",
      "datacite:hasDescription": ""
    }
  ],
  "heliport:has_Automation": [
    {
      "heliport:automation_id": 1,
      "datacite:uri": "https://doi.org/10.26434/chemrxiv-2021-04-01-09-14-34.295524+00:00",
      "rdfs:label": "PaN Research Project"
    }
  ],
  "heliport:has_Results": [
    {
      "heliport:results_id": 1,
      "datacite:uri": "https://doi.org/10.26434/chemrxiv-2021-04-01-09-14-34.295524+00:00",
      "rdfs:label": "PaN Research Project"
    }
  ]
}

```

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

# Requirements and Expectations

- HELIPOINT 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 HELIPOINT can also answer our scientists' most important questions, such as:

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

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



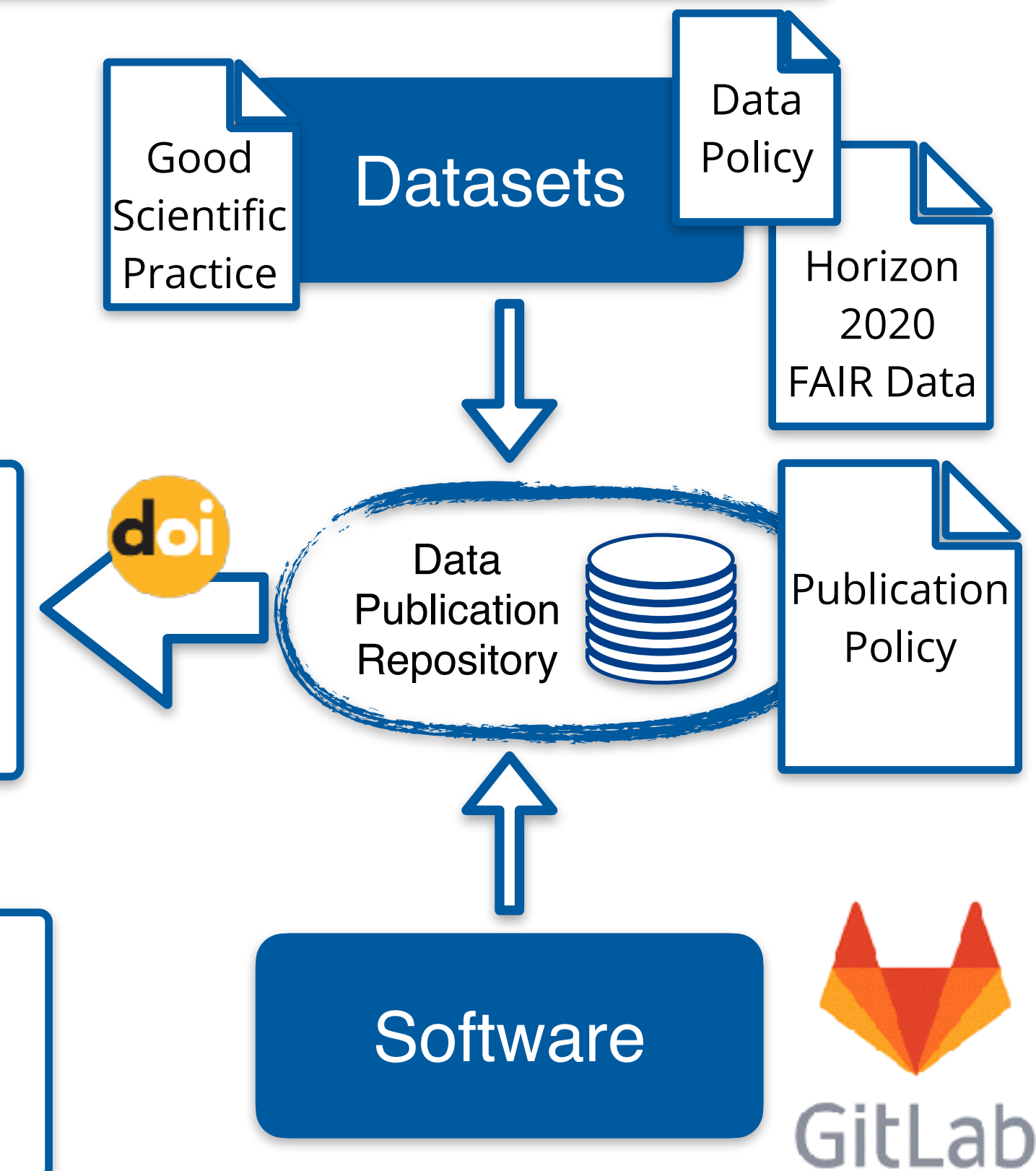
What are the necessary steps towards a full comprehensible and FAIR research experiment ensuring data provenance?

And how we can support them?!



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

