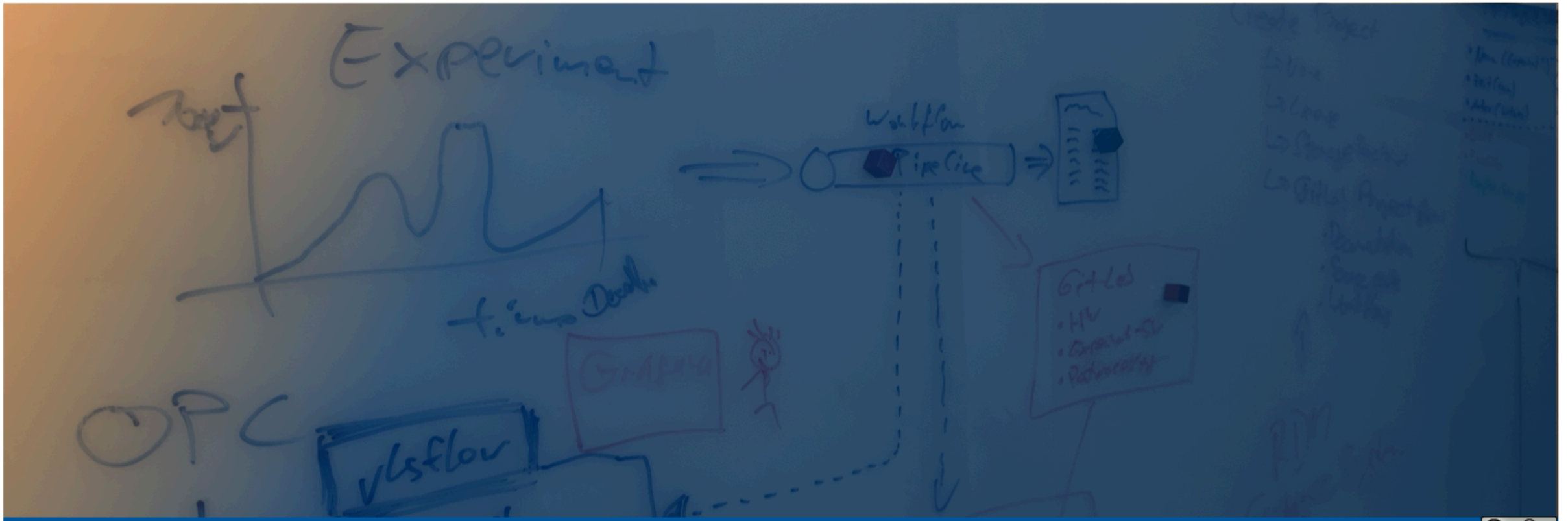
HELIPORT

JÜLICH FORSCHUNGSZENTRUM
HI JENA Helmholtz Institute Jena
HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF



Open Research Project Guidance System: HELIPORT

Helmholtz Open Science Forum: Research Software // February 6th, 2024

Tobias Huste, Oliver Knodel, Martin Voigt, Robert Ufer, David Pape, Mani Lokamani, Jeffrey Kelling, Stefan E. Müller, Thomas Gruber, Guido Juckeland, Alexander Kessler, Chien-Li Lee , Joachim Hein, Bernd Schuller // contact: o.knodel@hzdr.de

DRESDEN concept

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.

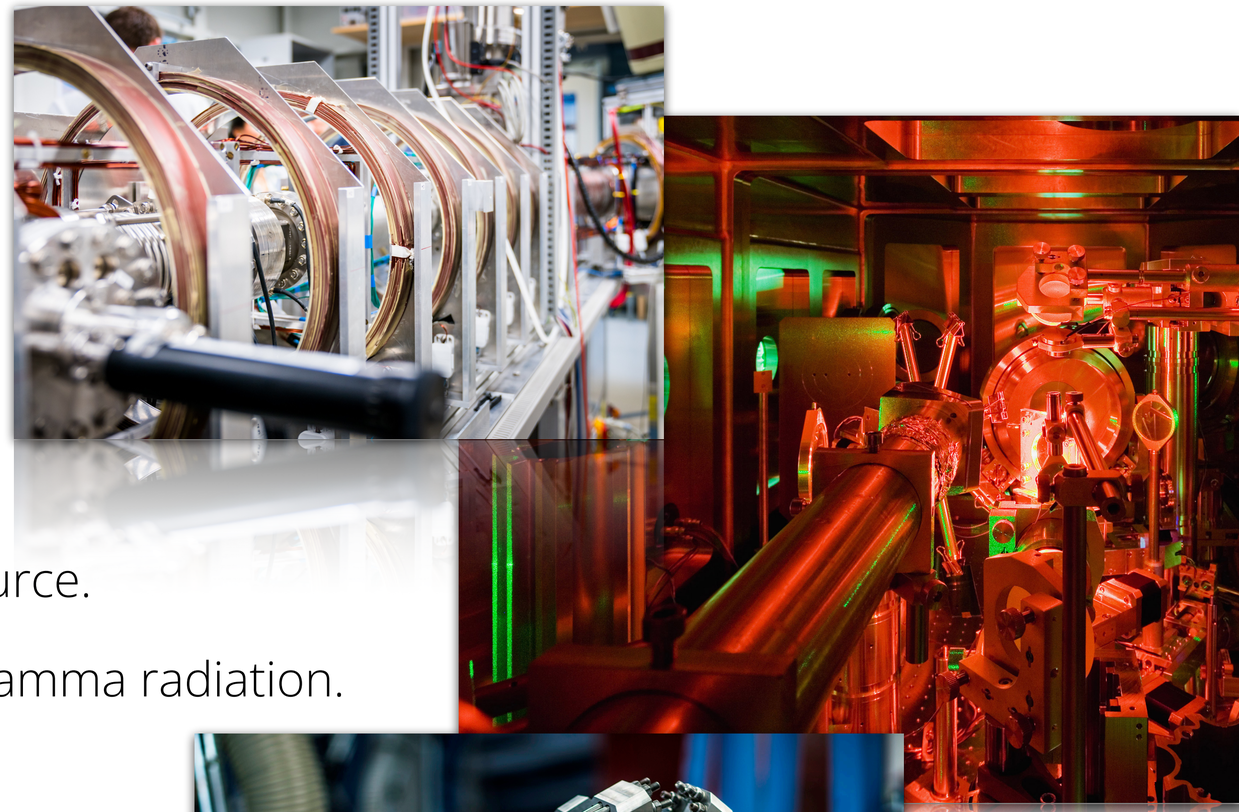
— Positrons, protons, neutrons as well as X-ray and gamma radiation.

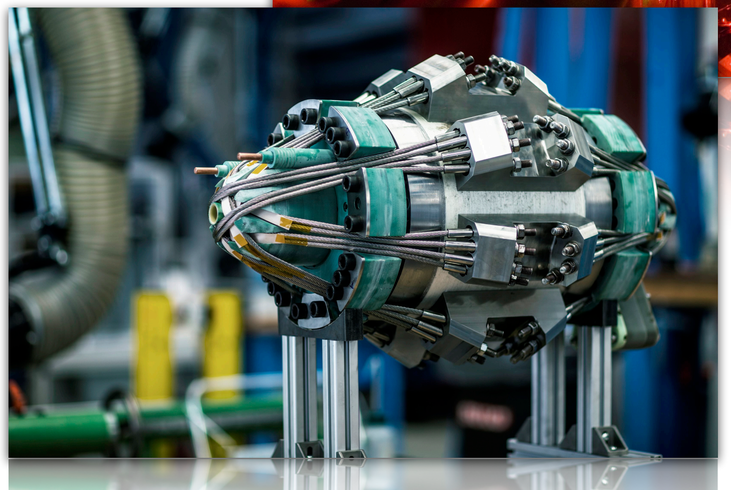
Dresden High Magnetic Field Laboratory (HLD)

— Europe’s highest pulsed magnetic fields.

Ion Beam Center (IBC)

— Nanoscale surface analysis and modification.

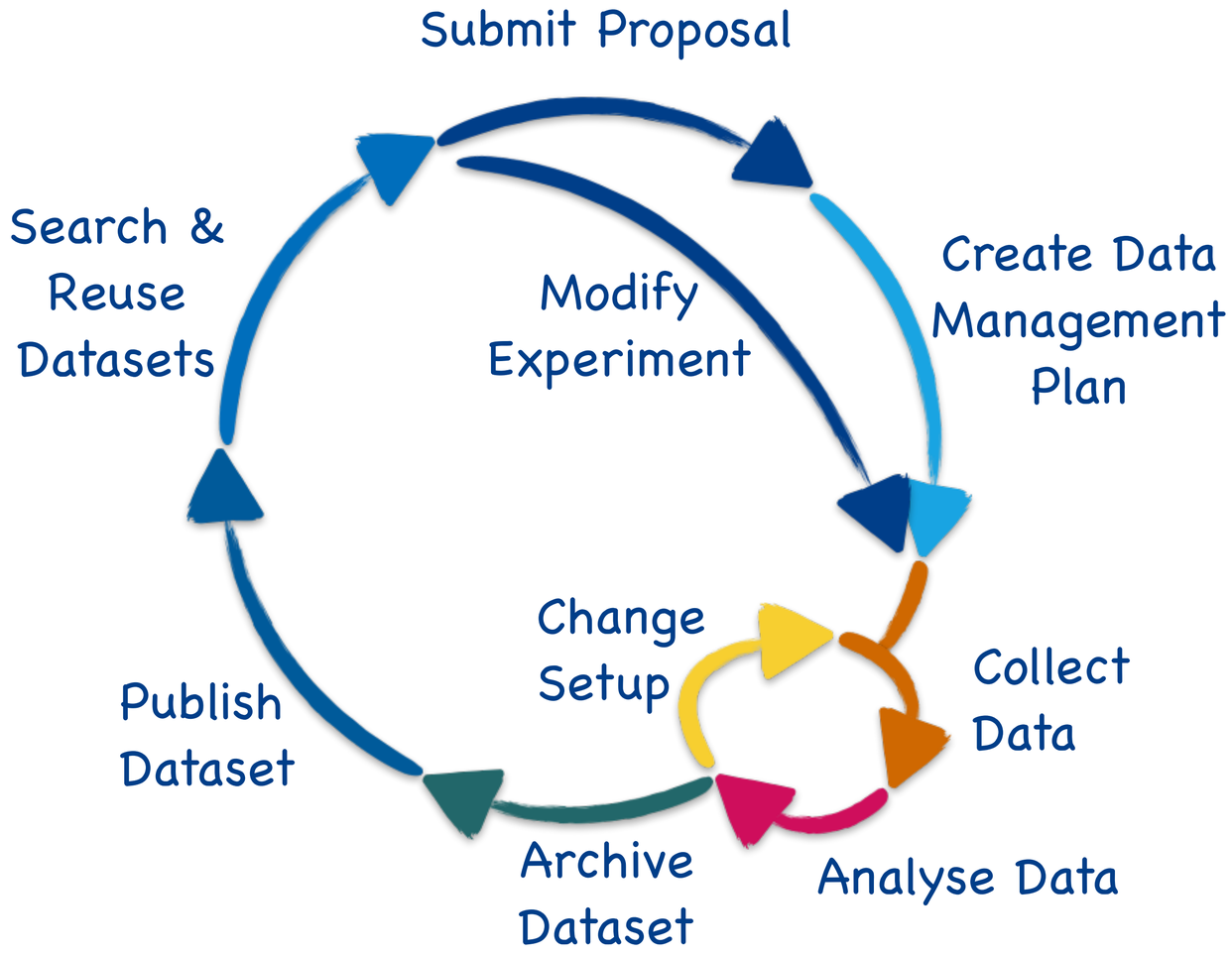




HZDR

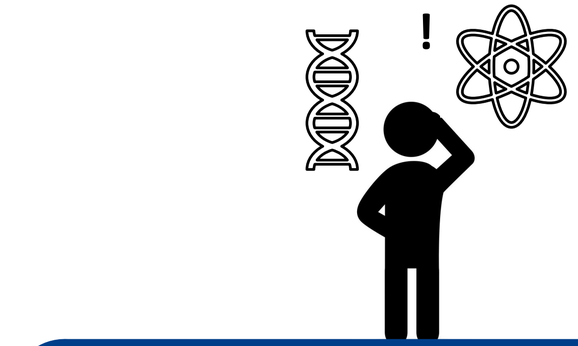
Our Challenge: An End-to-End Digital Data Lifecycle

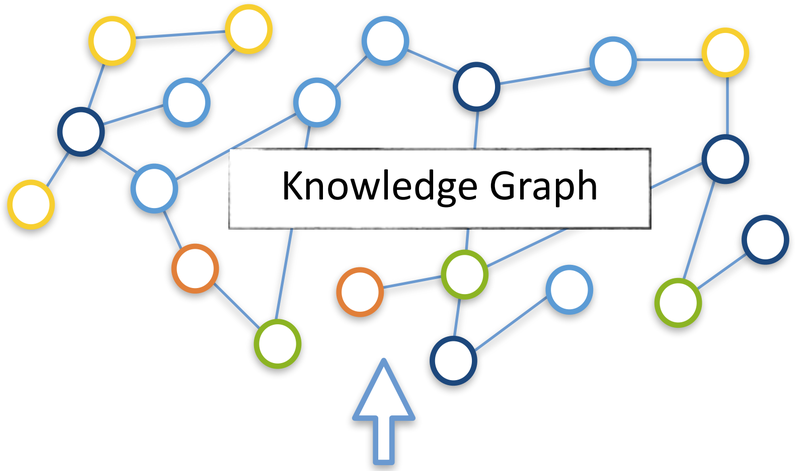
- We support many steps of our different research experiment (matter, energy and health) with tools:
  - electronic lab books,
  - interactive analysis,
  - publication of datasets,
  - scientific workflow management,
  - Handle generation and management.
- A uniform and smooth access to and between 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.