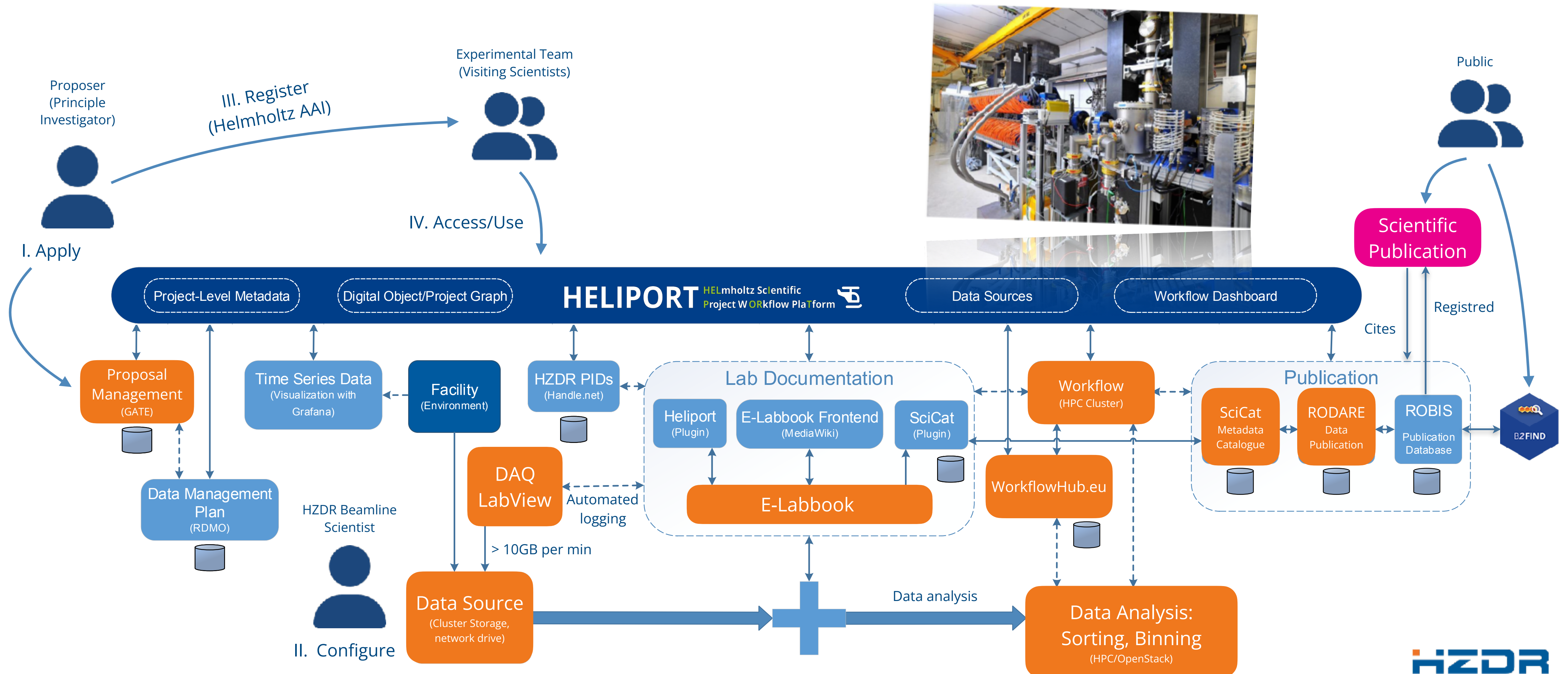
Where are data, software and how can I gain **access** to both of them?





# Example: HELIPORT Supported Experiment at The TELBE Beamline

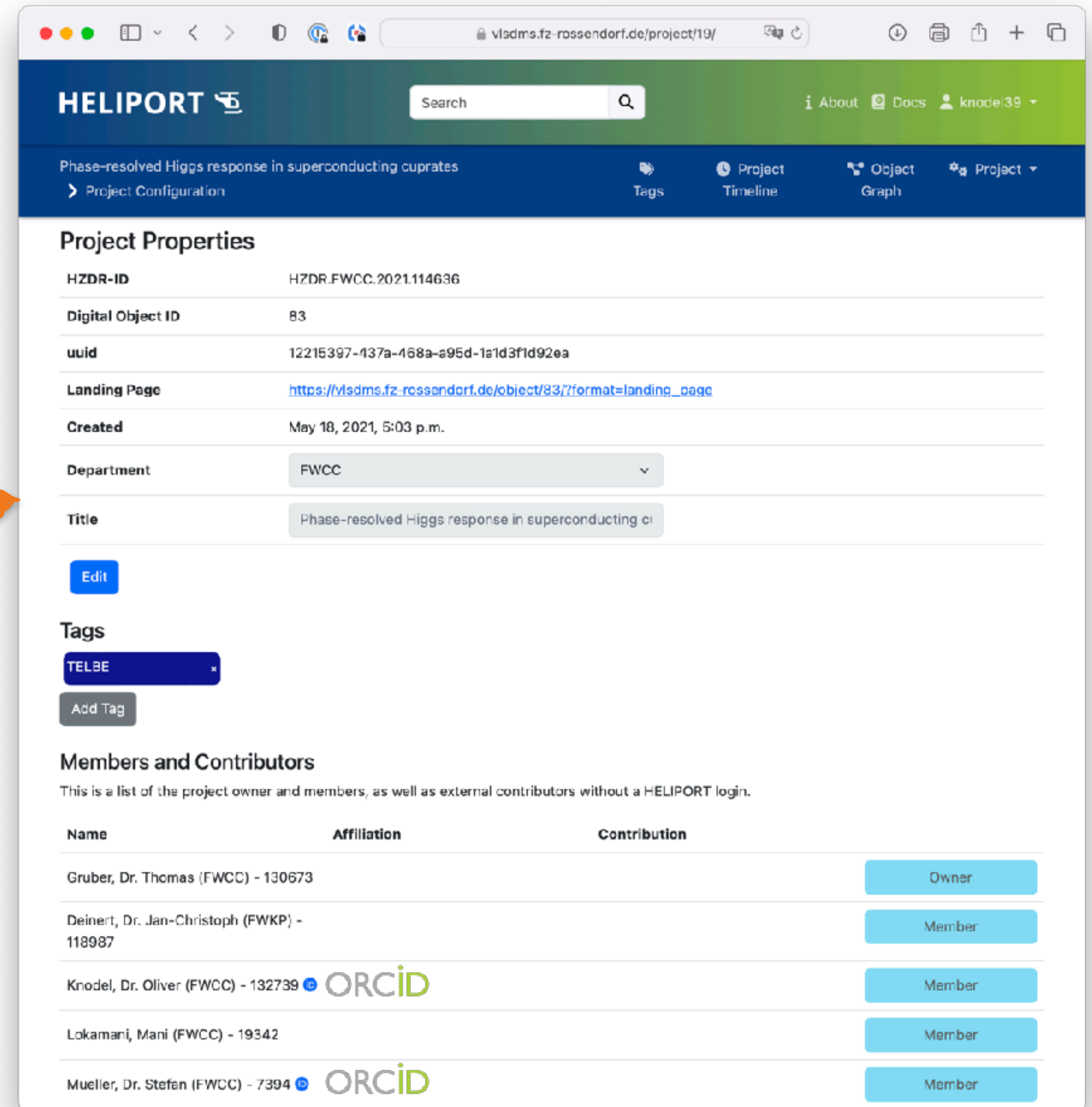
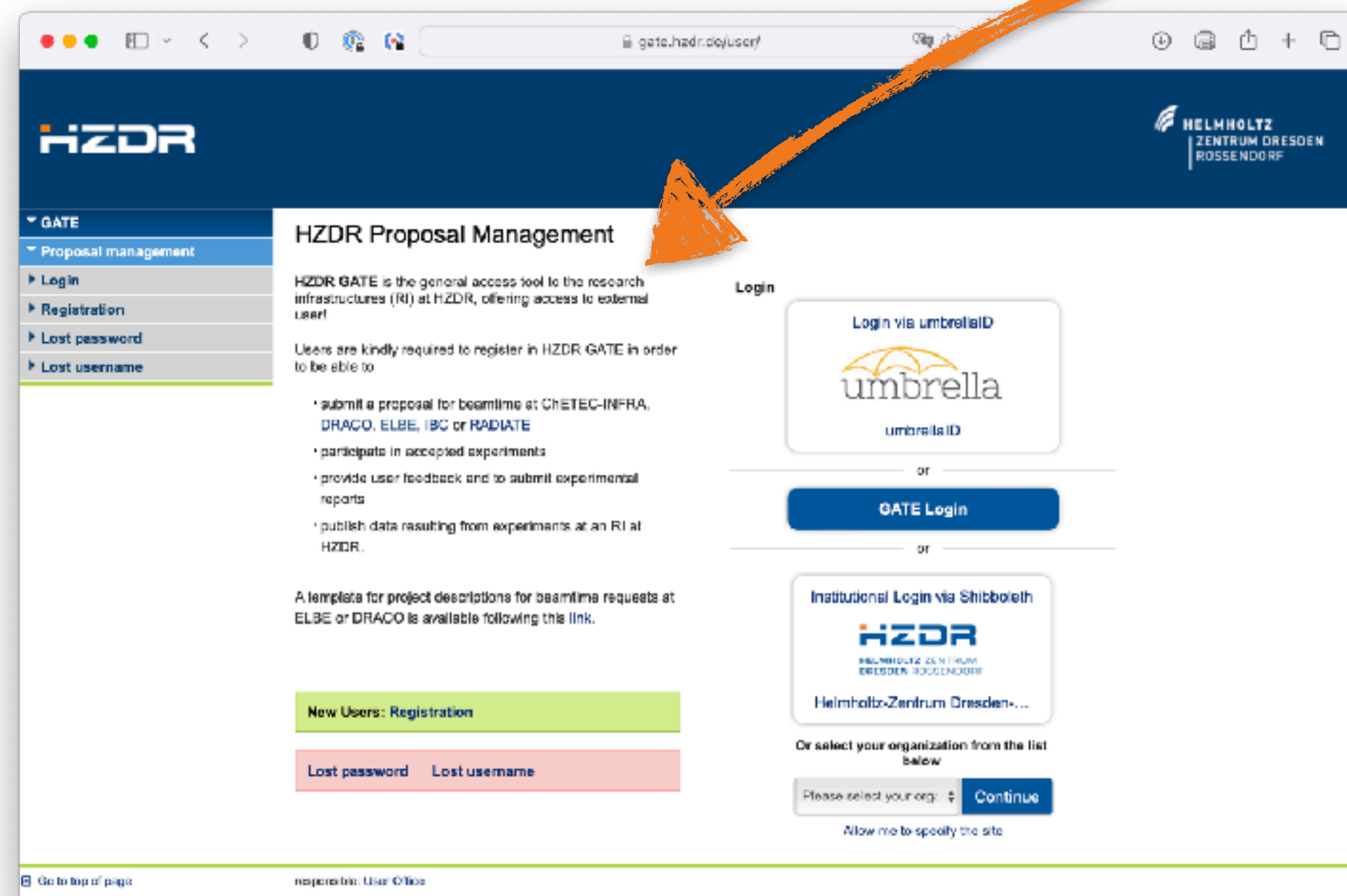
- HELIPORT provides access to our services and guides **external/visiting scientists** through the entire experiment.
- This requires the experiment to be mapped to systems in the HZDR infrastructure.



# I. Proposal Submission

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

- 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 HELIPOINT 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.

The left screenshot shows the 'Project List' page. It features a table with columns for Project Name, Last Modified, and Owner. The table lists several projects, including 'Semantic x-Lab', 'gELBE Projects', and 'Phase-resolved Higgs response in superconducting cuprates'. Each row has an 'Open' button. A 'Create Project' button is at the bottom left.

Project Name	Last Modified	Owner	Action
Semantic x-Lab	Jul 11, 2023	Voigt, Martin (FWCC-D) - 141575	Open
gELBE Projects	Apr 24, 2023	Mueller, Dr. Stefan (FWCC) - 7394	Open
gELBE beamtime 21102205-ST	Sep 11, 2023	Mueller, Dr. Stefan (FWCC) - 7394	Open
gELBE beamtime 21202619-ST	Sep 11, 2023	Mueller, Dr. Stefan (FWCC) - 7394	Open
Example parent project	Apr 24, 2023	Voigt, Martin (FWCC-D) - 141575	Open
ML Ops Project	Jun 06, 2023	Knodel, Dr. Oliver (FWCC) - 132739	Open
SOTA on Uncertainties	May 23, 2023	Pape, David (FWCC) - 139658	Open
Phase-resolved Higgs response in superconducting cuprates	May 23, 2023	Gruber, Thomas (FWCC-D) - 141575	Open
Digital Twin Showcase	Jun 07, 2023	Voigt, Martin (FWCC-D) - 141575	Open
Beamtime Dashboard Test	May 31, 2022	Voigt, Martin (FWCC-D) - 141575	Open
Rodare Data Publication Project	Aug 09, 2022	Knodel, Dr. Oliver (FWCC) - 132739	Open

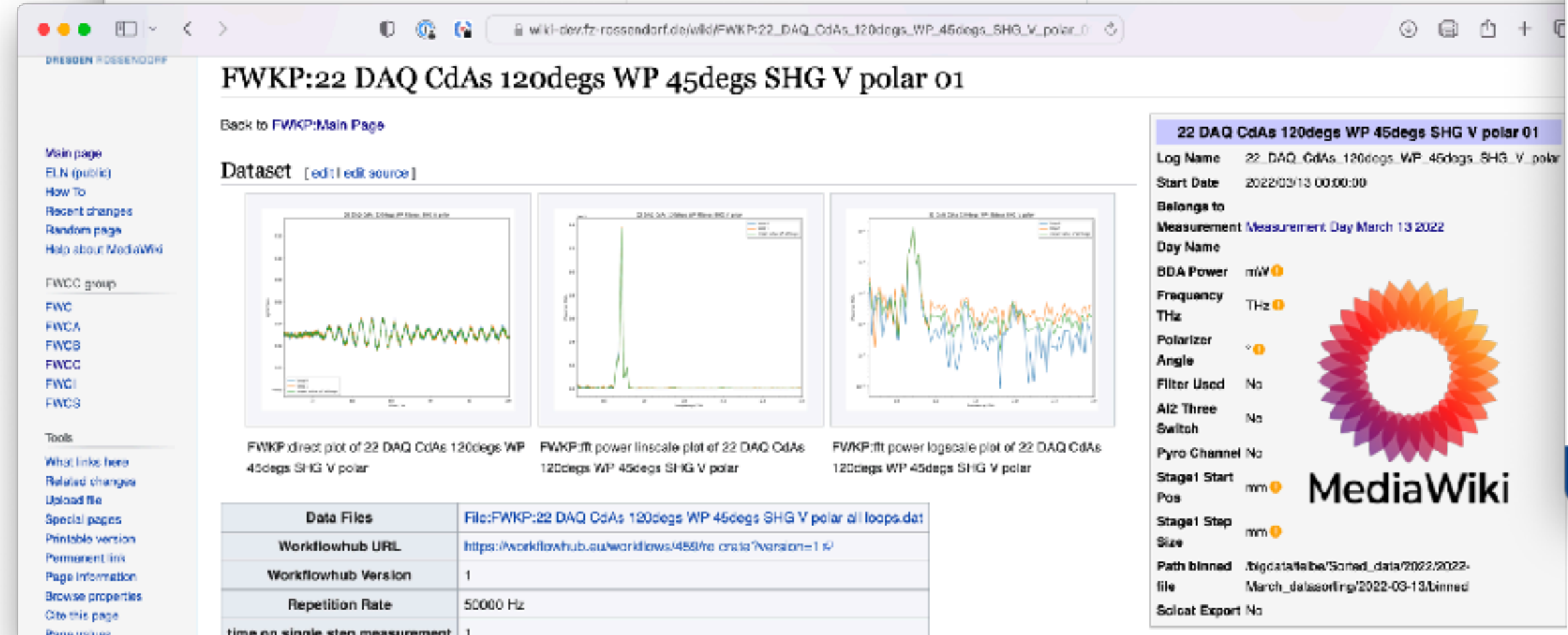
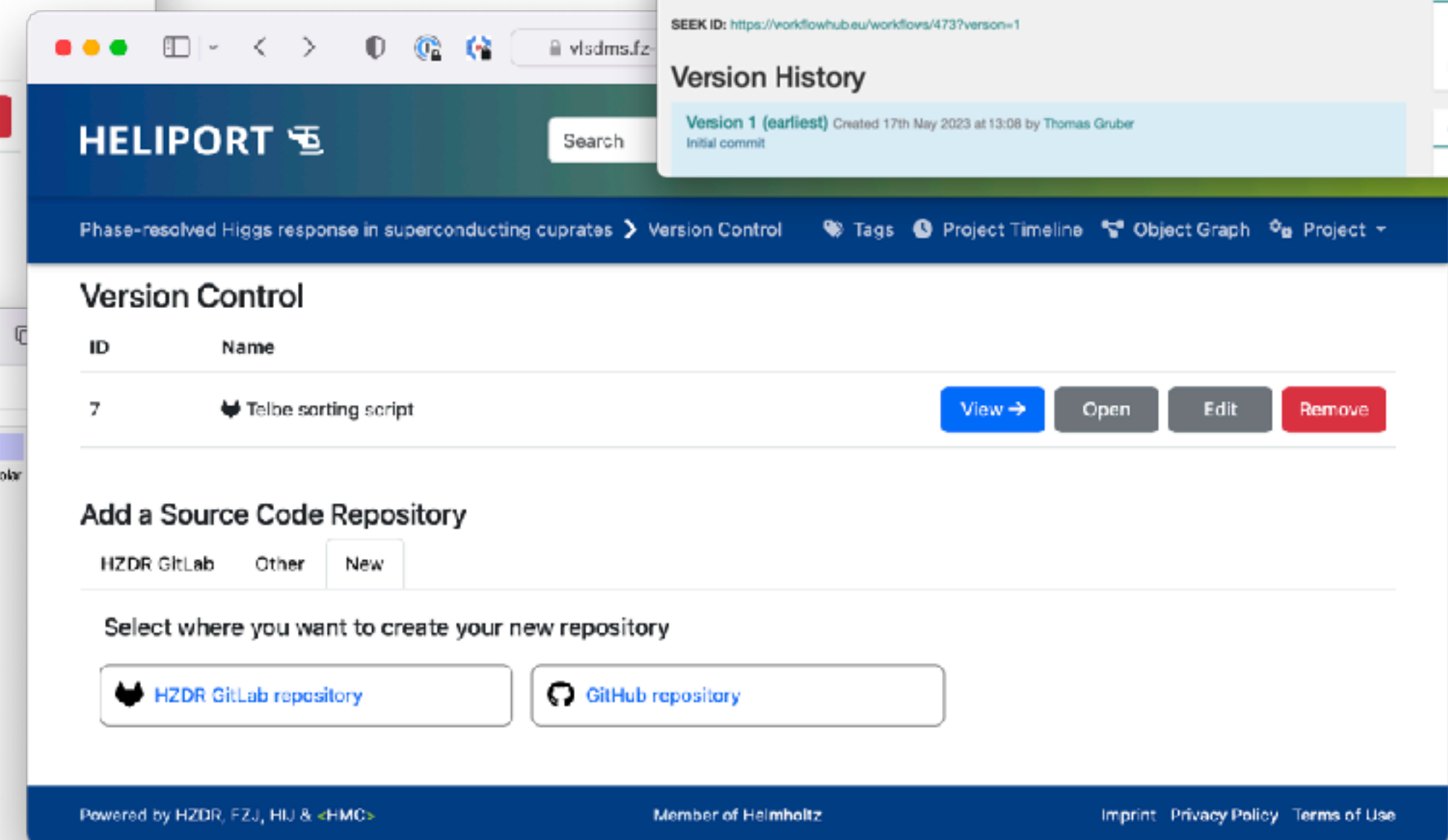
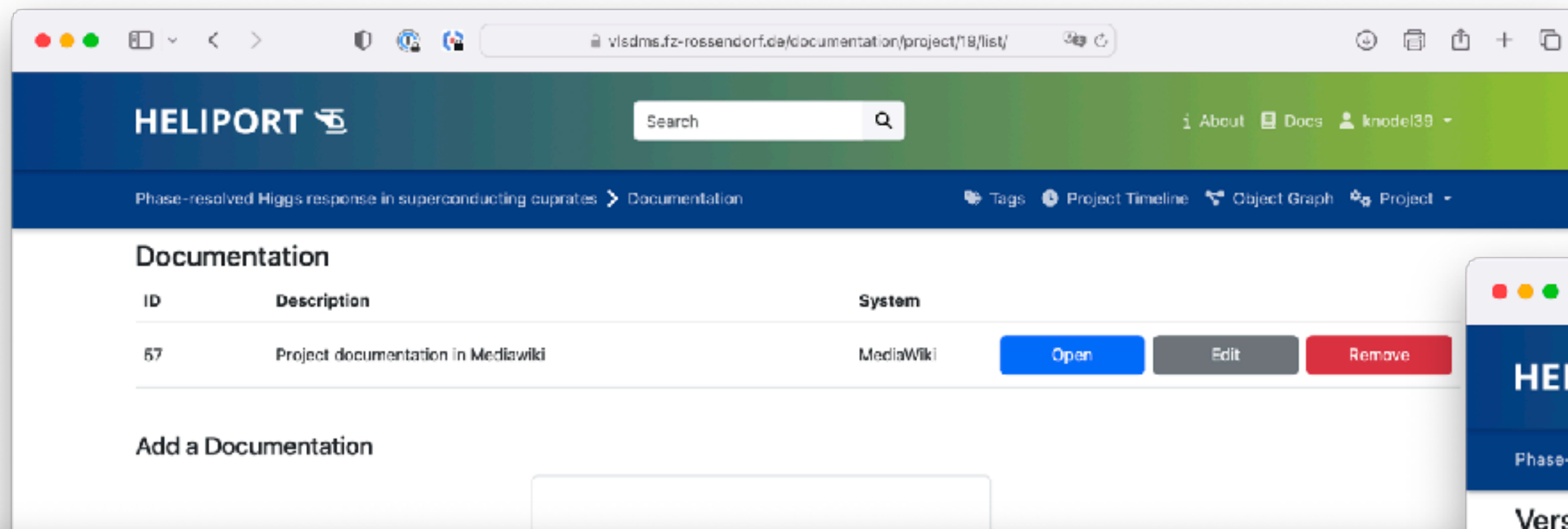
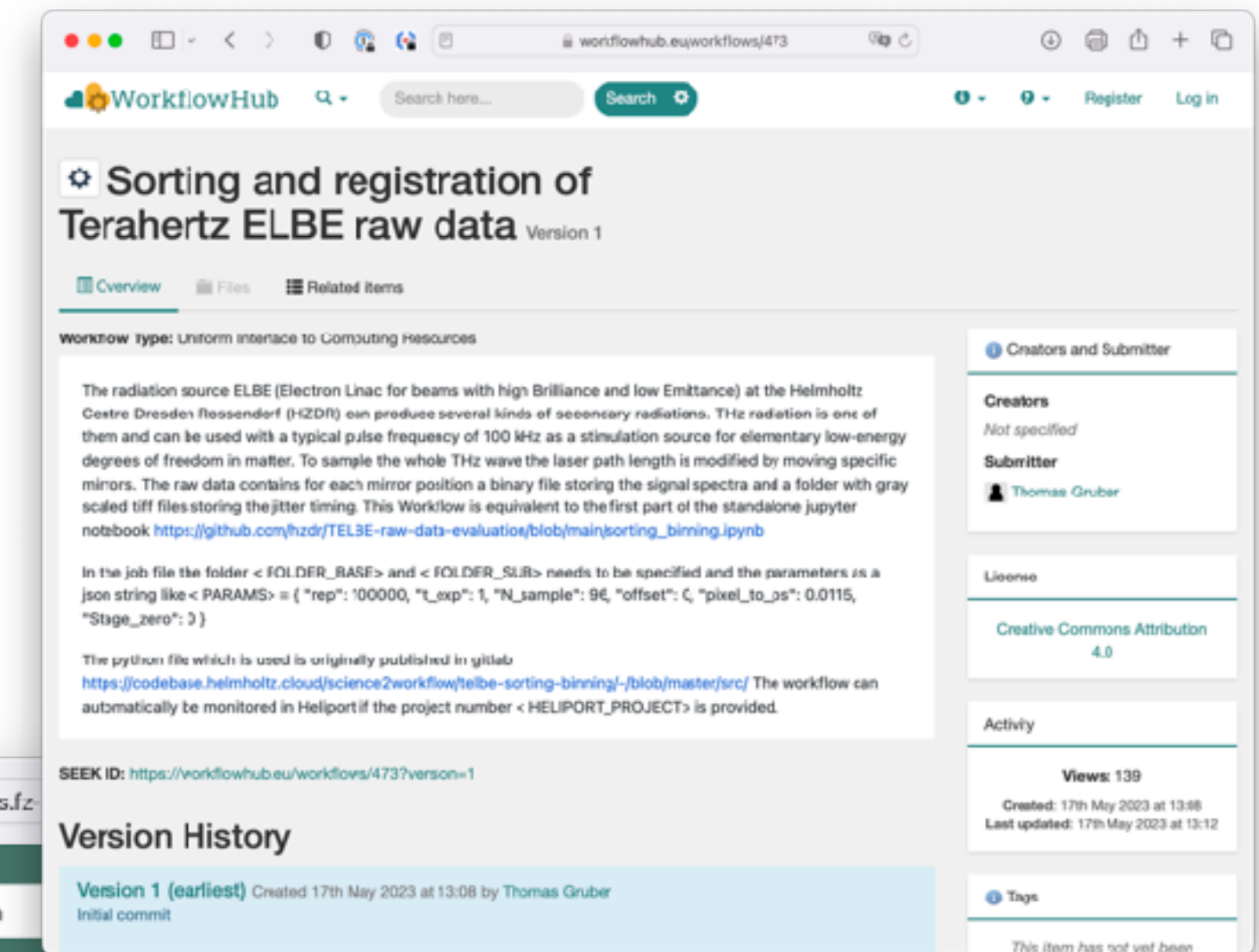