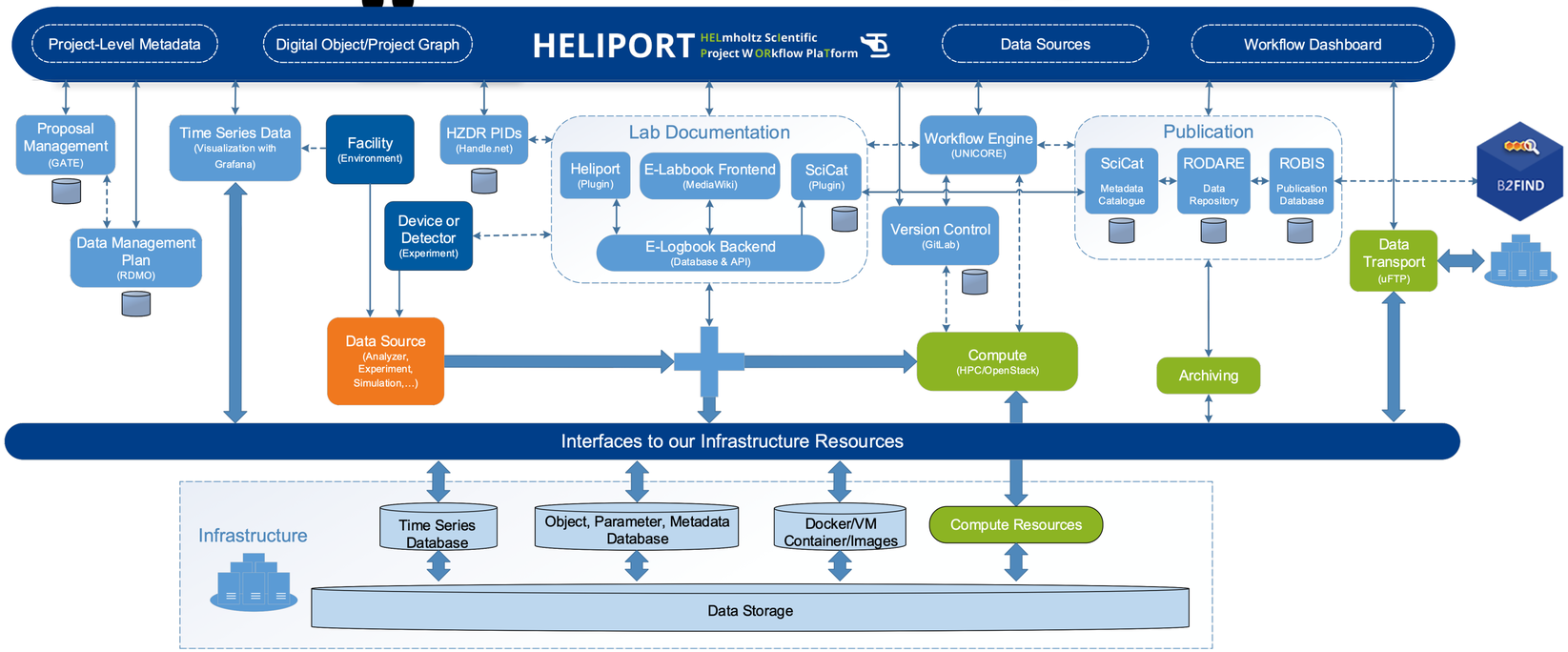


Our Observations and Experiences

- Our IT infrastructures can support various experiments, but they are complex…
- Scientists often don't know which services are available and how to use them.
- An overarching system guiding our scientists (and visitors) through the lifecycle of their research project is essential.
- In the future we can provide an overall Helmholtz-wide knowledge graph!







DRESDEN concept HZDR

HELIPORT

HELmholtz Sc**I**entific
Project W**OR**kflow Pla**T**form

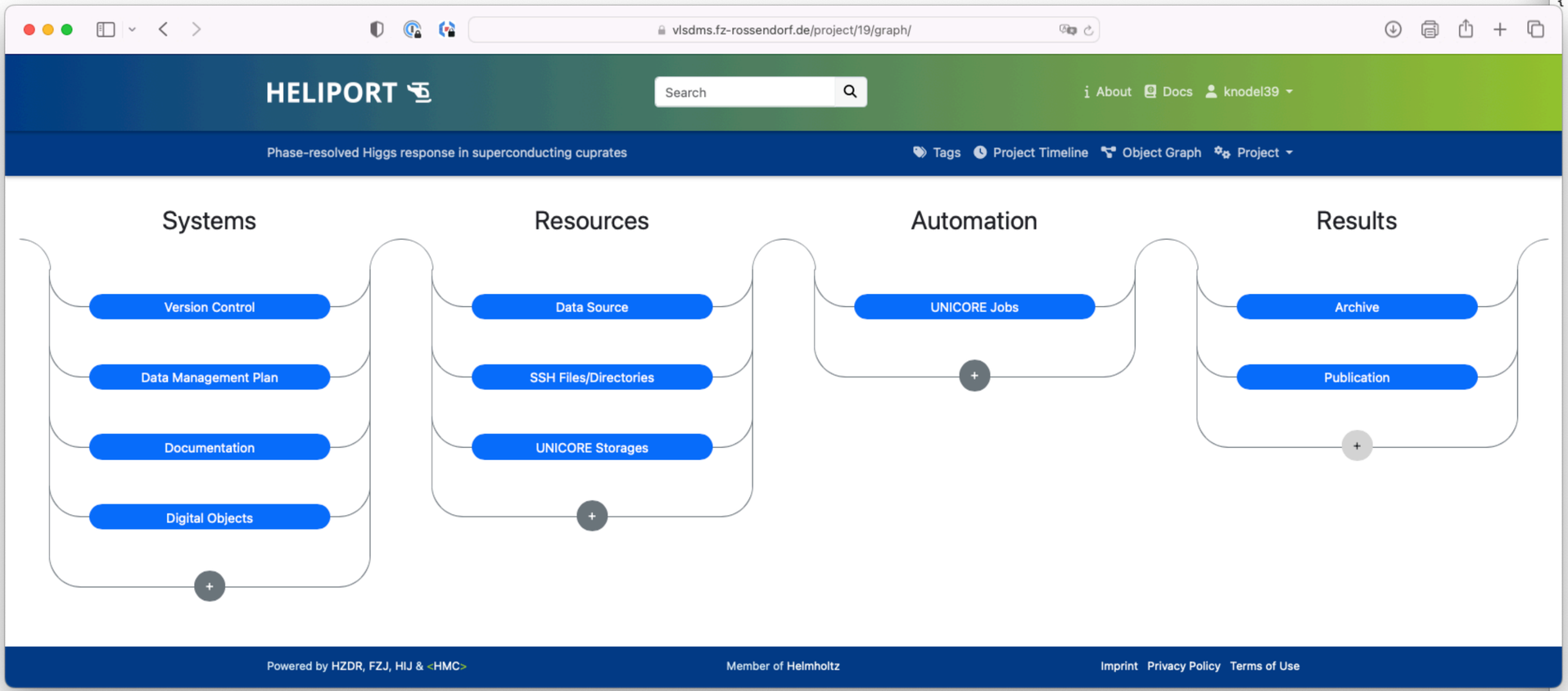
> The HELIPORT project aims at developing a platform which accommodates the **complete life cycle** of a scientific project and links all corresponding programs, systems and workflows to create a more **FAIR and comprehensible** project description.

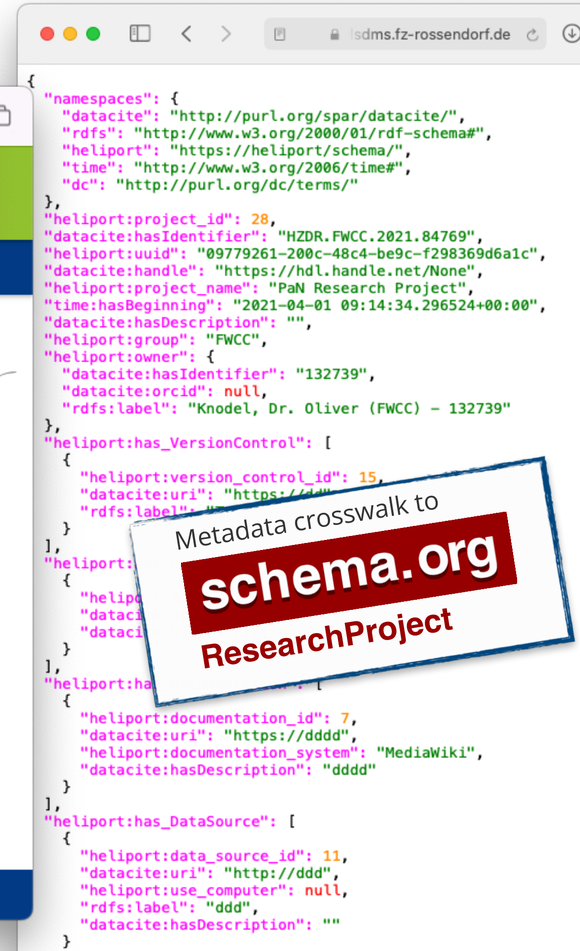
Project Members:

HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF
HI JENA Helmholtz Institute Jena
JÜLICH FORSCHUNGSZENTRUM

Funded by:

<HMC> HELMHOLTZ Metadata Collaboration





The Motivation to Develop HELIPORT

— HELIPORT was originally intended to provide only the **proposal's metadata**, to allow the assignment of resources.

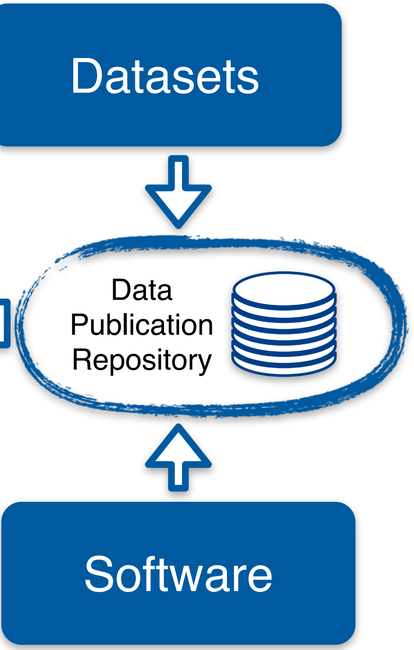
— Over time, we realised that HELIPORT can also answer our scientists' most important questions, such as:

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

How we can reduce computations and save energy?

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

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



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

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

HELIPORT JÜLICH FORSCHUNGSZENTRUM HI JENA HZDR

HELIPORT Features

- Entry point for experiments and scientific projects
- User and group authorisation/management
- Overview of systems and services involved in an experiment
- Provision of metadata from proposal systems (e.g. GATE)
- Registration of and access to internal file systems
- Automated transfer of metadata between involved systems/services
- Background data publication of datasets (e.g. Zenodo, Rodare)
- Integration of reproducible computational workflows
- HPC cluster access (slurm, UNICORE)
- Digital object and handle management with graph visualisation
- Timeline representing changes
- HELIPORT REST API
- Authentication via Helmholtz ID