The right screenshot shows the 'Project Dashboard' for the project 'Phase-resolved Higgs response in superconducting cuprates'. It displays a workflow diagram with four main stages: Systems, Resources, Automation, and Results. Each stage contains several components connected by lines and plus signs, indicating a flow or relationship between them.

- Systems:** Version Control, Data Management Plan, Documentation, Digital Objects.
- Resources:** Data Source, SSH Files/Directories, UNICORE Storages.
- Automation:** UNICORE Jobs.
- Results:** Archive, Publication.

# III. Resources: Documentation and Repositories

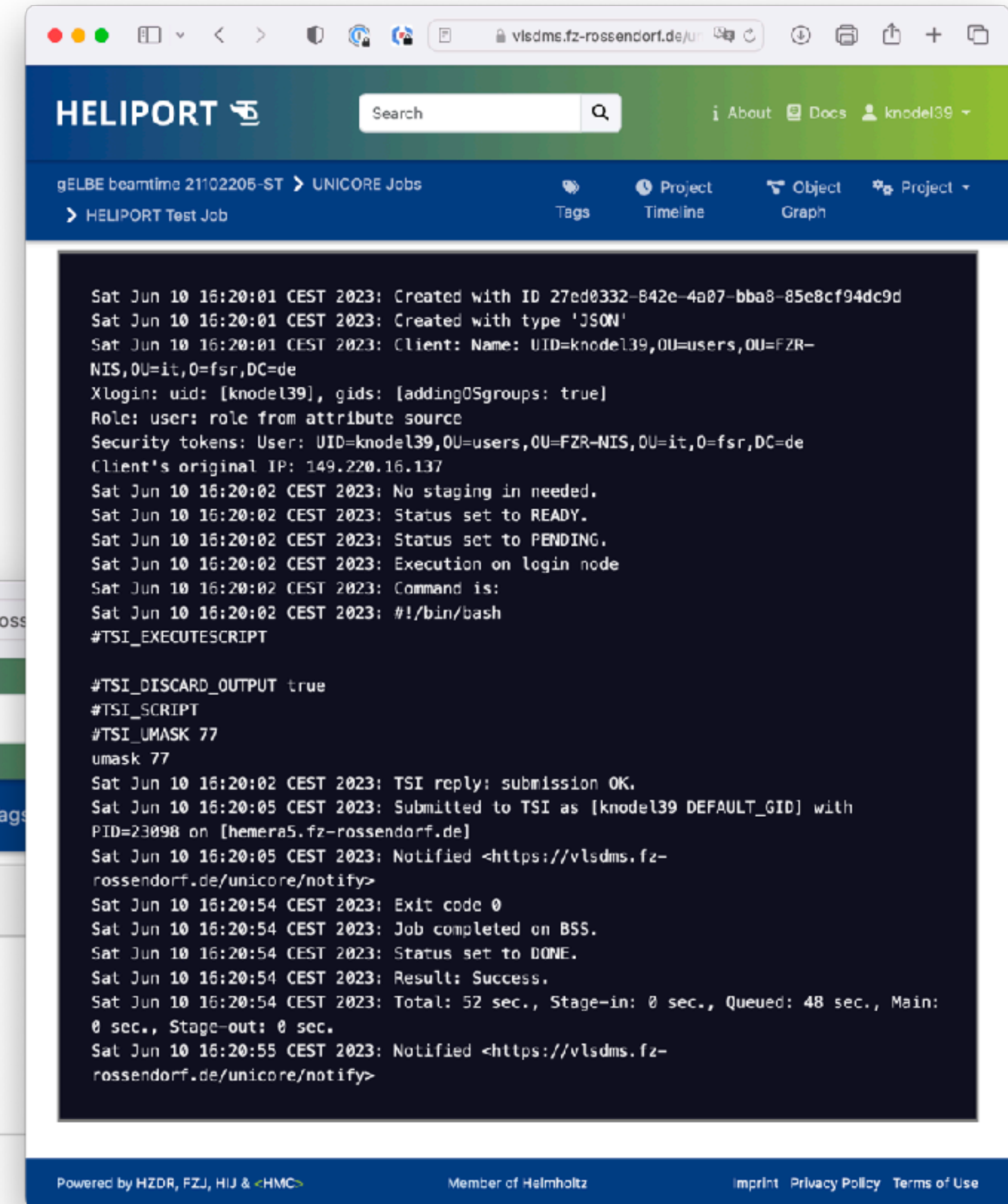
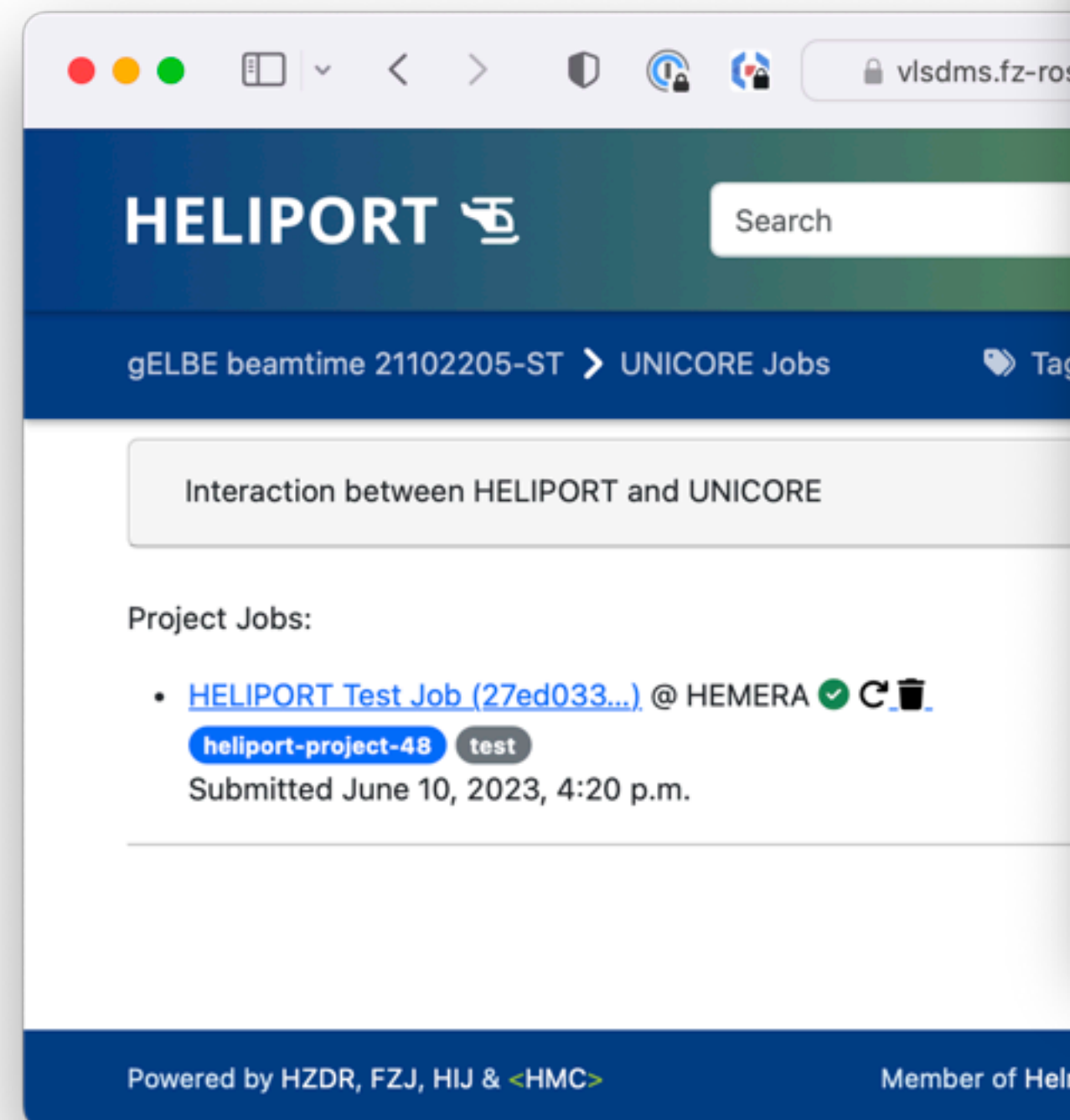
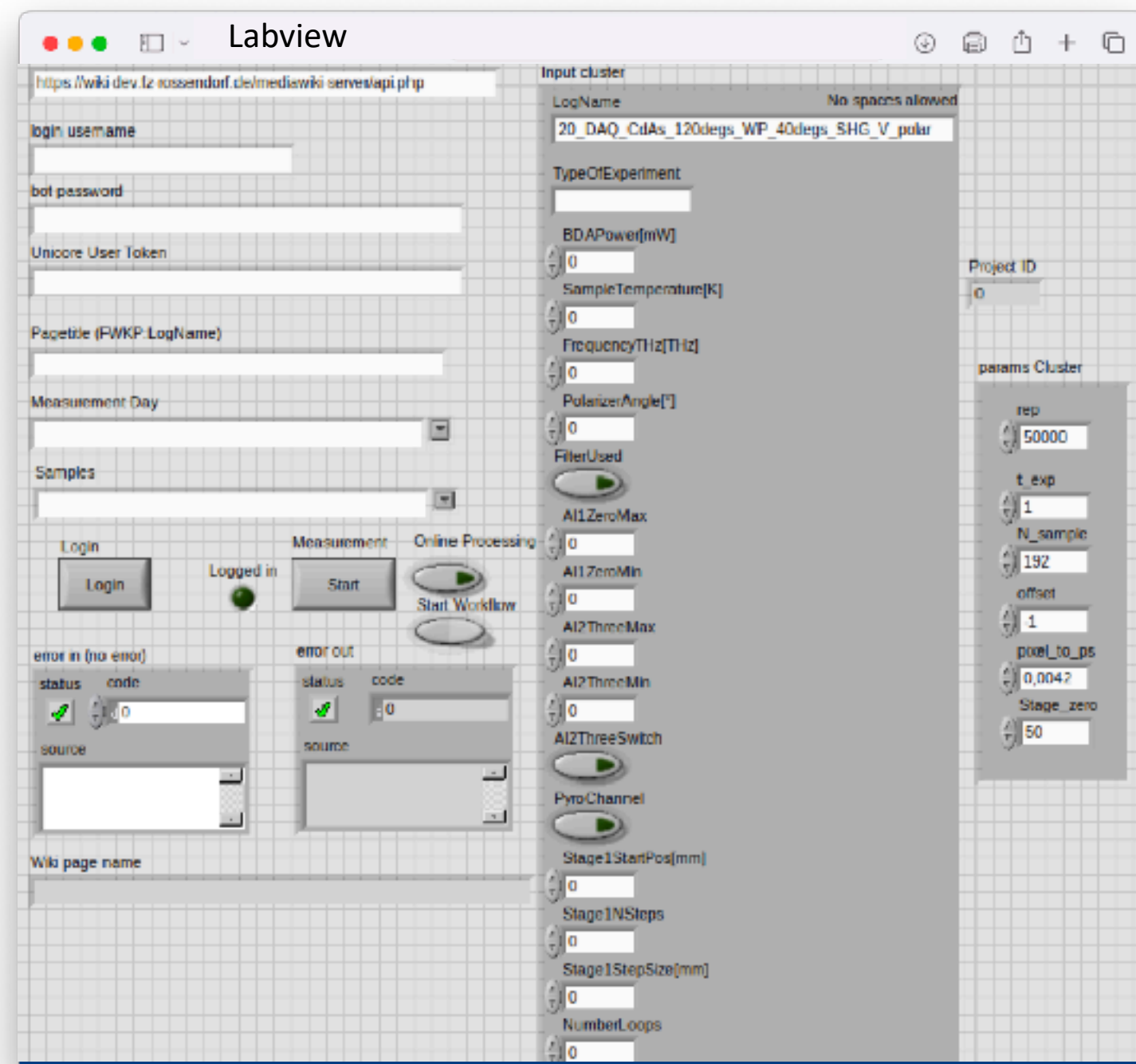
The documentation section is typically used to refer to all internal and external systems or services used:

- E-Labbook (Mediawiki),
- GitLab, Github, Workflowhub, ...



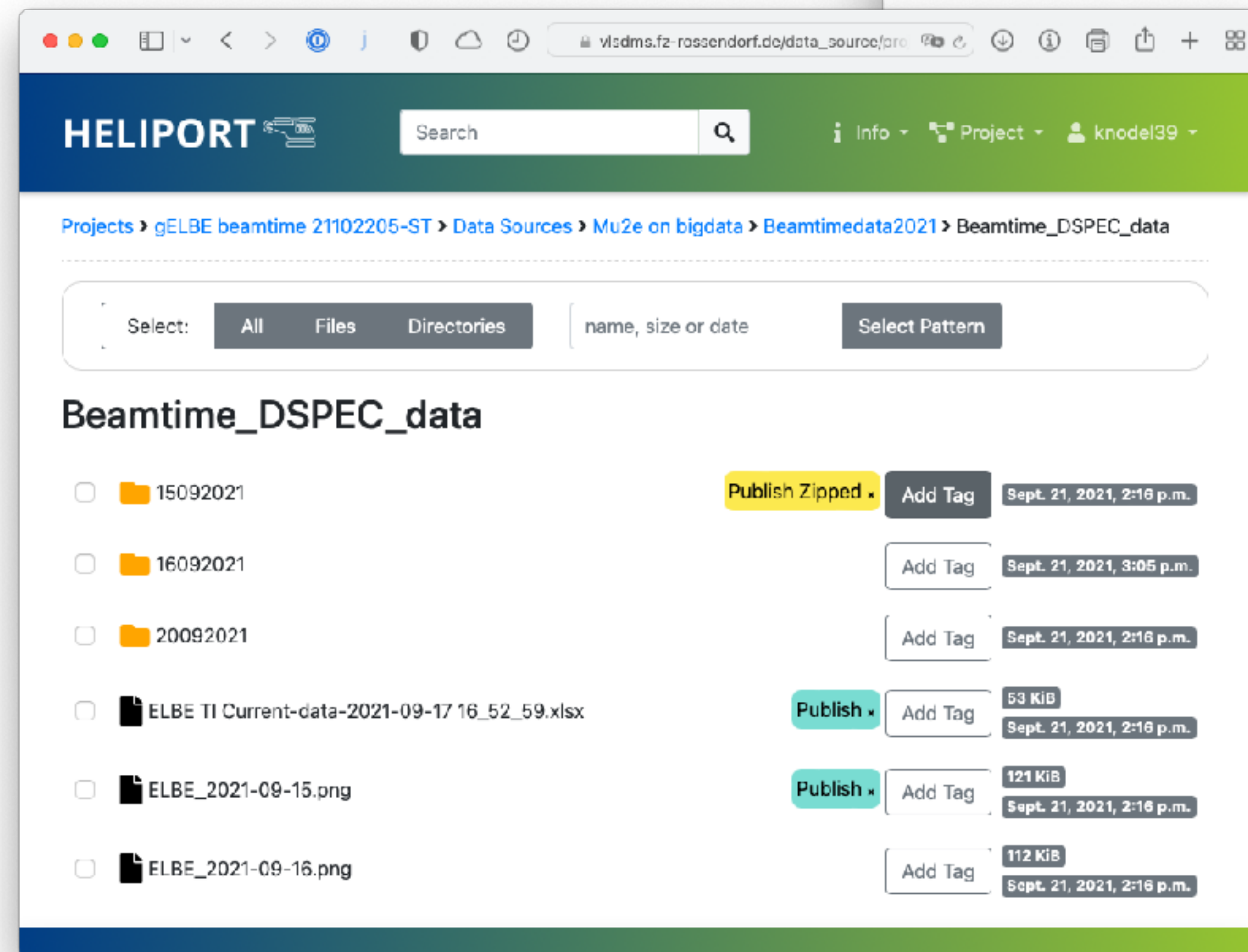
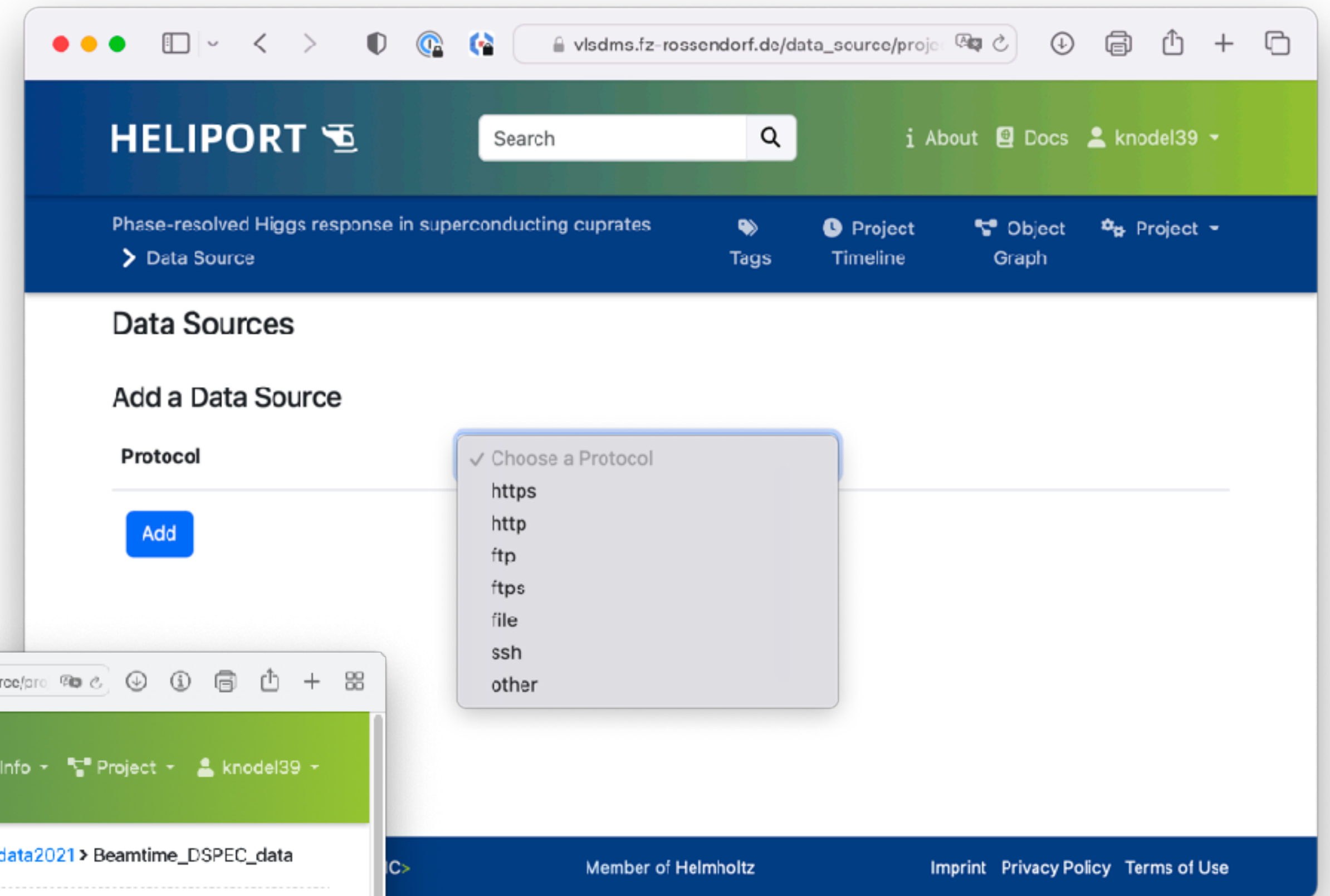
# IV. Detector Control and Workflows

- 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 HELIPOINT as **data source**.
- Each **project member** has access to the files and folders using HELIPOINT.
- 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 HELIPOINT.

The image illustrates the integration of HELIPOINT and RODARE in a data publication workflow. It features several overlapping screenshots:

- HELIPORT Project Configuration:** Shows project details for "Phase-resolved Higgs response in superconducting cuprates", including HZDR-ID, Digital Object ID, and a list of members and contributors.
- HELIPORT Systems Diagram:** A flowchart showing the integration of "Version Control", "Data Management Plan", and "Documentation" with "Data Source", "SSH Files/Directories", and "UNICORE Storages", leading to "Automation" via "UNICORE Jobs".
- HELIPORT File Selection:** A screenshot of the "Second Day" file selection interface, showing a list of files with checkboxes and "Add Tag" buttons.
- RODARE Publication Record:** A screenshot of the RODARE website showing the publication record for "Phase-resolved Higgs response in superconducting cuprates", including the DOI (10.14278/rodare.1289), keywords, and a file tree structure.

Orange arrows indicate the flow of data and metadata from the HELIPORT project configuration and file selection stages into the RODARE publication record.





# Metadata Catalogue SciCat and Data Repository RODARE (Draft)

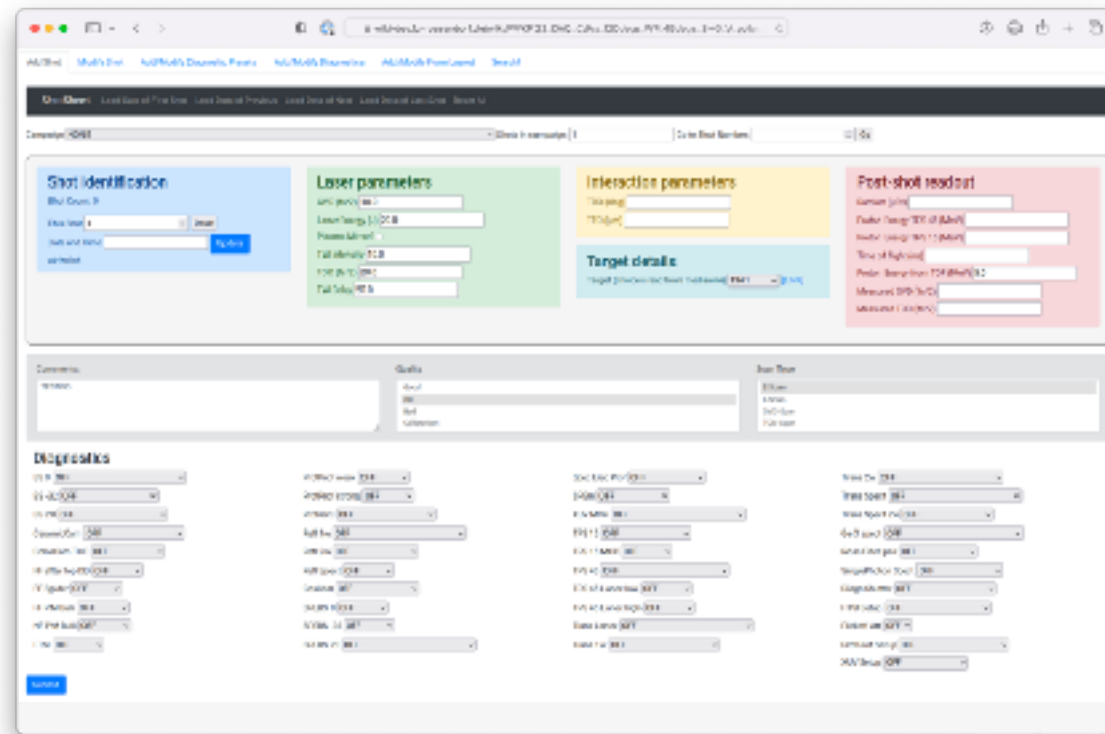