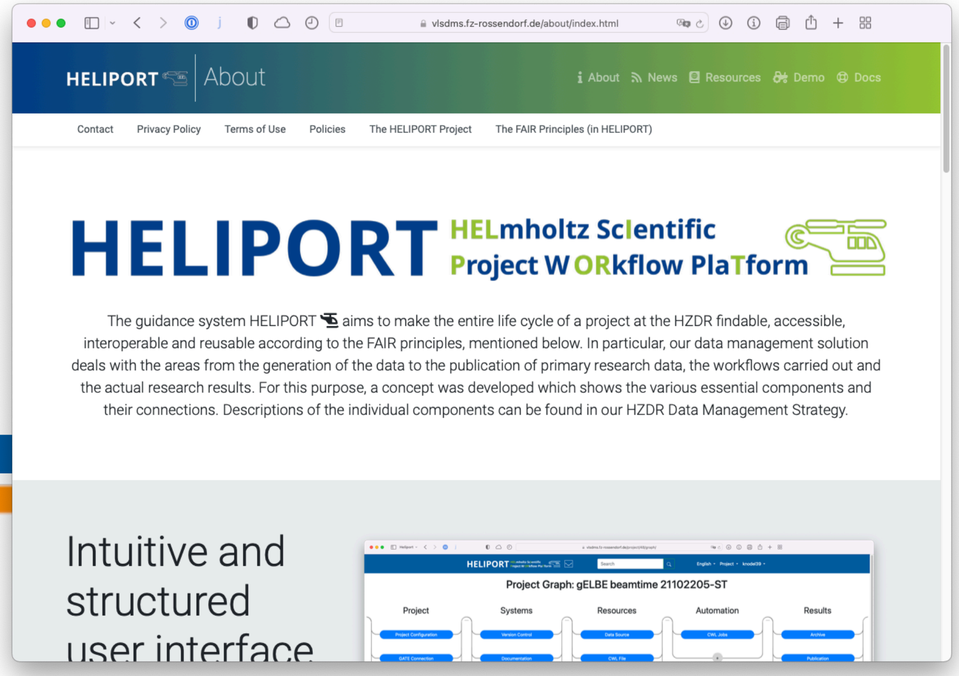
HELMHOLTZ ID

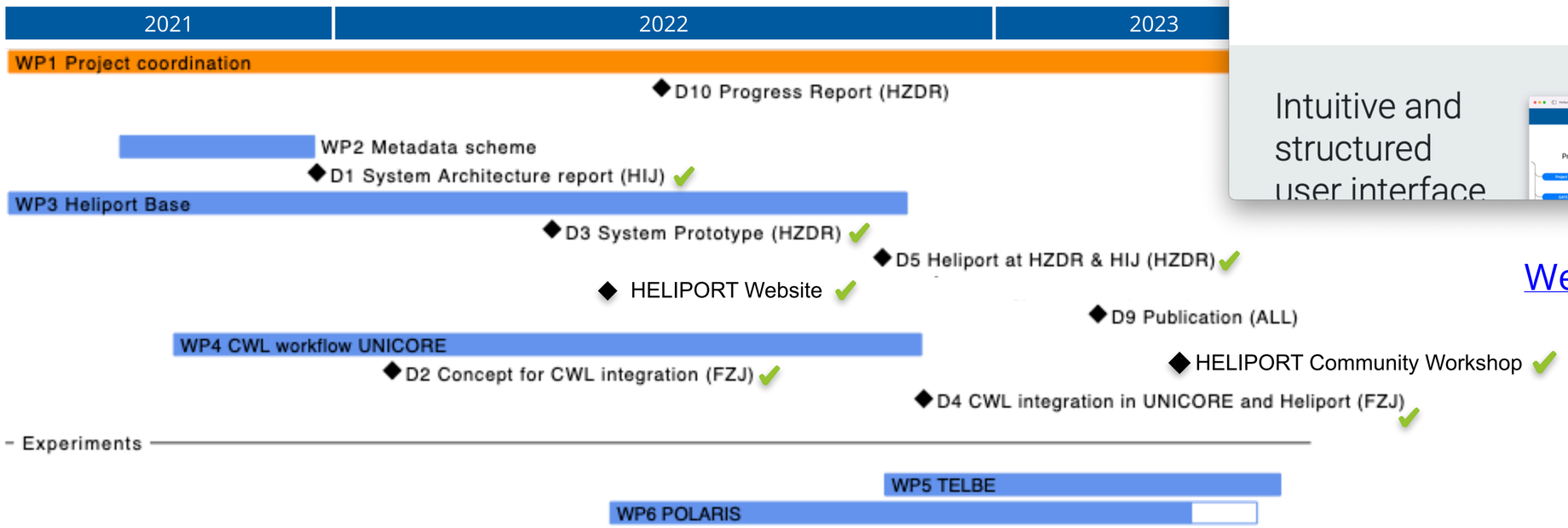
HELIPORT

JÜLICH FORSCHUNGSZENTRUM HI JENA HZDR

Timescale for the HMC-Project HELIPORT

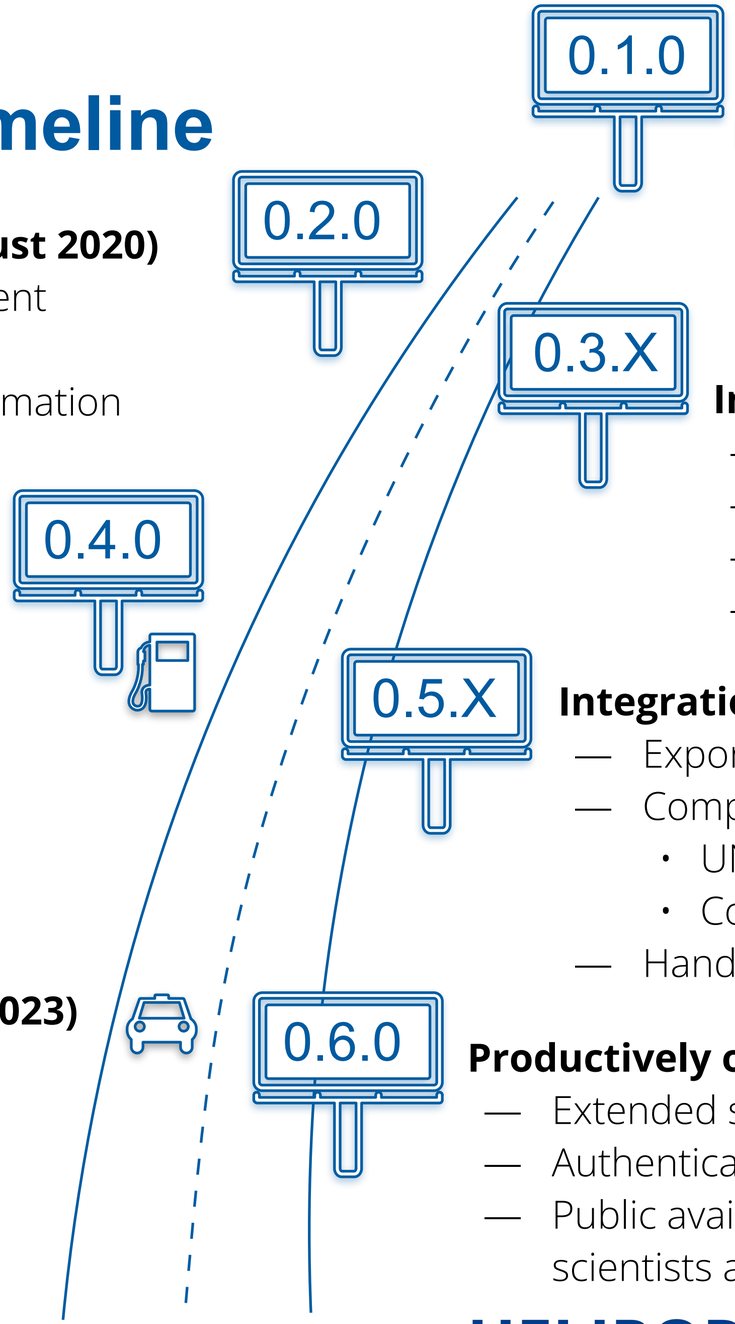
- The deliverables and our prototype are available on our website.
- We are in contact with different Helmholtz centers, universities and European partners and build a HELIPORT community.
- Overview of work packages and milestones:





[Website: heliport.hzdr.de](heliport.hzdr.de)

Heliport (Project) Timeline



Initial Version (June 2020)
- DMS Projects and proposal information from the **HZDR GATE proposal** database
- Webinterface with user authentication (LDAP)

First Draft: Project Plan (August 2020)
- Project and user management
- Configurable stages
- **REST API** for proposal information
- CWL visualization prototype

Improved Project Plan (December 2020)
- Configurable stages and modules
- Infrastructure and database updates
- Daily proposal database update
- Advanced logging and monitoring

Modular Structure (July 2021)
- Official start of the HMC founded Heliport project:

  HMC | HELMHOLTZ METADATA COLLABORATION
- Redesign to provide modular and highly configurable system

Integration of various Apps and Features
- Export for (different) metadata schemas
- Computational/scientific workflow execution
  - UNICORE support ()
  - Computing job management and monitoring
- Handle management with public landing pages

HELIPORT Community Workshop (July 2023)

GFZ Helmholtz-Zentrum Potsdam · HZDR Helmholtz Zentrum Dresden Rossendorf · Jülich Forschungszentrum · GSI · HI JENA

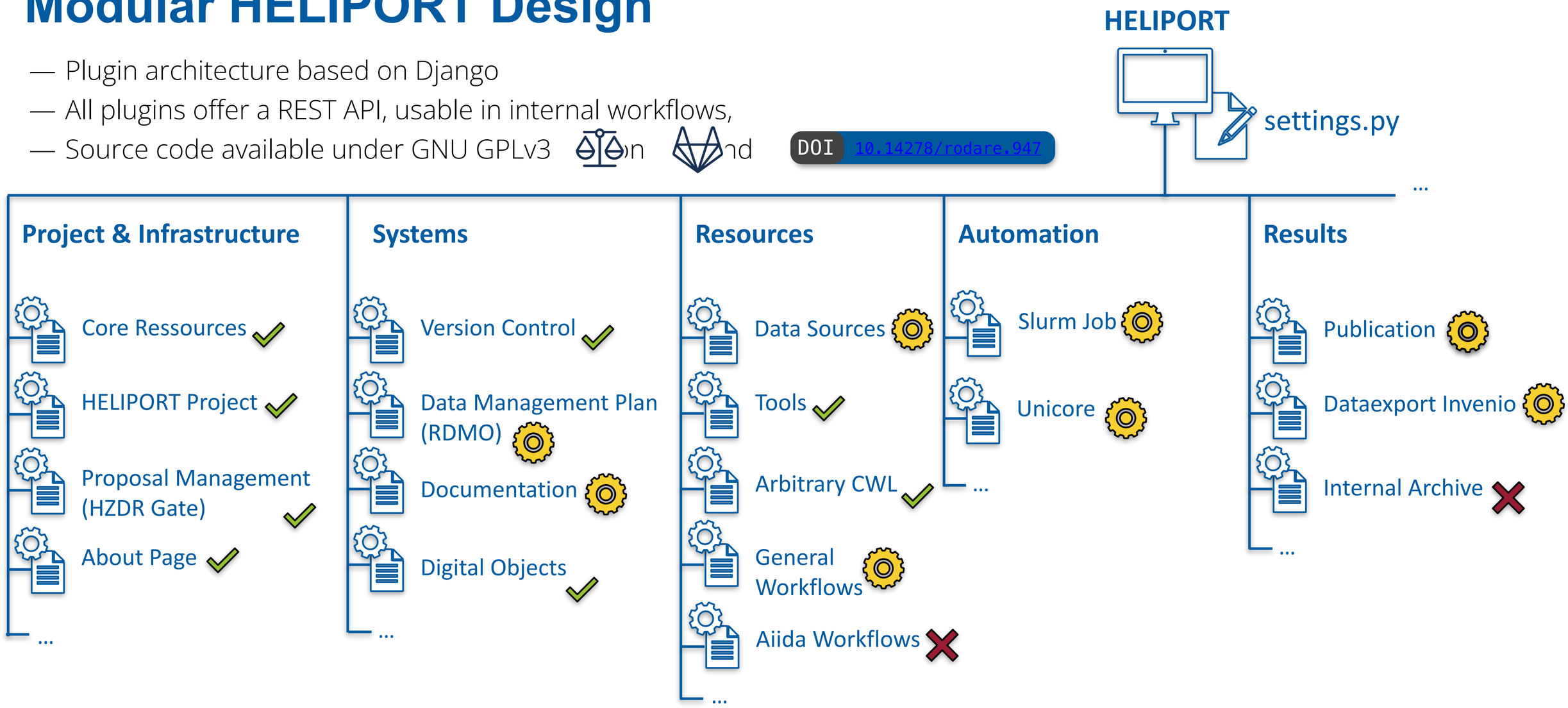
Productively operating HELIPORT for different RIs
- Extended support for a proposal system (GATE)
- Authentication with OpenID Connect (Helmholtz ID)
- Public available HELIPORT instance for remote/visiting scientists at HZDR

HELIPORT · Jülich Forschungszentrum · HI JENA · HZDR

Modular HELIPORT Design

— Plugin architecture based on Django
— All plugins offer a REST API, usable in internal workflows,
— Source code available under GNU GPLv3 on GitLab

DOI 10.14278/rodare.947

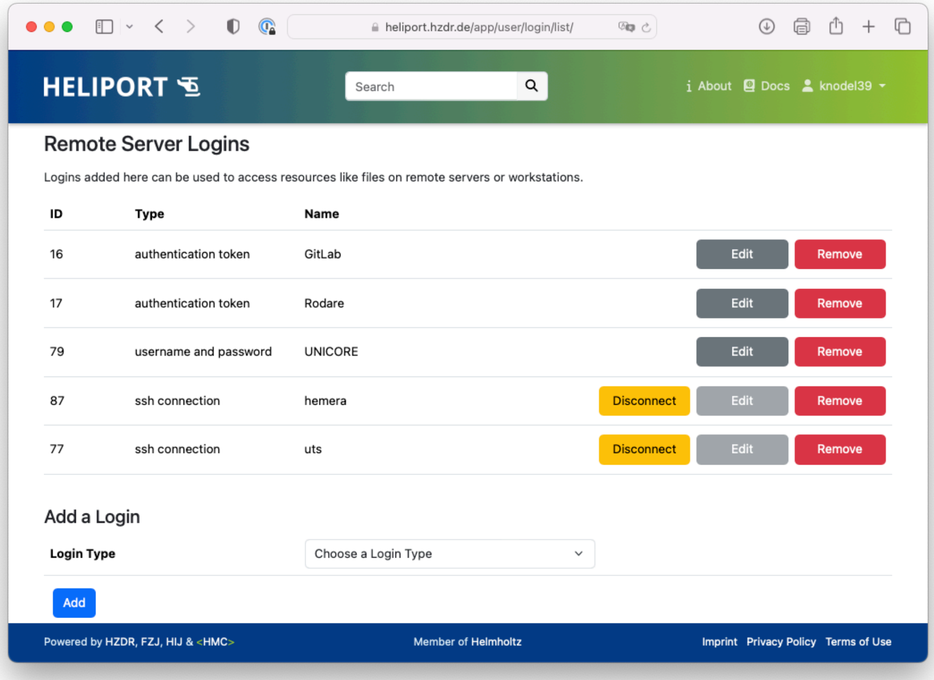


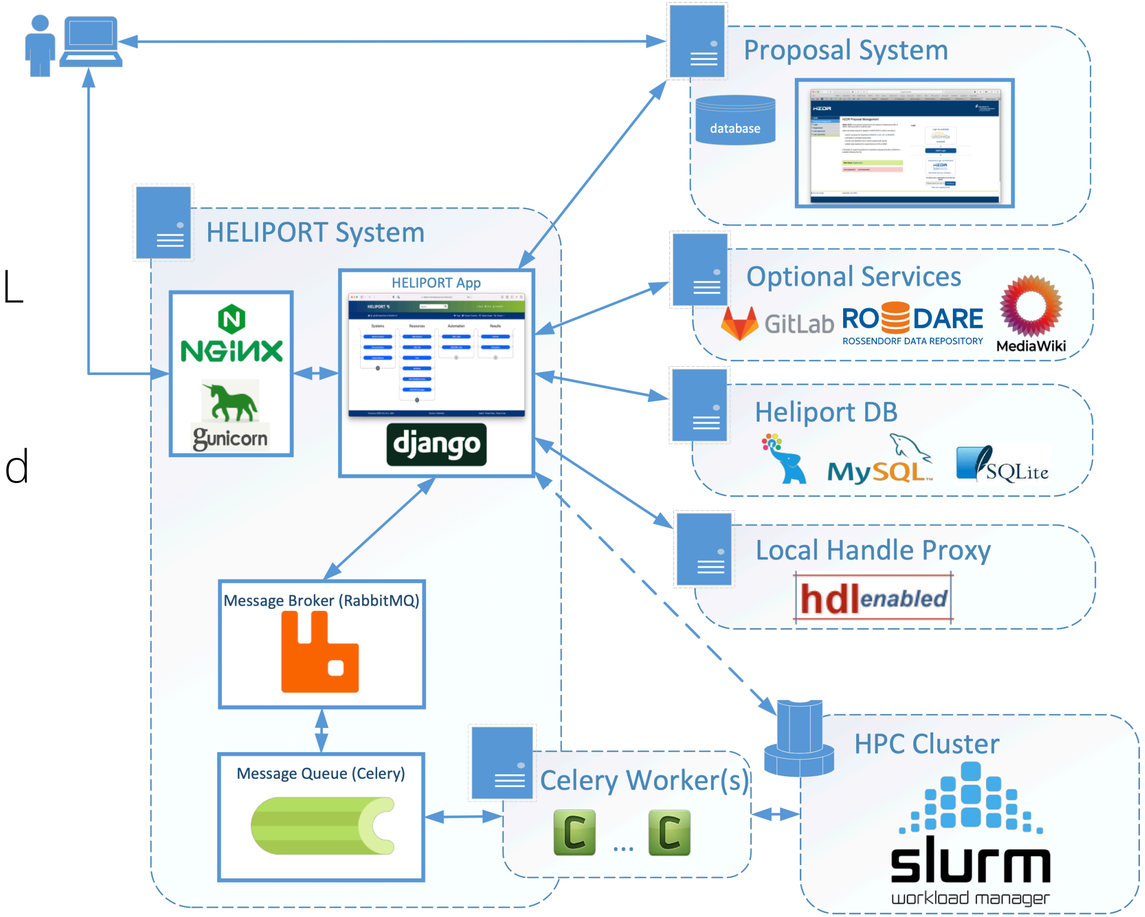
HELIPORT
JÜLICH FORSCHUNGSZENTRUM HI JENA HZDR

HELIPORT Infrastructure

— The HELIPORT web app is based on Django:
  - Heliport communicates with various system through REST APIs,
  - The project-level metadata is stored in a PostgreSQL database and can be exported in various metadata schemes.

— Computational workflows are managed in HELIPORT and executed on HPC clusters using slurm or UNICORE.



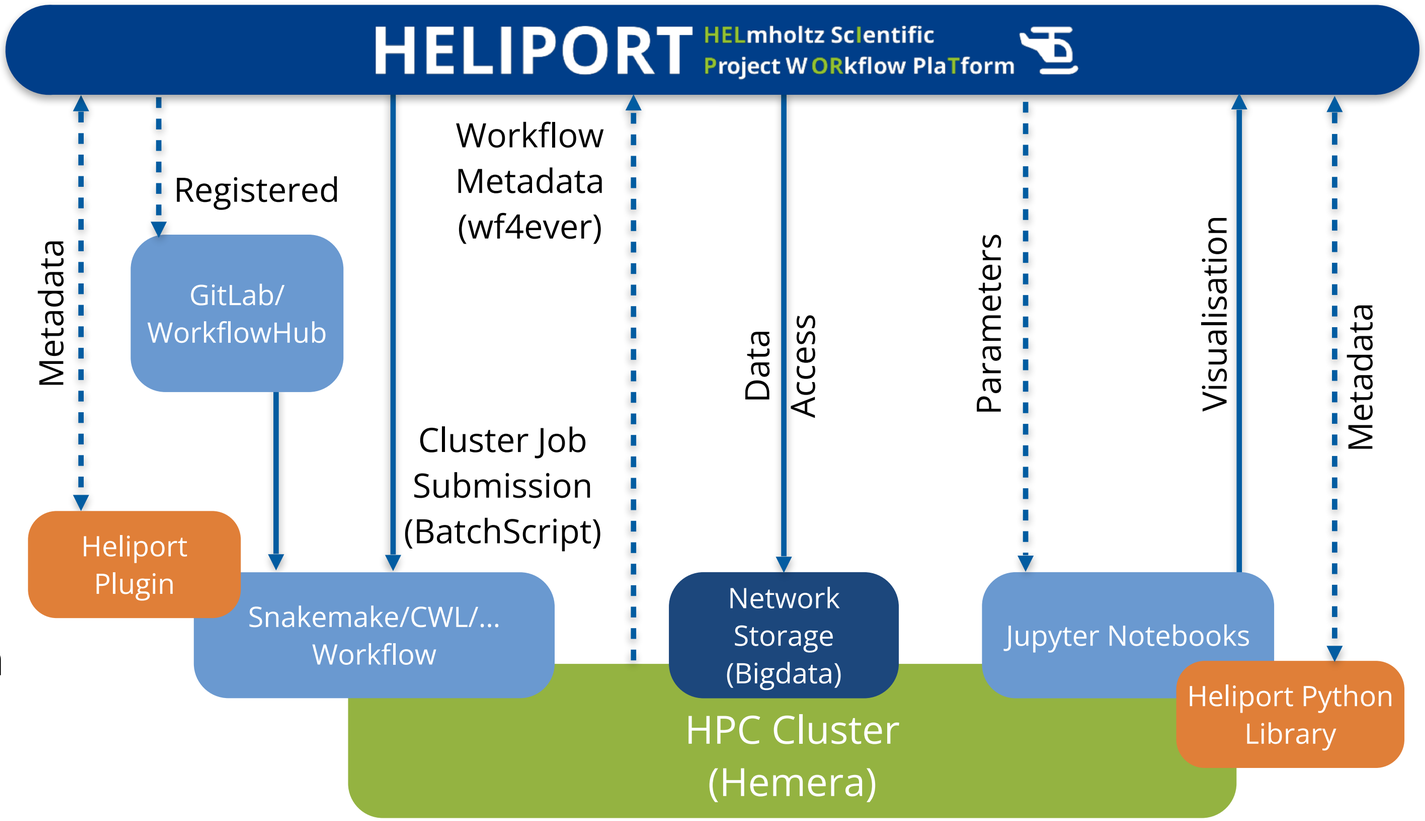


HELIPORT

JÜLICH FORSCHUNGSZENTRUM HI JENA HZDR

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**

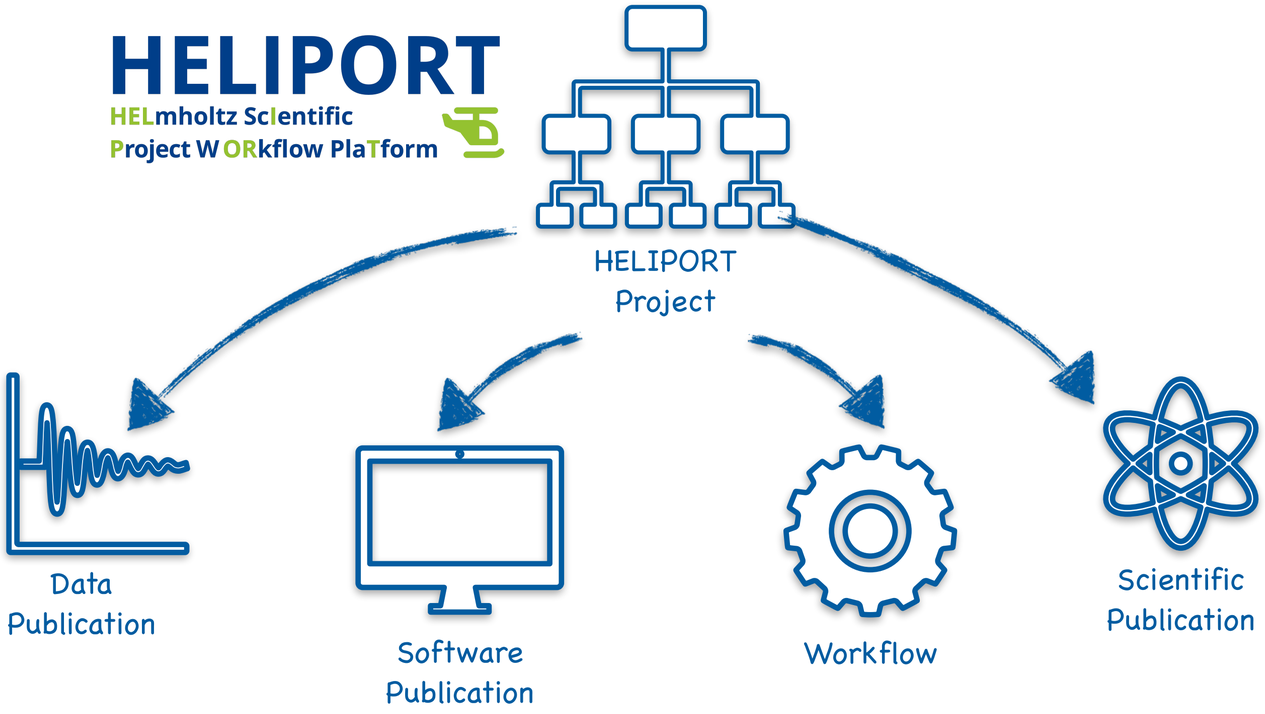


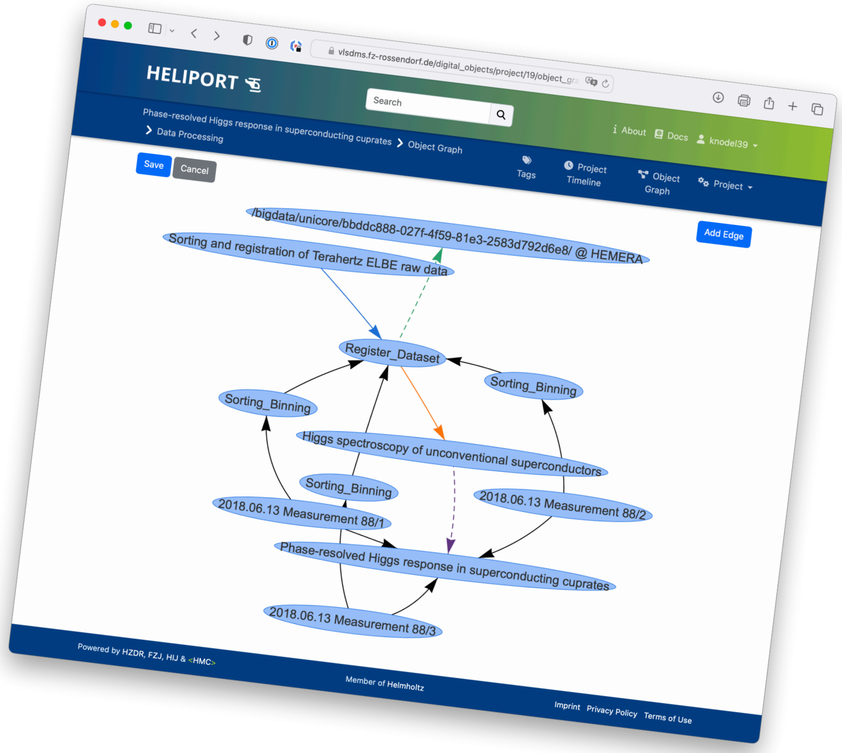
Conclusions

— HELIPORT describes and collects all metadata from all services and systems involved in an scientific experiment.

— Such an approach is desirable and leads us to a fully **FAIR** and **comprehensible** research project.

— The computational workflows are essential to keep track of everything what happened during the experiment.