## Curated Metadata Source

## Public Metadata Catalogue

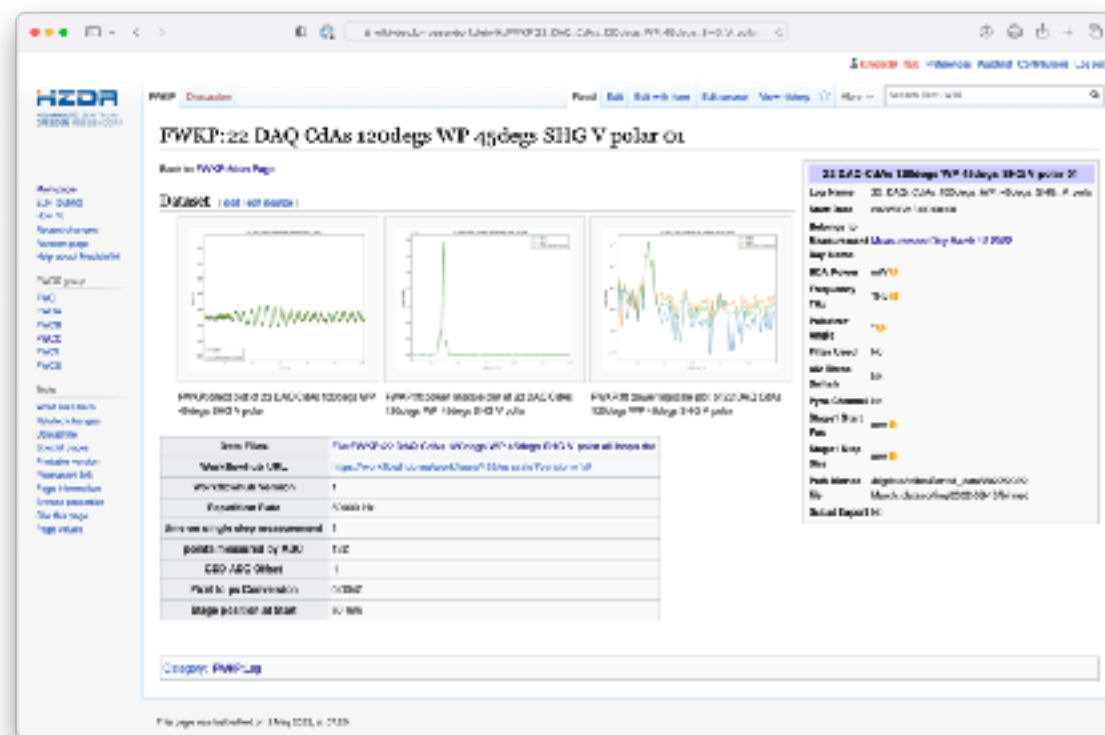
## Data Access

Metadata from Experiment/Simulation

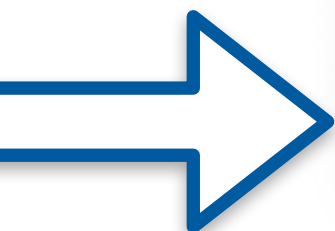
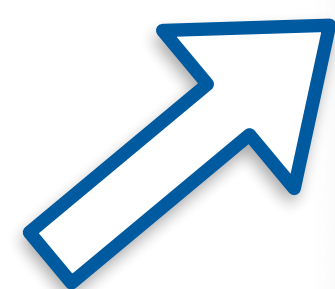
ExperimentLogging app (ExL)



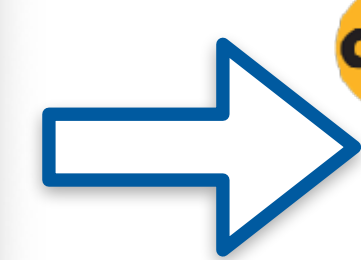
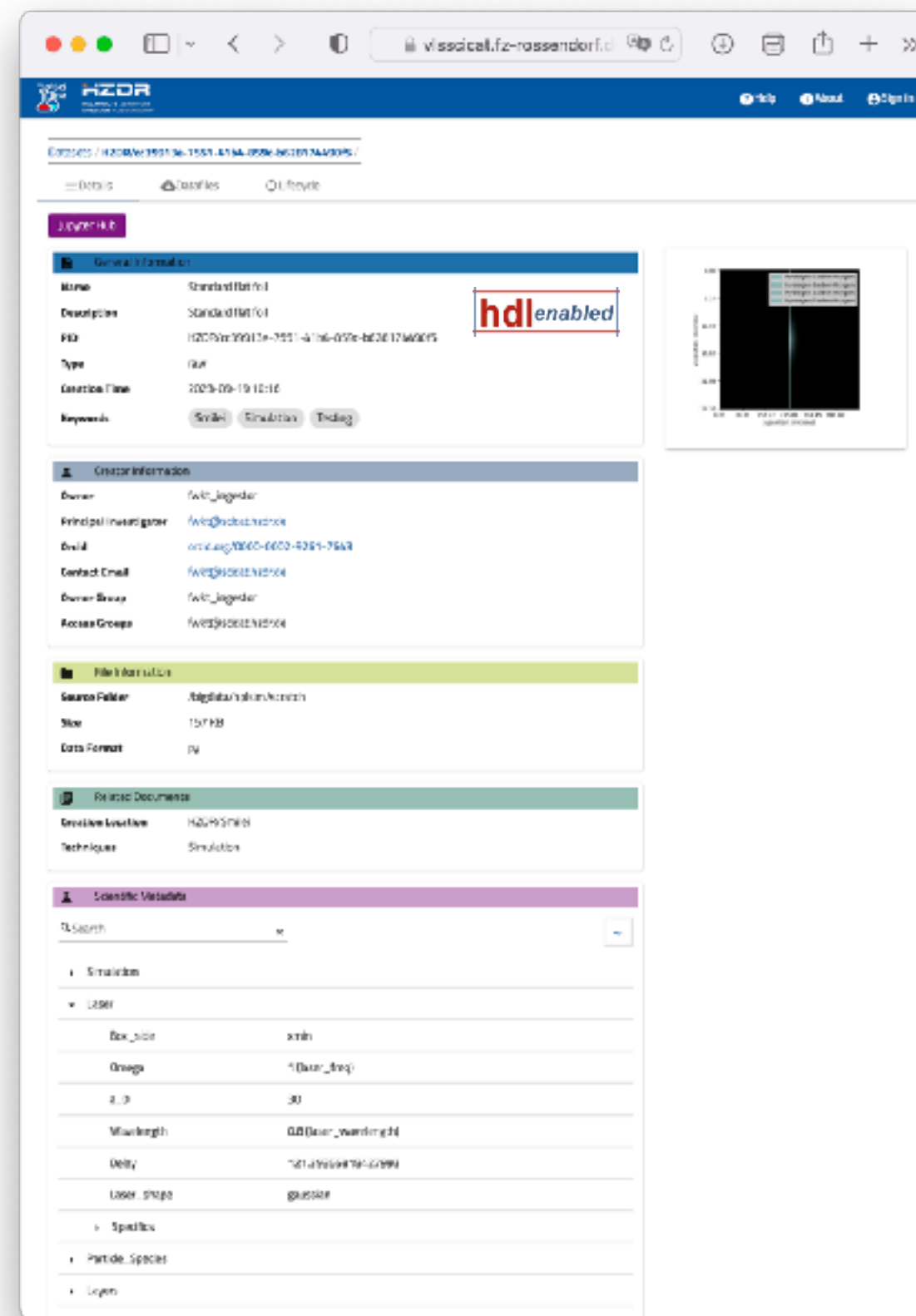
E-Logbook



Direct API Call



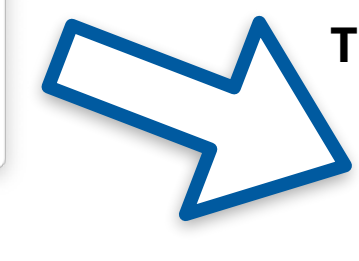
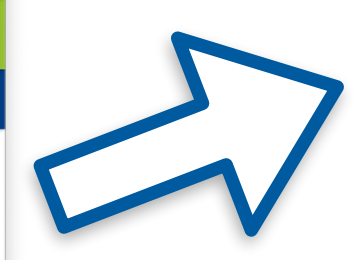
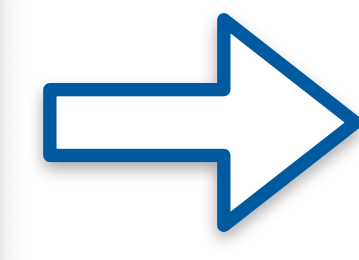
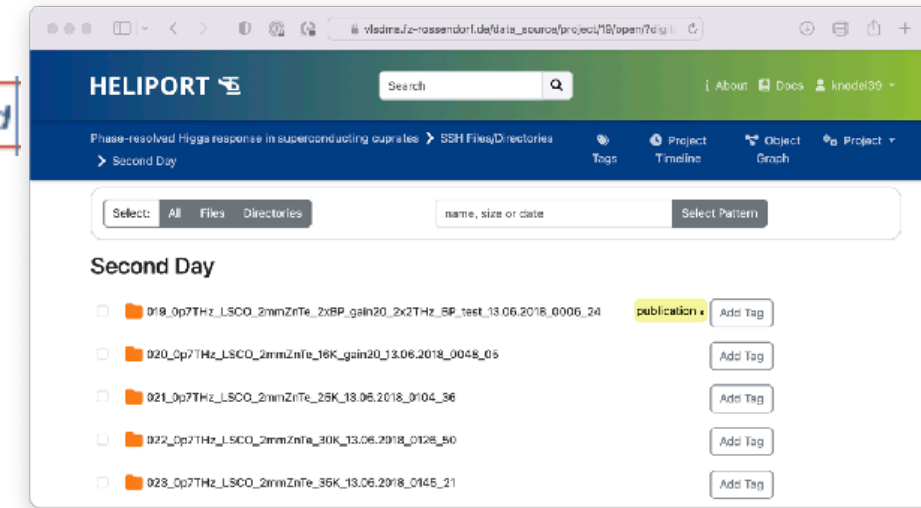
SciCat



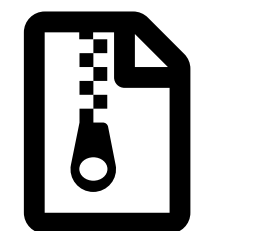
RODARE



HELIPORT



Dataset



Filesystem

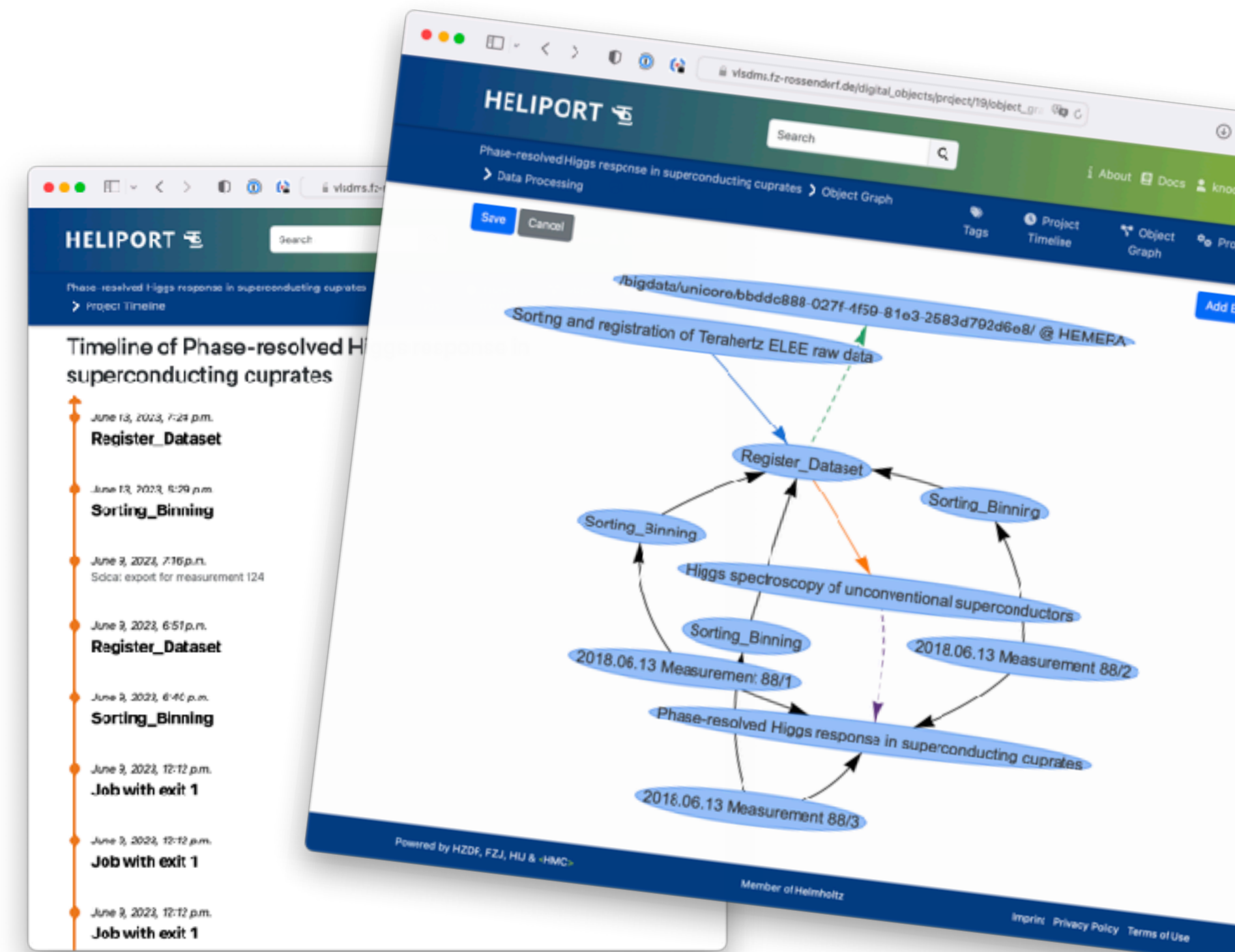
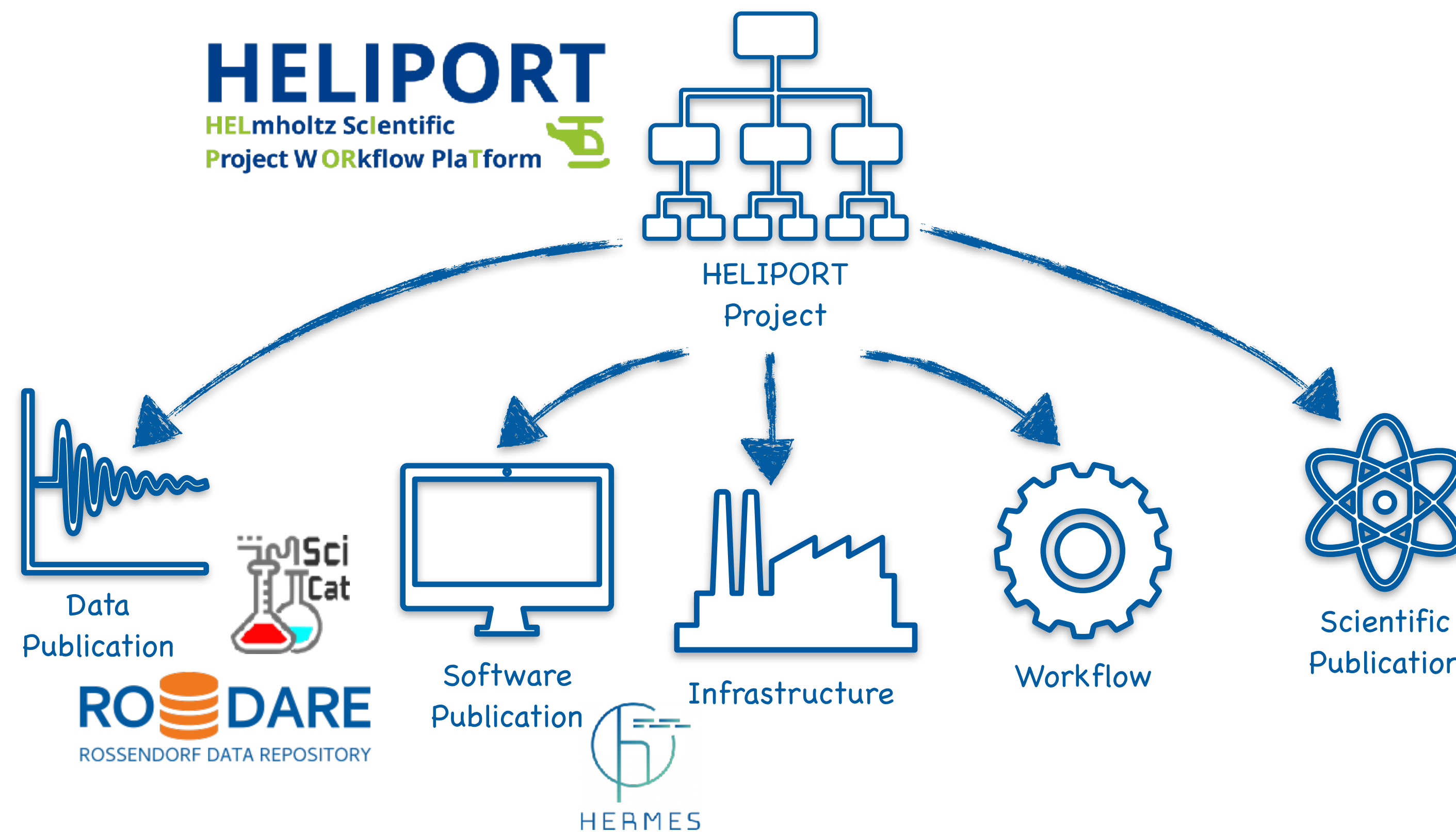


Tape Archive



# 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](https://heliport.hzdr.de)

Repository: [codebase.helmholtz.cloud/heliport](https://codebase.helmholtz.cloud/heliport/heliport)



Intuitive and structured user interface



## listProjects

Showing the most general HELIPORT project properties. Request more detailed information by appending 'uris' to the url. Go to a specific project by appending its id! to the url. For more information on how to authenticate look in HELIPORT (user > settings)

### QUERY PARAMETERS

Parameter	Type	Description
limit	integer	Number of results to return per page.