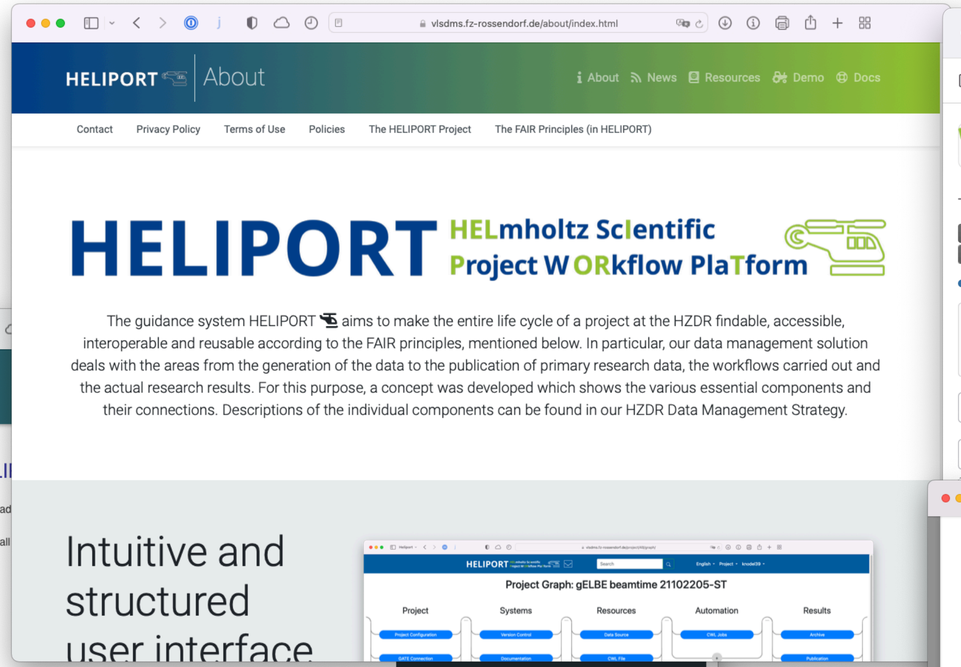


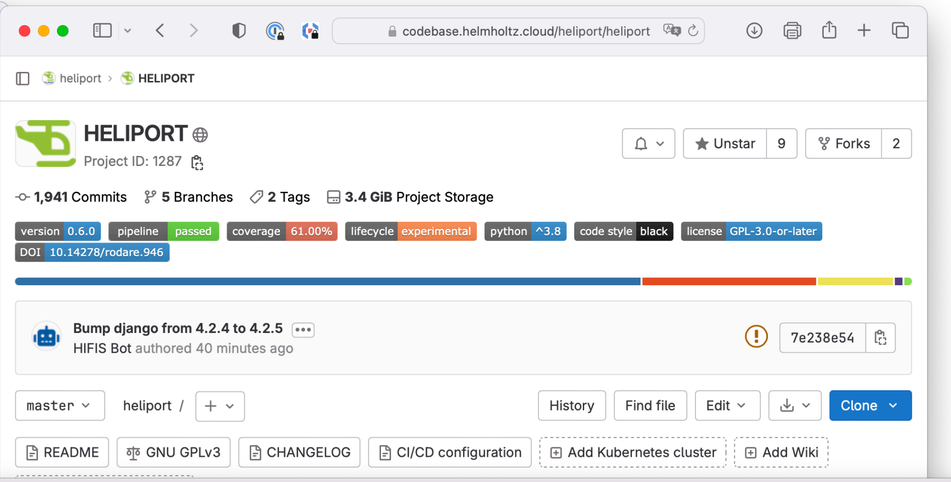


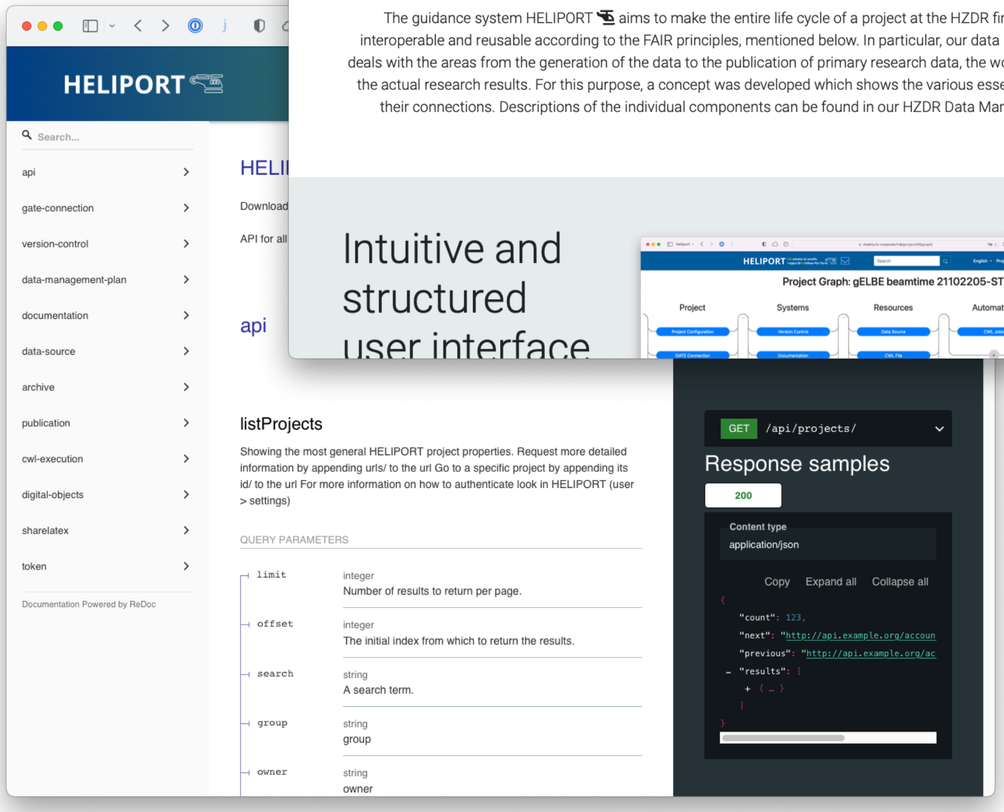
Resources

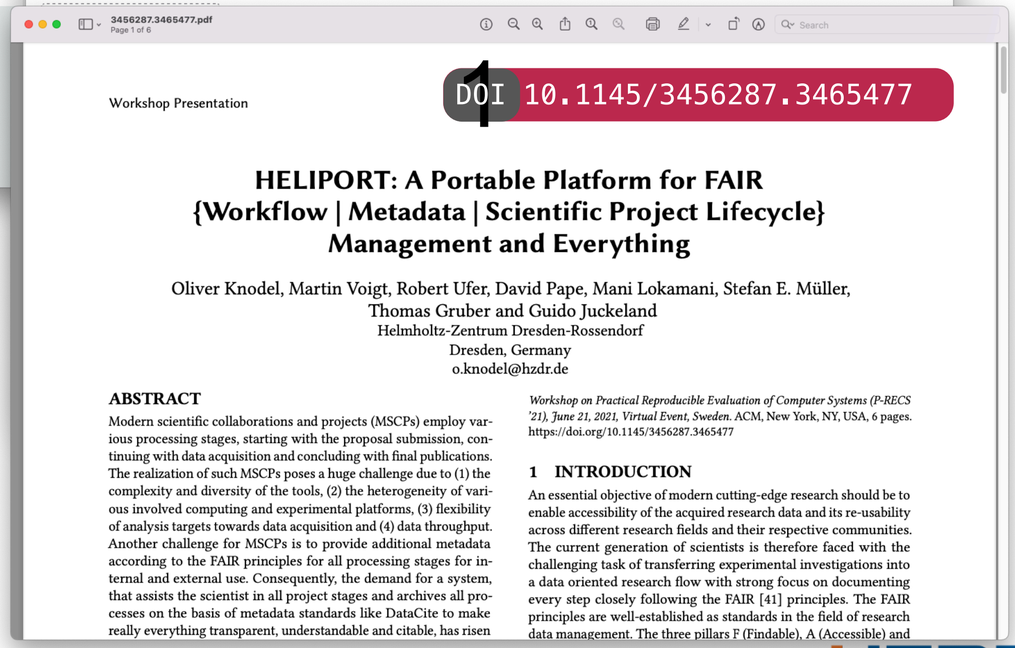
Website: heliport.hzdr.de

Repository: codebase.helmholtz.cloud/heliport









API Doc: heliport.hzdr.de/redoc/

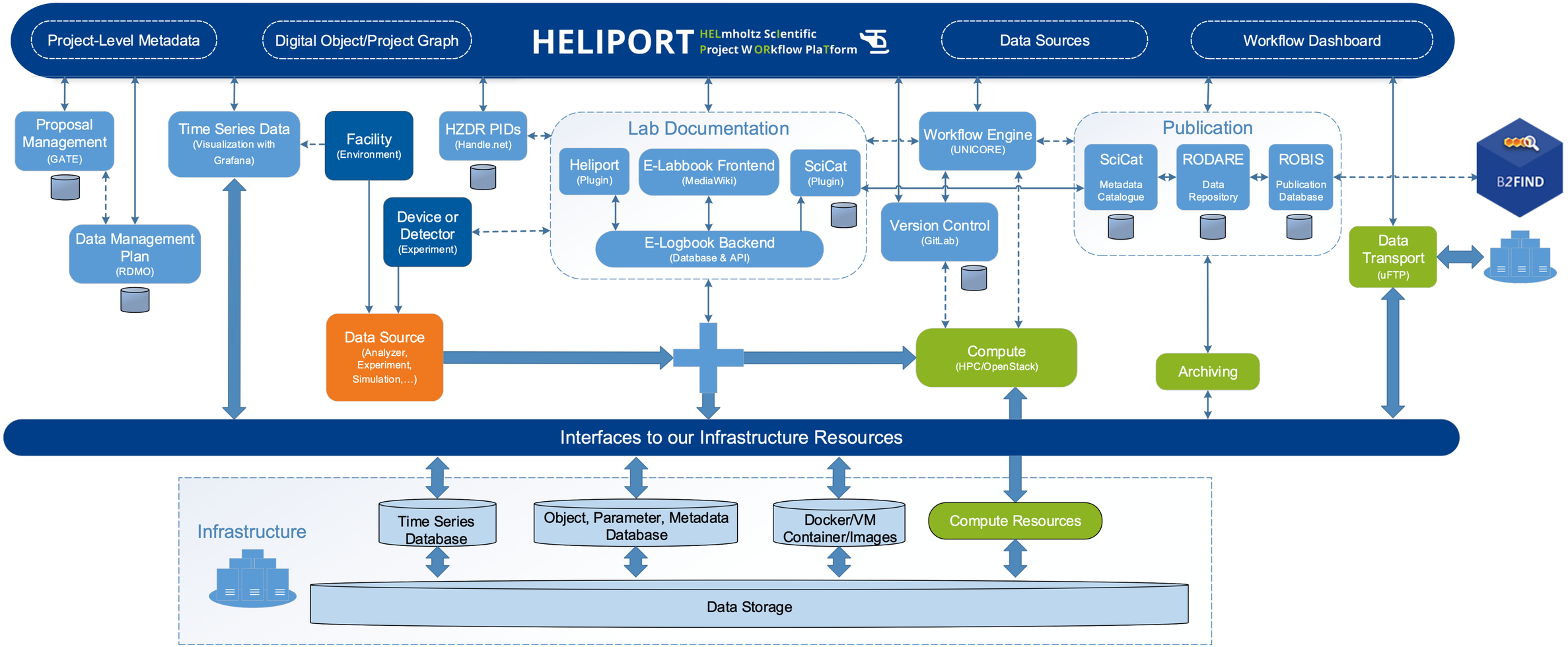
HELIPORT

JÜLICH FORSCHUNGSZENTRUM
HI JENA Helmholtz Institute Jena
HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

Appendix

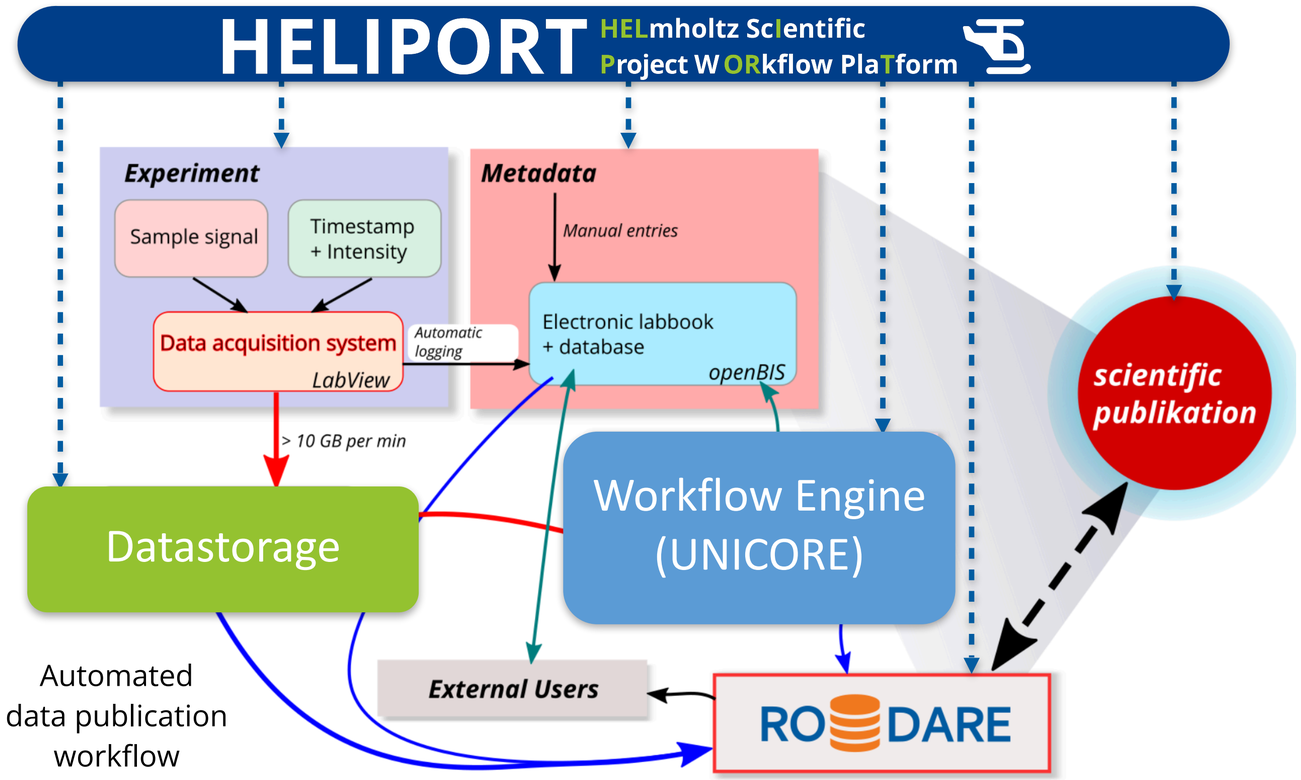
DRESDEN concept

HZDR Research (Infrastructure) Landscape



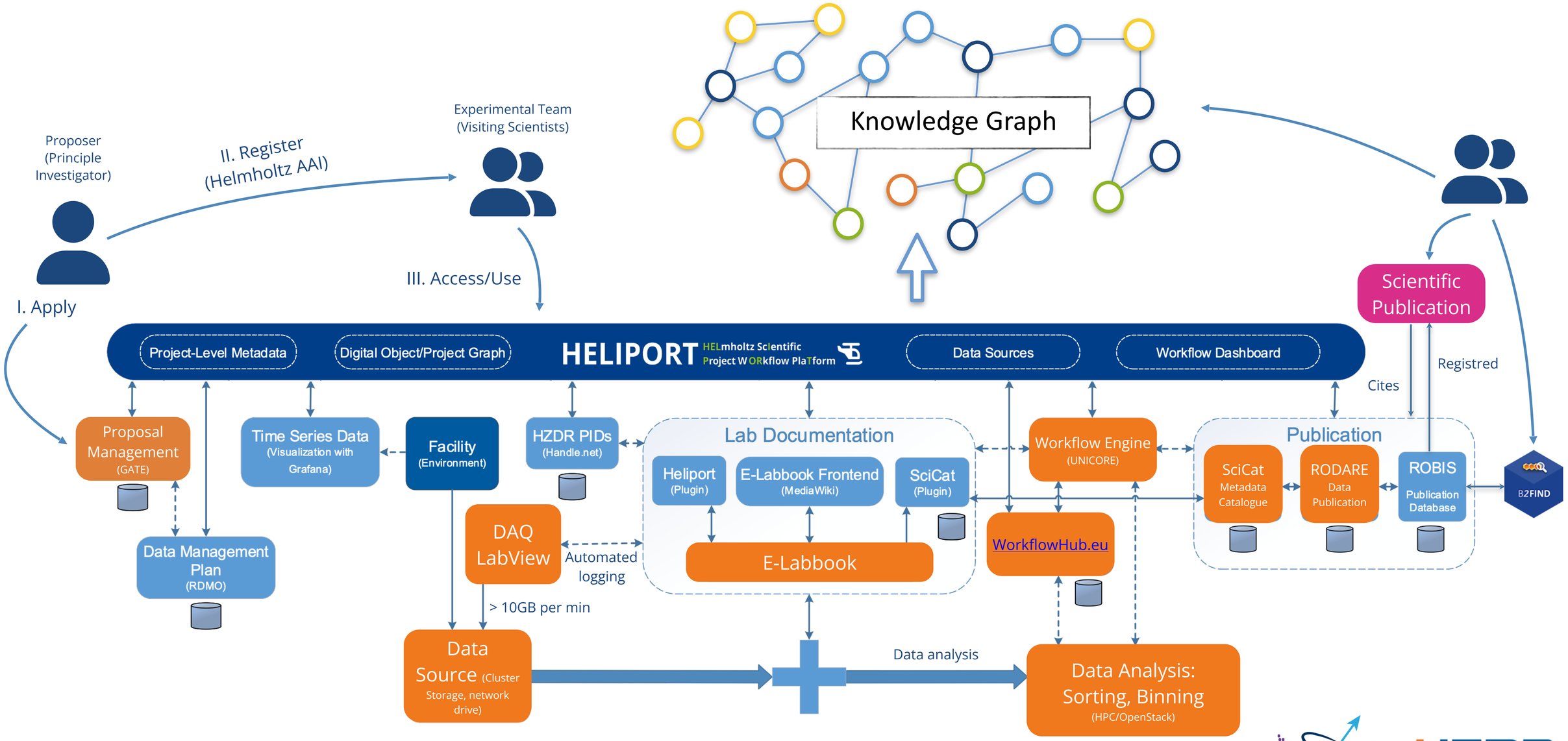
TELBE Data Flow

- 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



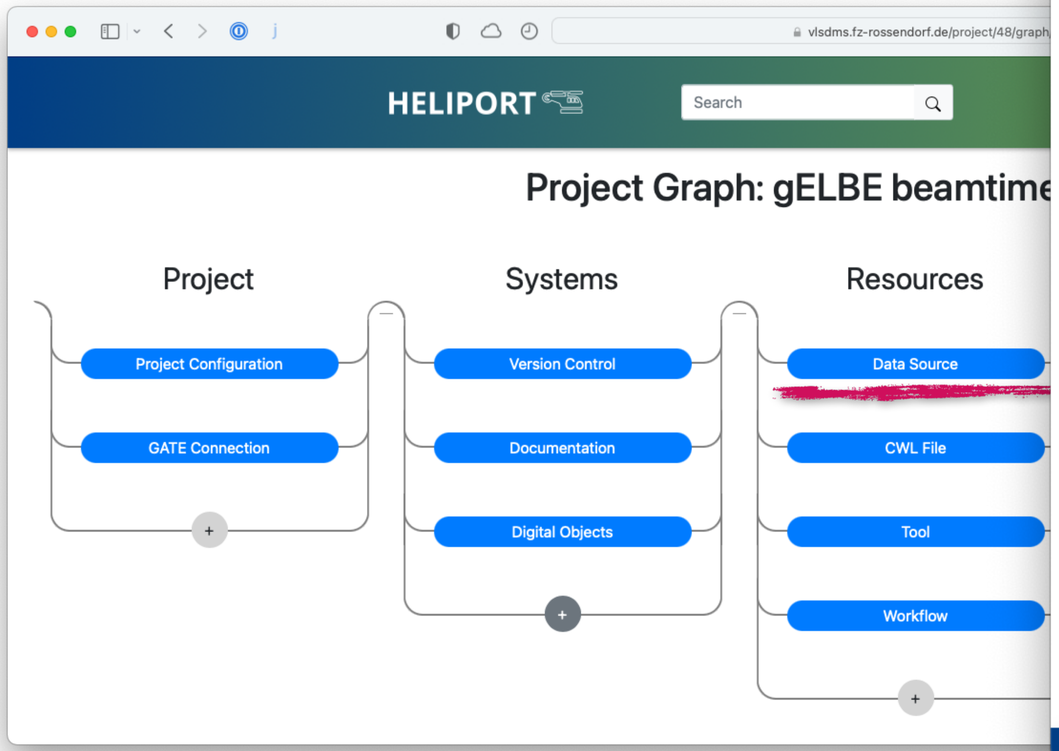
Deinert, Jan-Christophi. (2021, November 18). TELBE data analysis workflow and the PaN training platform UX — Booklet of presentations from the PaN EOSC Symposium 2021. Zenodo. https://doi.org/10.5281/zenodo.5636331

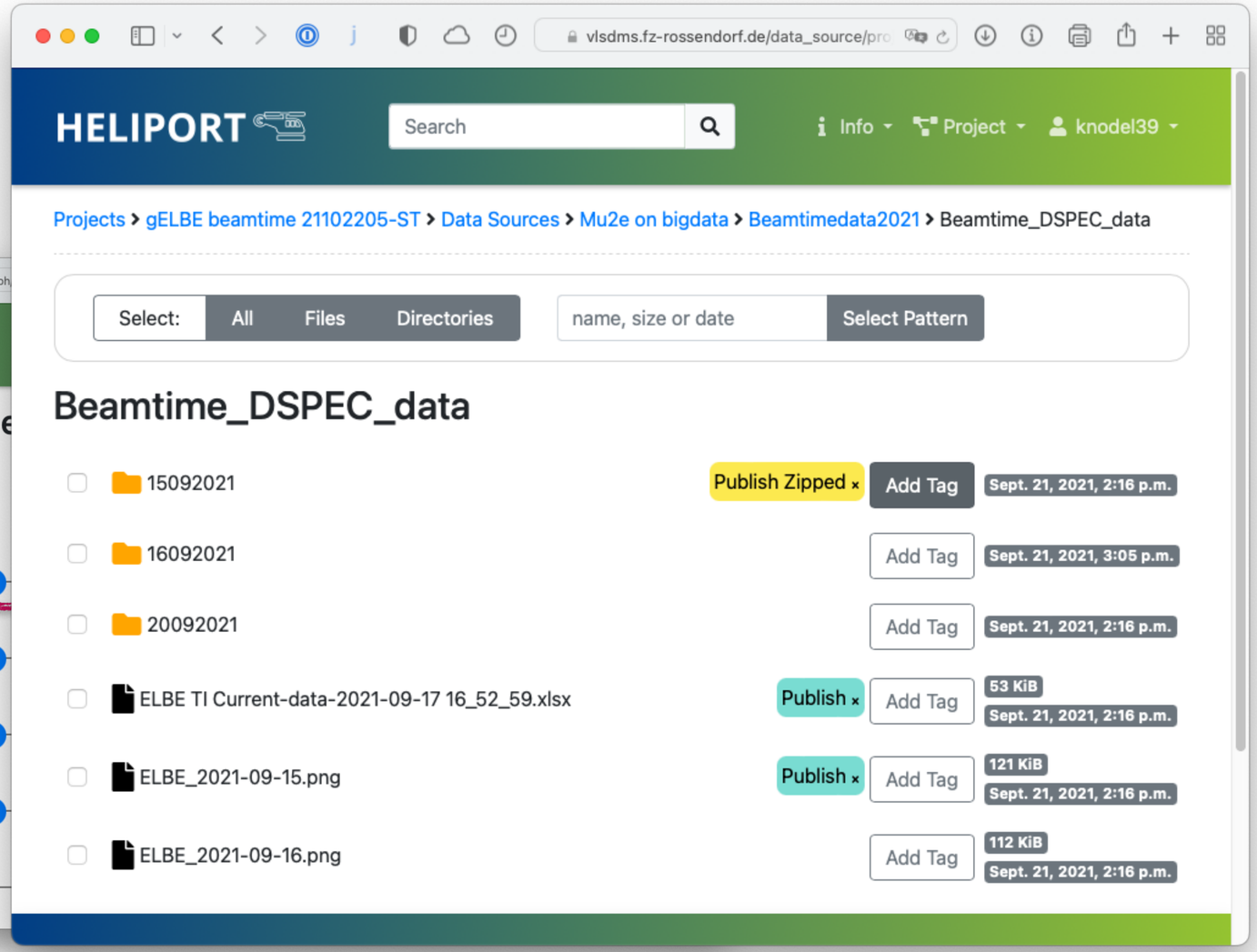
Mapping of the TELBE Resources to HELIPORT



Data Sources can be Registered…

…and files (selected for publication) can be transferred directly to the data publication.



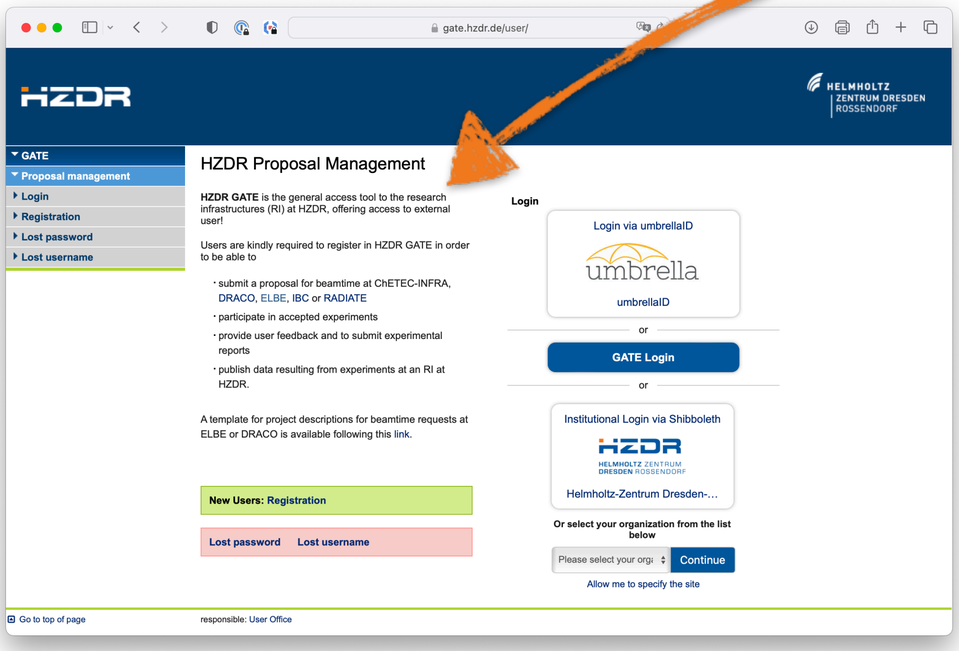


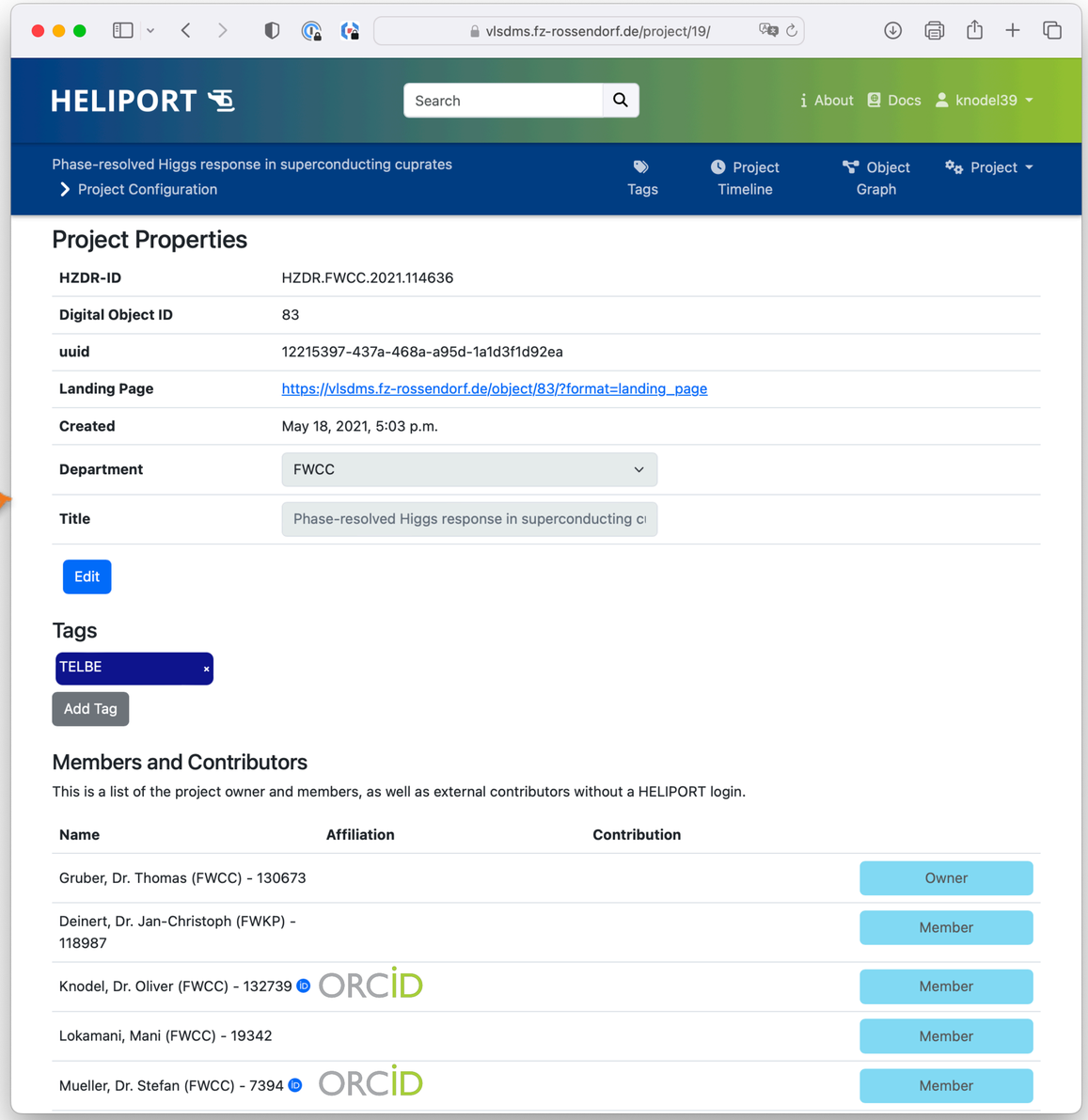
HELIPORT JÜLICH HI JENA HZDR
FORSCHUNGSZENTRUM

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)

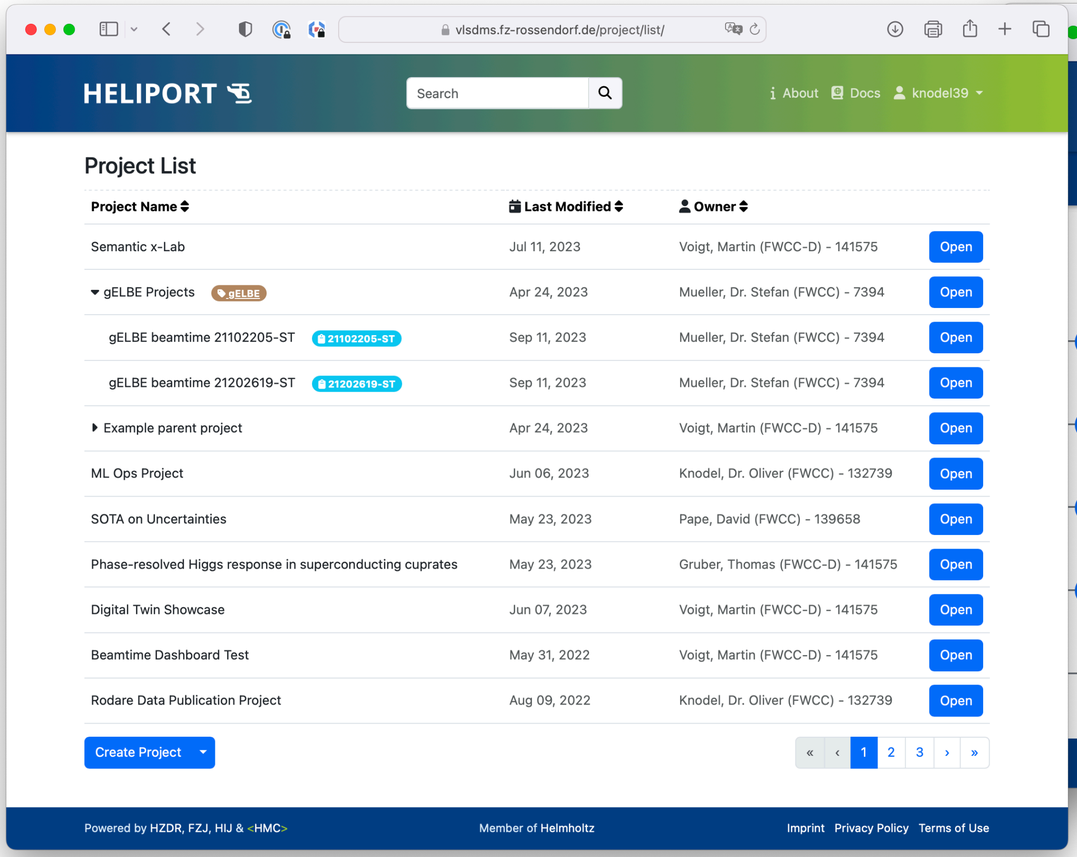


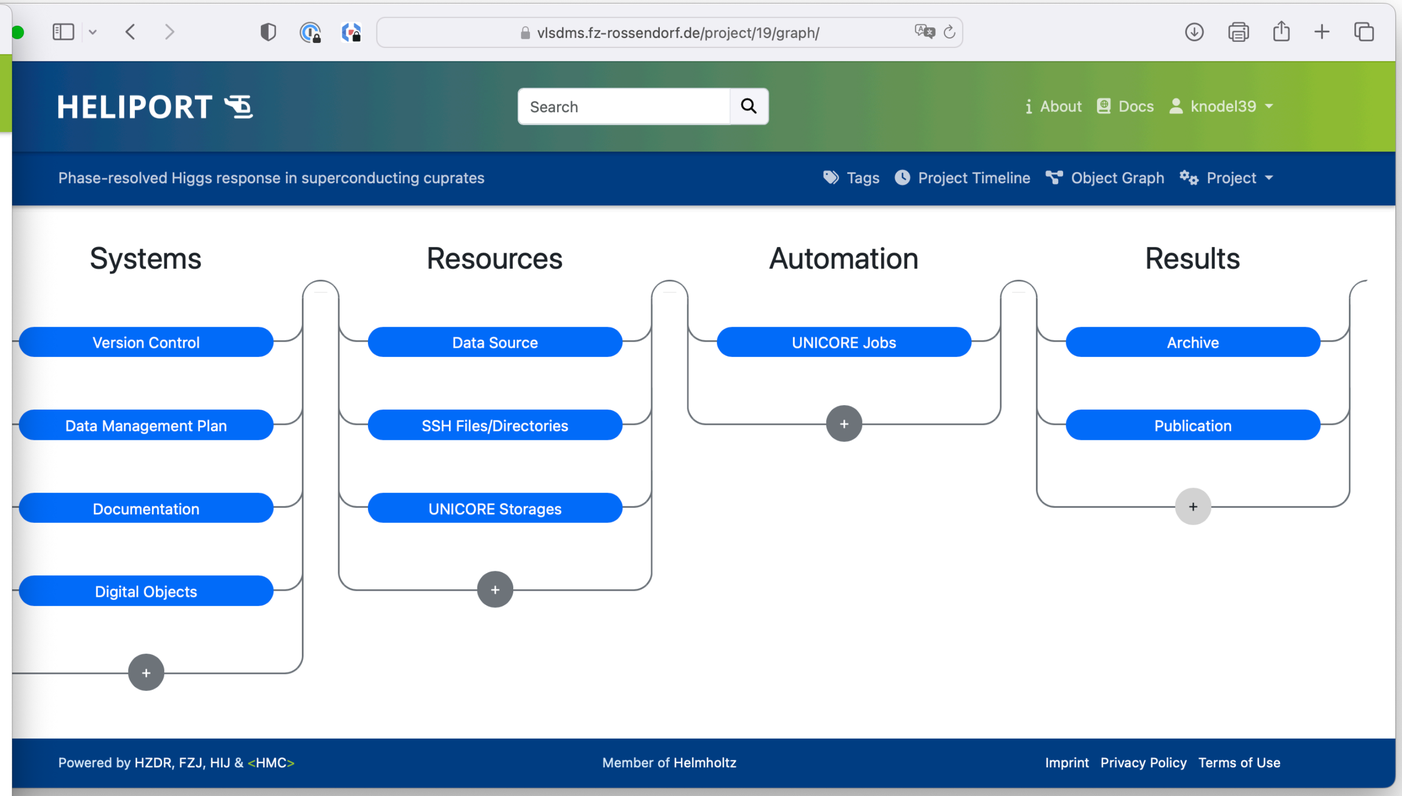


II. Project List and Dashboard

— 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.