offset	integer	The initial index from which to return the results.
search	string	A search term.
group	string	group
owner	string	owner

## Response samples

```

{
  "count": 123,
  "next": "https://api.heliport.hzdr.de/api/projects/?page=2",
  "previous": "https://api.heliport.hzdr.de/api/projects/?page=1",
  "results": [
    {
      "id": "1",
      "name": "Project 1",
      "description": "Project 1 description"
    }
  ]
}

```

Name	Last commit	Last update
.gitlab	Enable pydocstyle checks	1 month ago
.vscode	Fix djlint and prettier formatting and lint...	5 months ago
.yarn/releases	Frontend Package Management and Lint...	8 months ago
LICENSES	Resolve "Separate SVGs of workflow vis...	5 months ago
config	Resolve "Remove unused Flower stuff"	9 months ago
docs	Enable pydocstyle checks	1 month ago
heliport	Enable flake8-return- checks	3 days ago
heliport_config	Ruff: Enable E501 (lines too long)	1 month ago

modern scientific collaborations and projects (MSCPs) employ various processing stages, starting with the proposal submission, continuing with data acquisition and concluding with final publications. The realization of such MSCPs poses a huge challenge due to (1) the complexity and diversity of the tools, (2) the heterogeneity of various involved computing and experimental platforms, (3) flexibility of analysis targets towards data acquisition and (4) data throughput. Another challenge for MSCPs is to provide additional metadata according to the FAIR principles for all processing stages for internal and external use. Consequently, the demand for a system, that assists the scientist in all project stages and archives all processes on the basis of metadata standards like DataCite to make really everything transparent, understandable and citable, has risen

<https://doi.org/10.1145/3456287.3465477>

## 1 INTRODUCTION

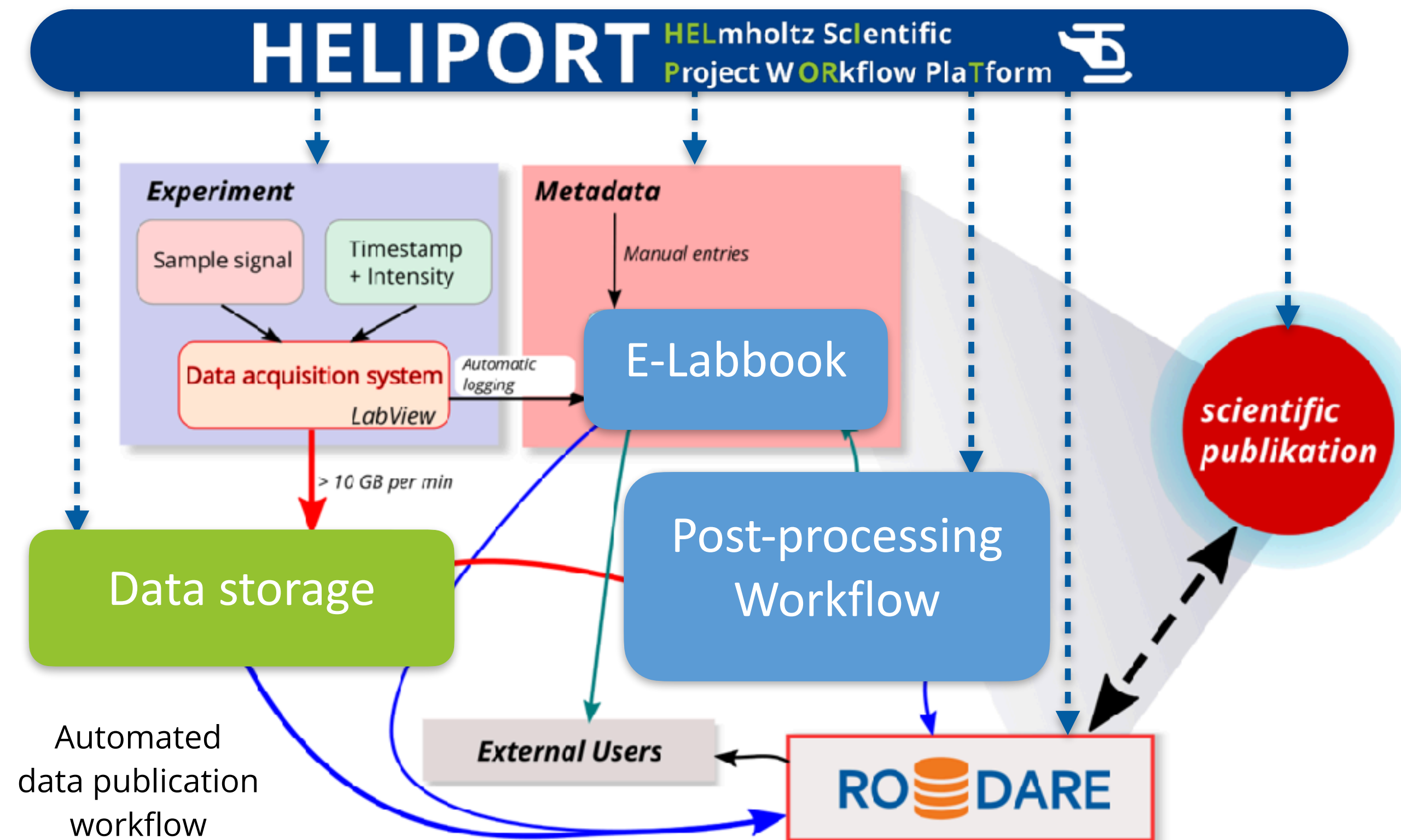
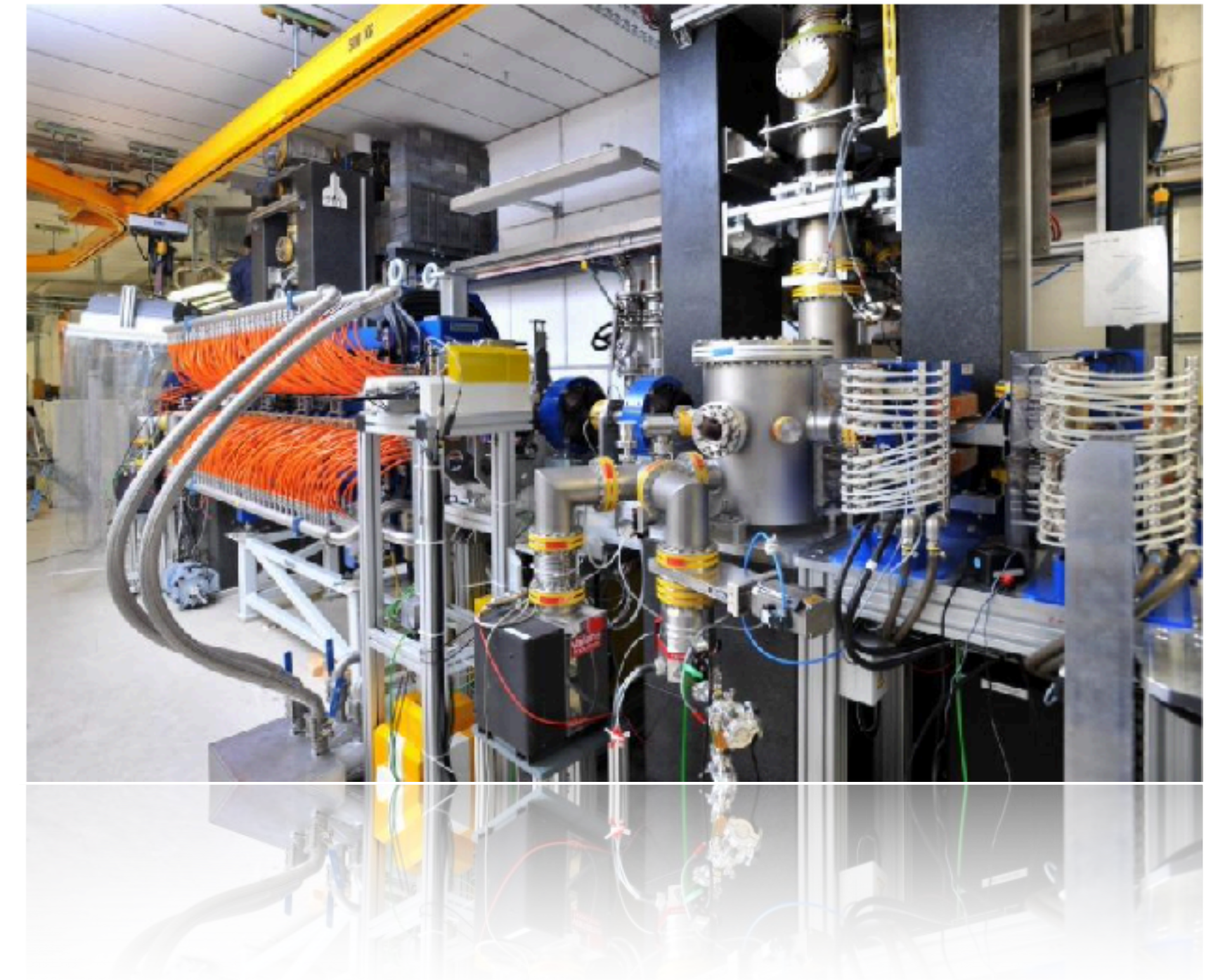
An essential objective of modern cutting-edge research should be to enable accessibility of the acquired research data and its re-usability across different research fields and their respective communities. The current generation of scientists is therefore faced with the challenging task of transferring experimental investigations into a data oriented research flow with strong focus on documenting every step closely following the FAIR [41] principles. The FAIR principles are well-established as standards in the field of research data management. The three pillars F (Findable), A (Accessible) and