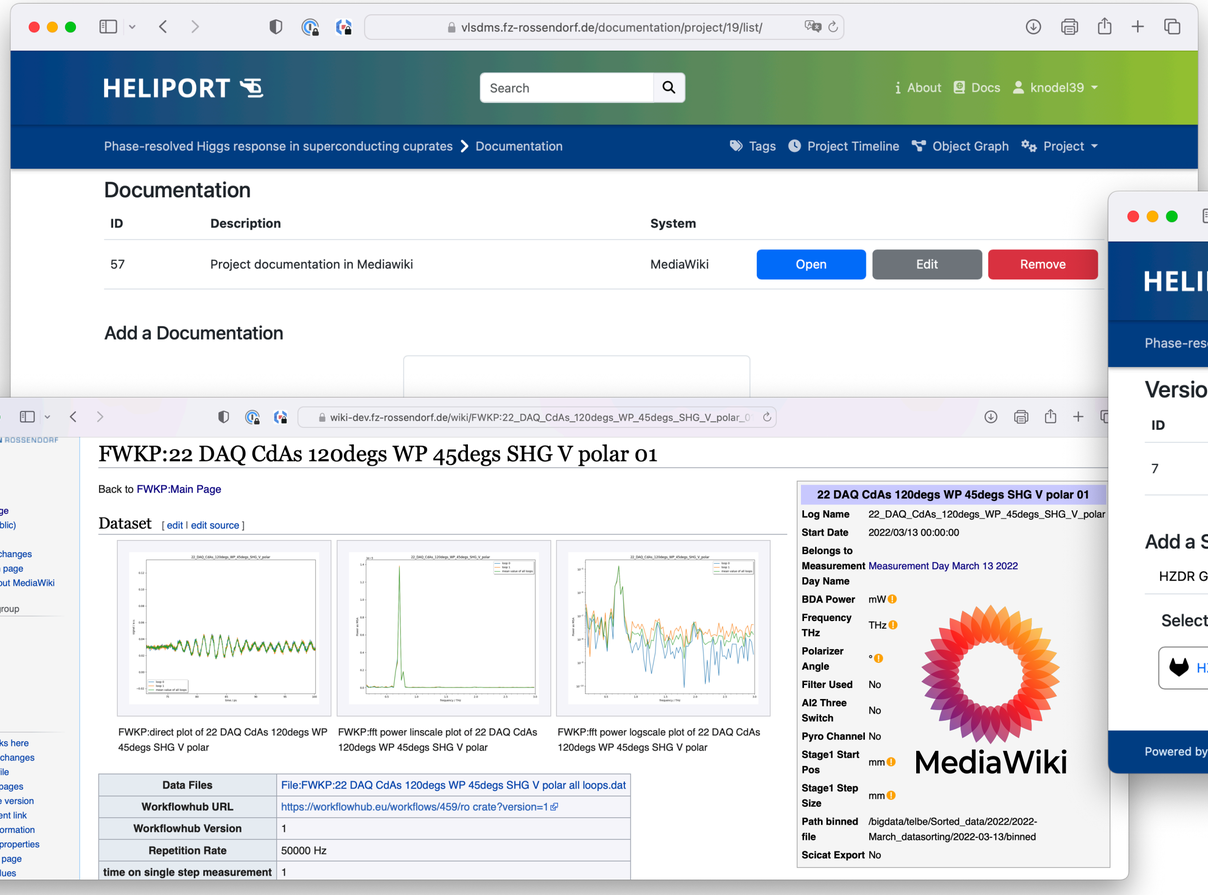


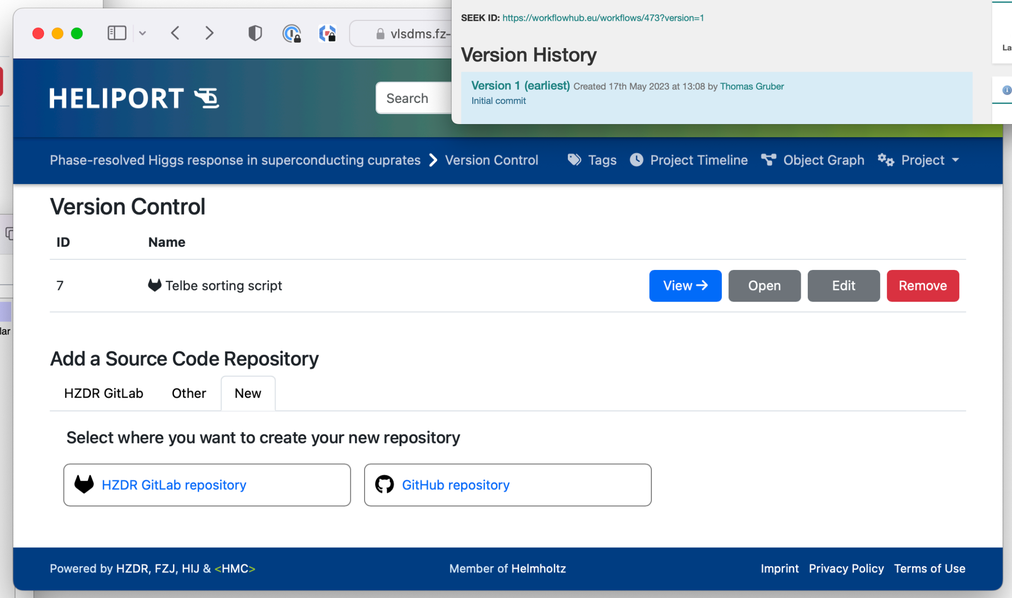


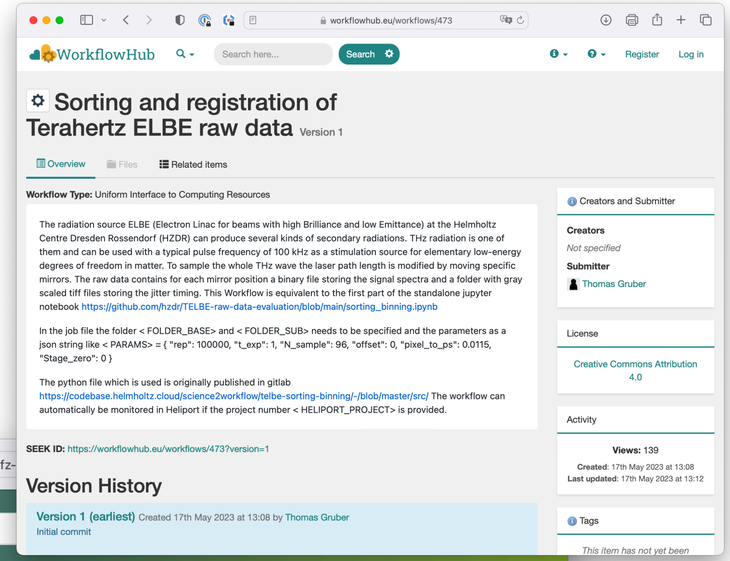
III. Resources: Documentation and Repositories

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

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

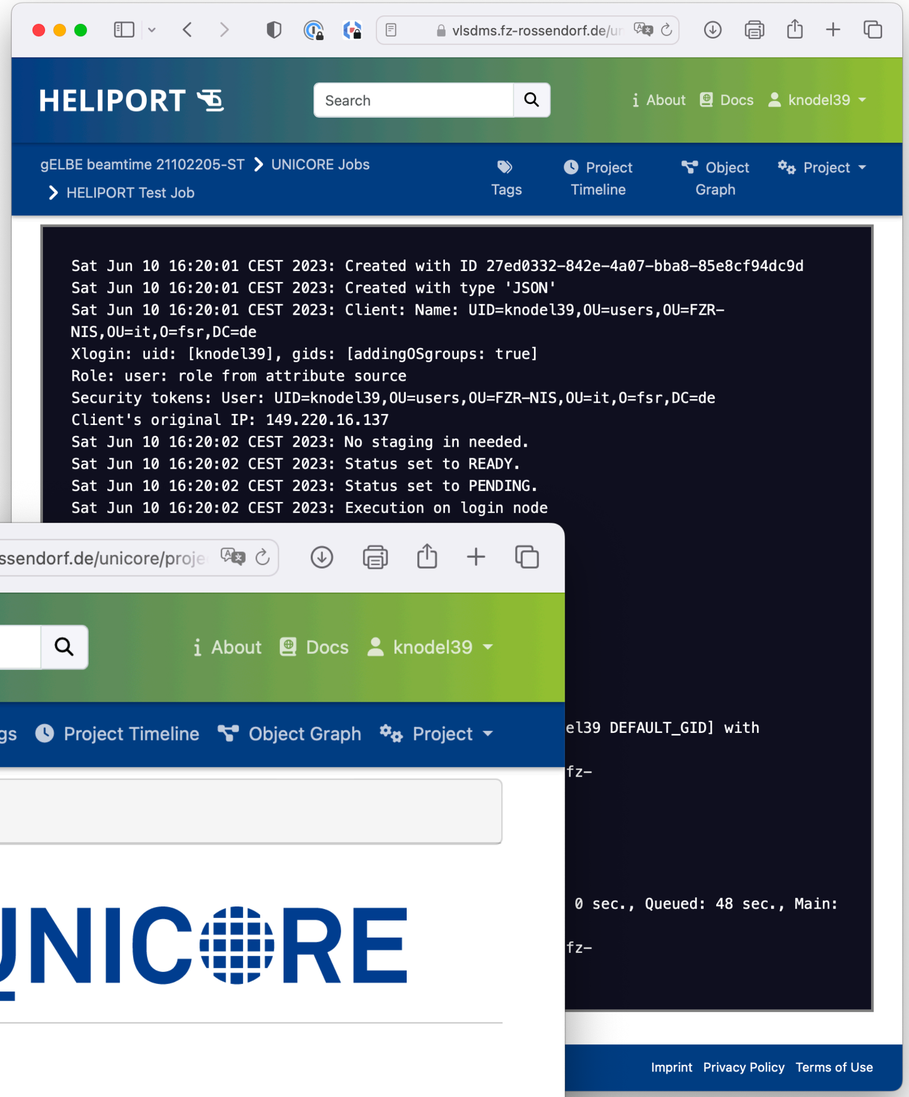


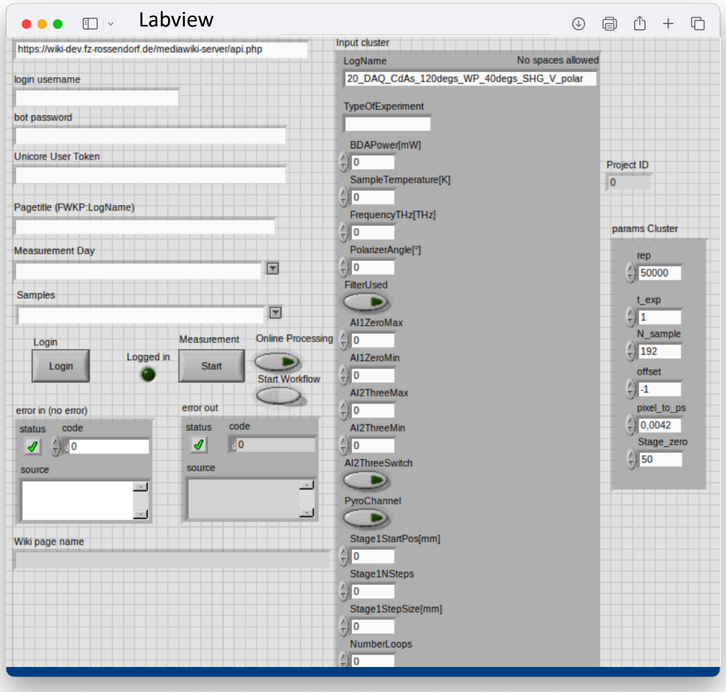


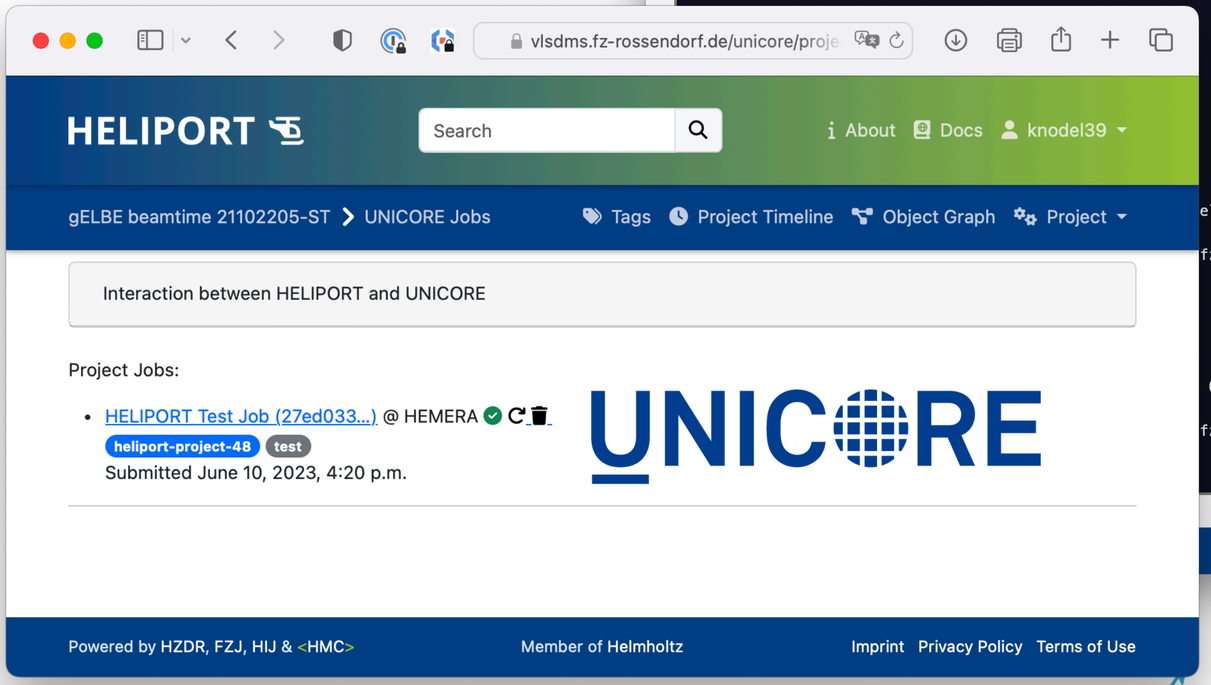


IV. Detector Control and Workflows

- The HELIPORT REST-API enables the transfer of metadata between HELIPORT and external systems (e.g. detector control in LabView).
- The integrated workflow management system (e.g. UNICORE) provides metadata for the provenance information required by HELIPORT.
- Workflows (on our HPC cluster) can be accessed by any project member directly in the HELIPORT web frontend.