# 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.



# Documentation Ecosystem at HZDR

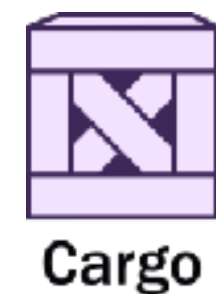
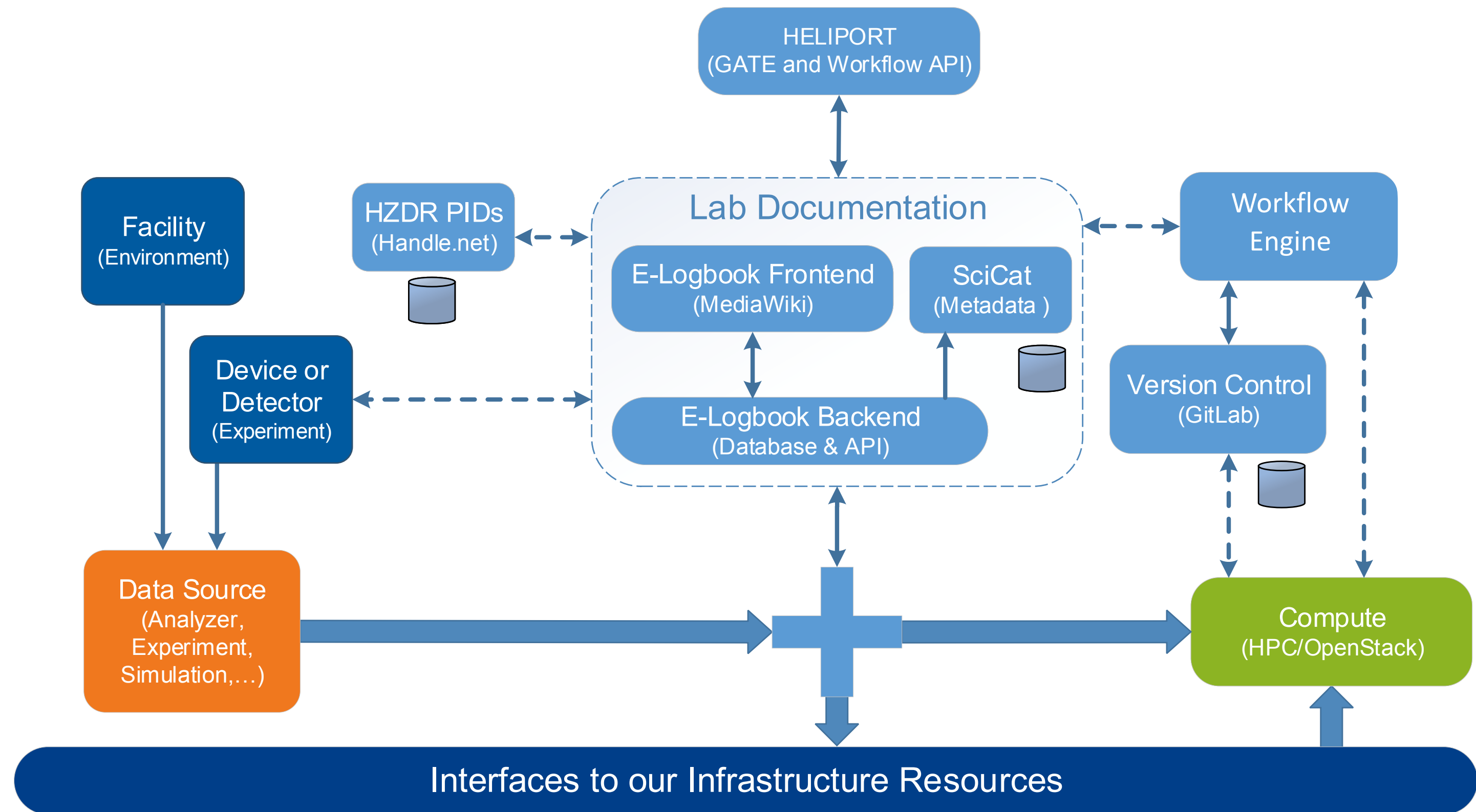
— Our **Electronic Lab Documentation** is a central database gathering information from various systems:

- 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).

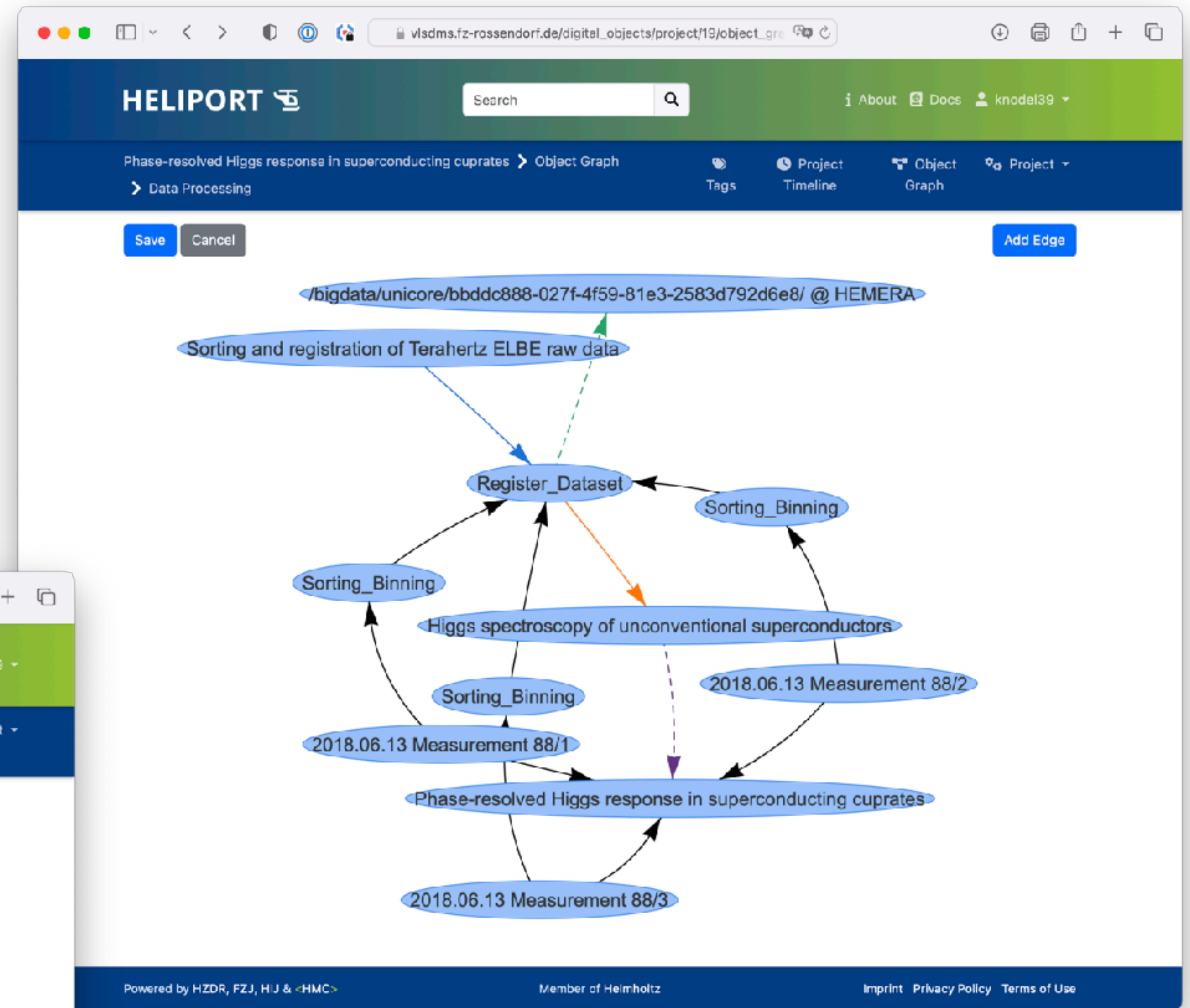
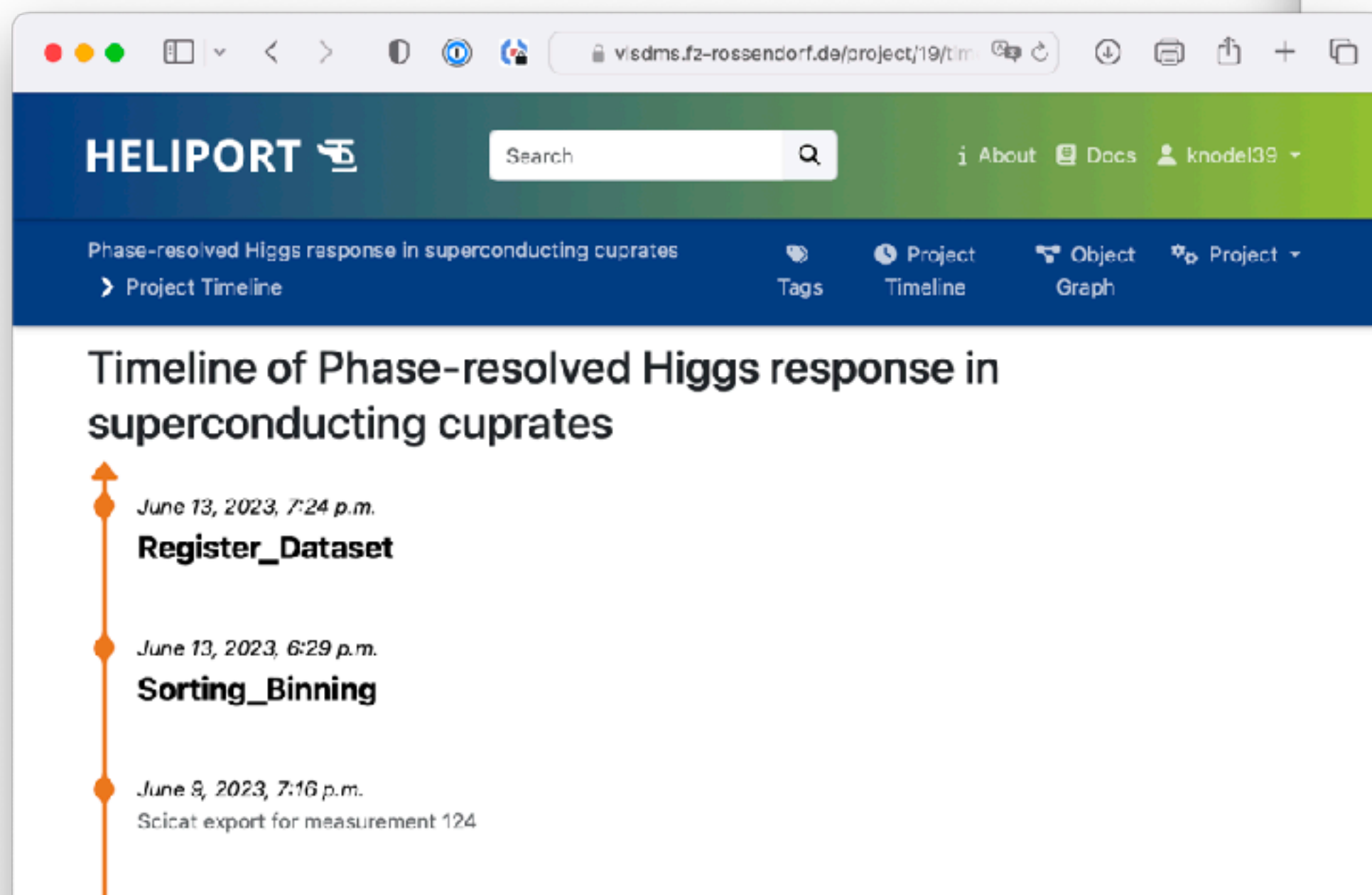
— The system is build on-top:



Cognitive Process Designer

# 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**