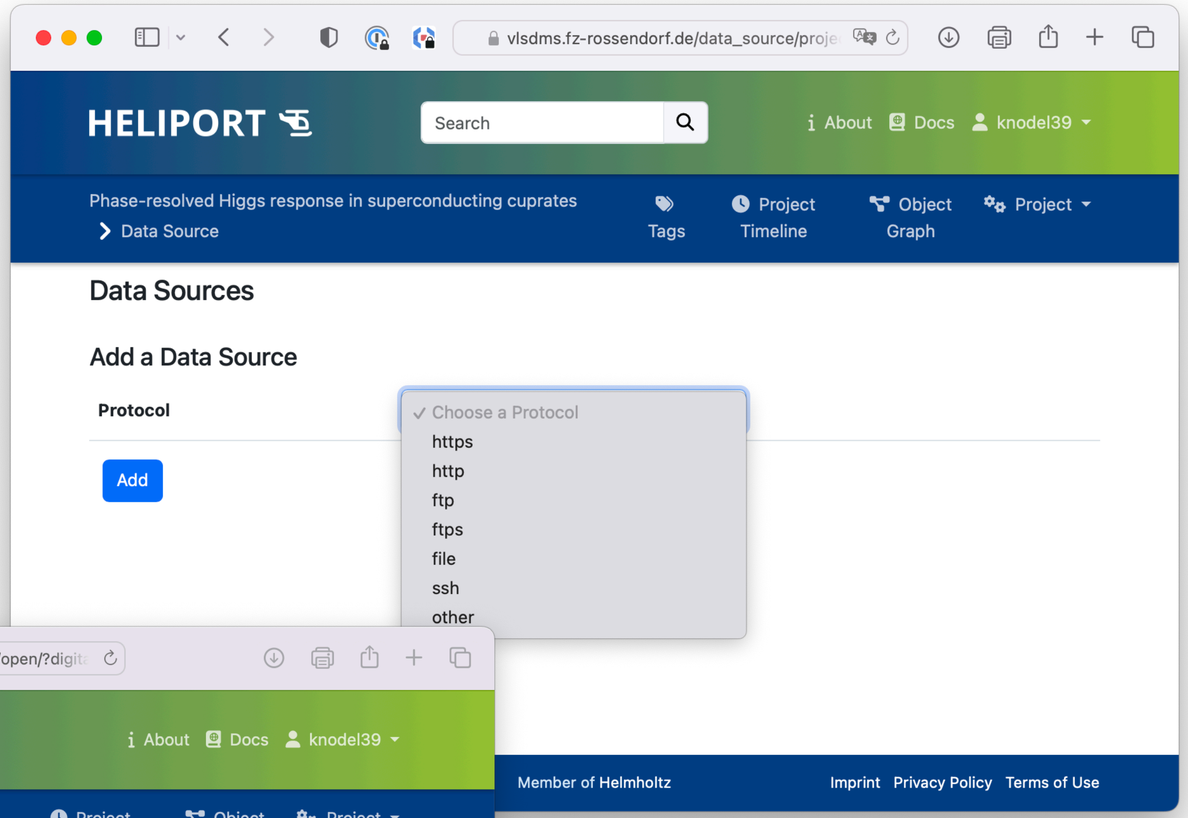


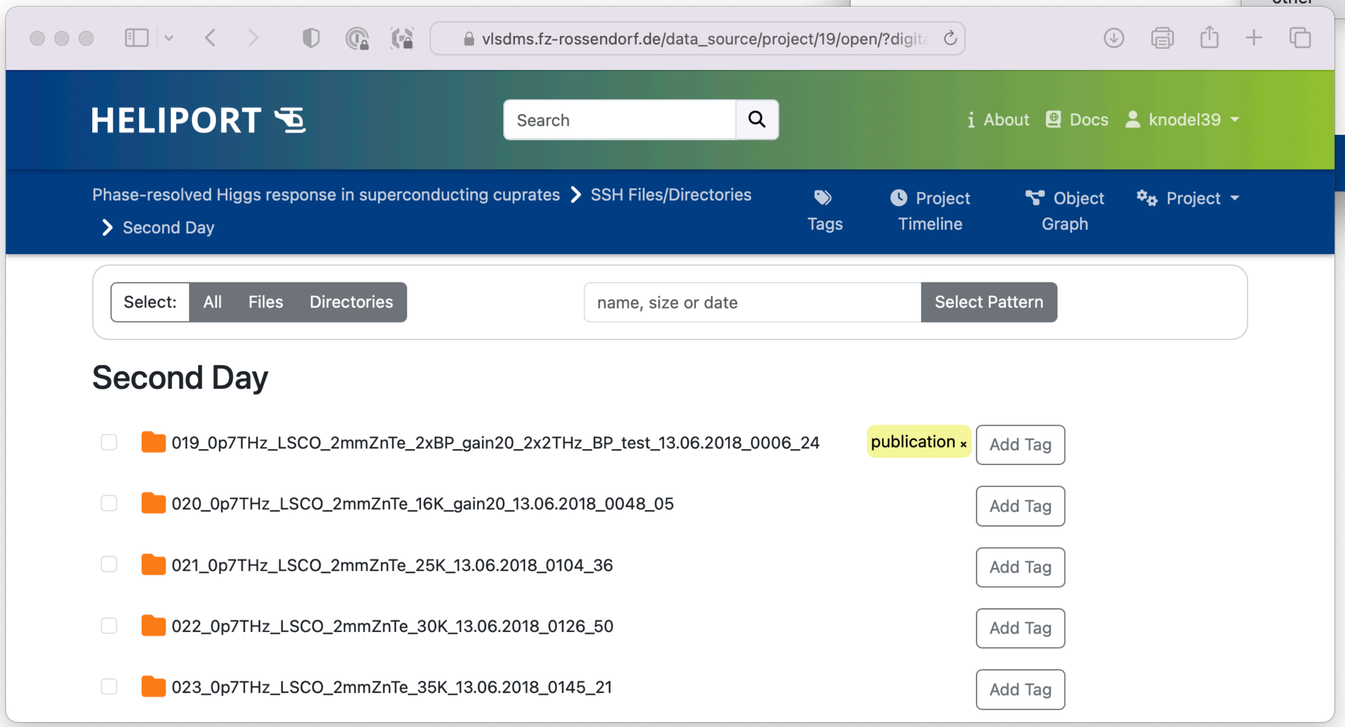
V. Data Sources

— Folders and Files in our internal filesystem can be registered in HELIPORT as data source.

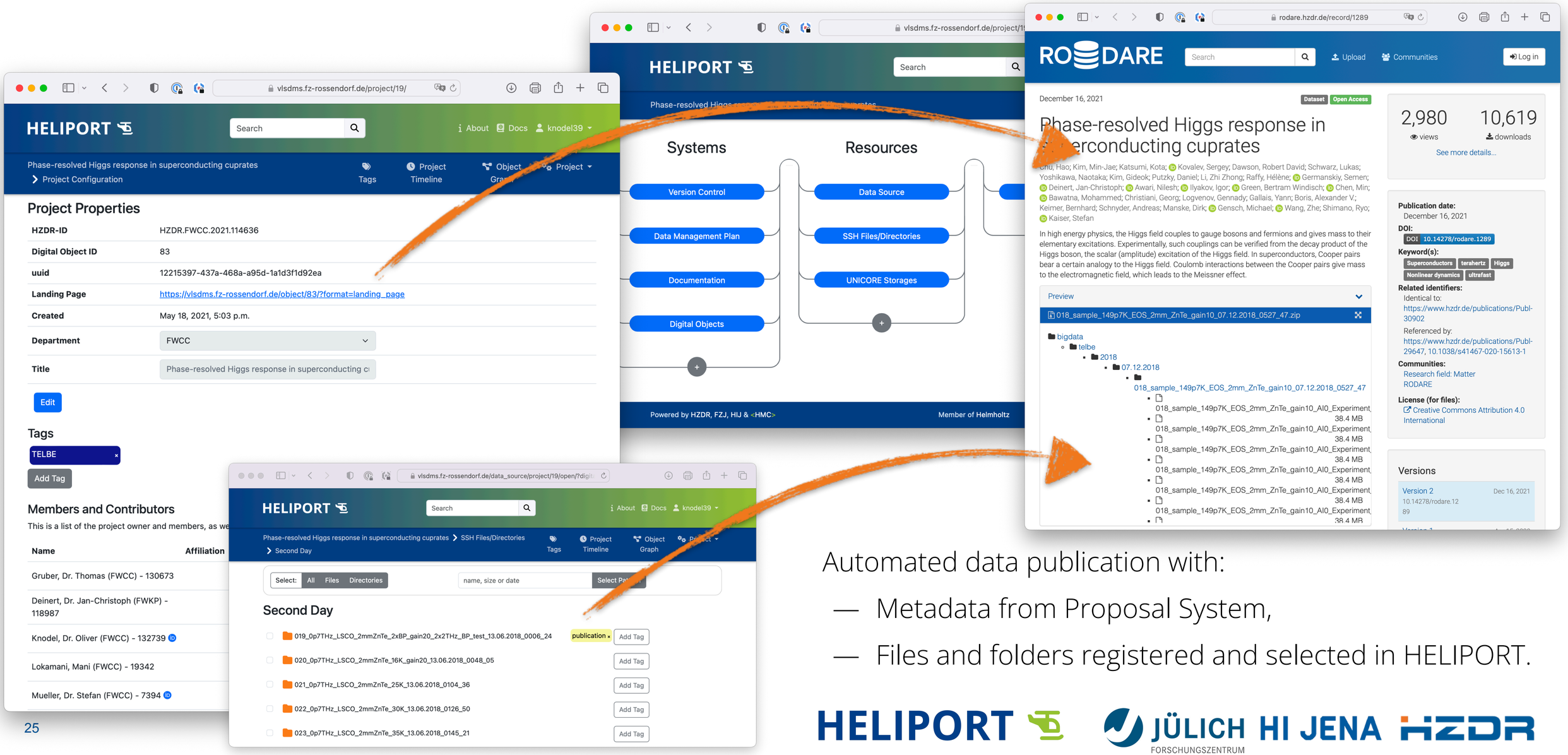
— Each member of a HELIPORT project has access to the files and folders.

— 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:

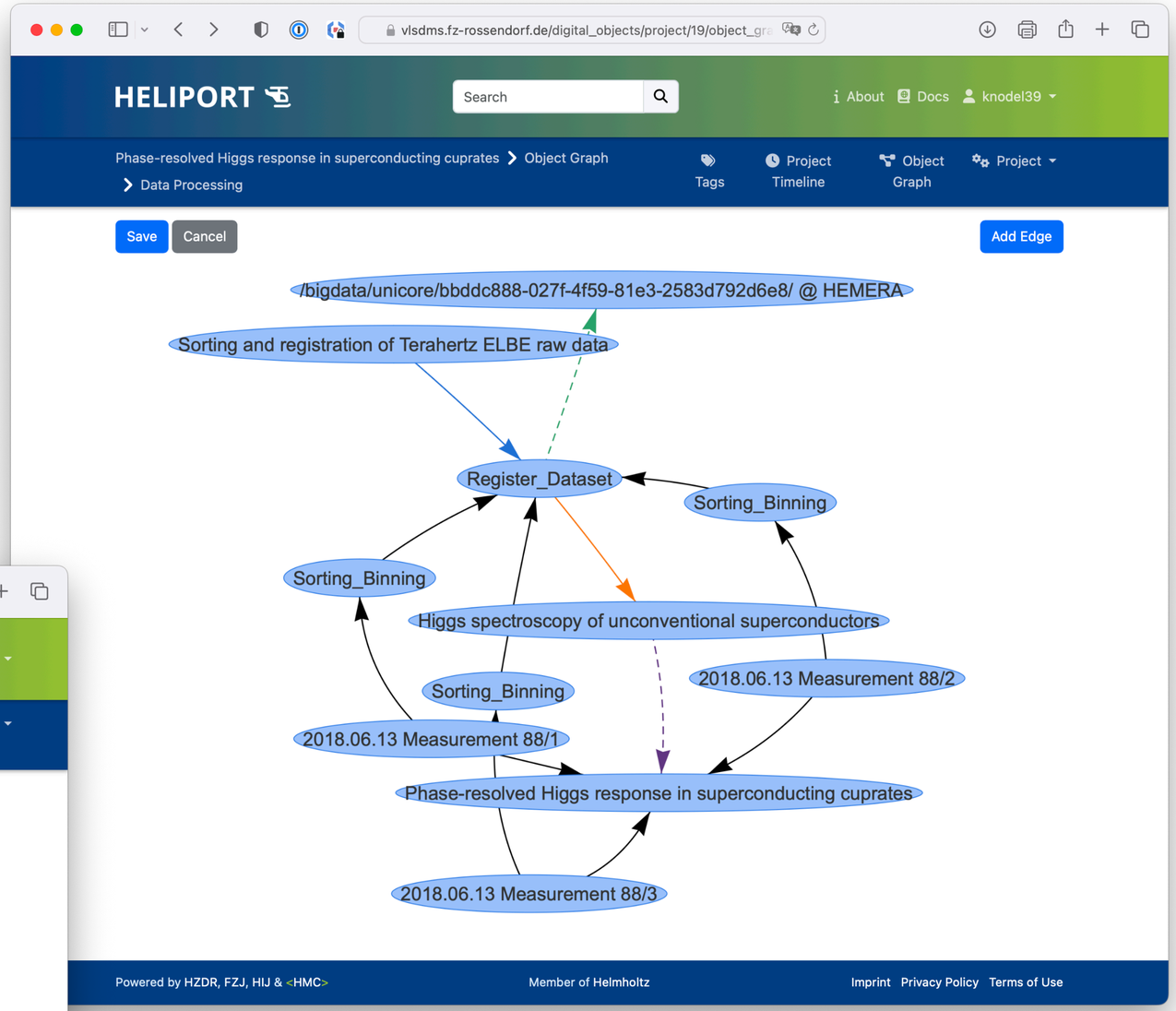
- Metadata from Proposal System,
- Files and folders registered and selected in HELIPORT.

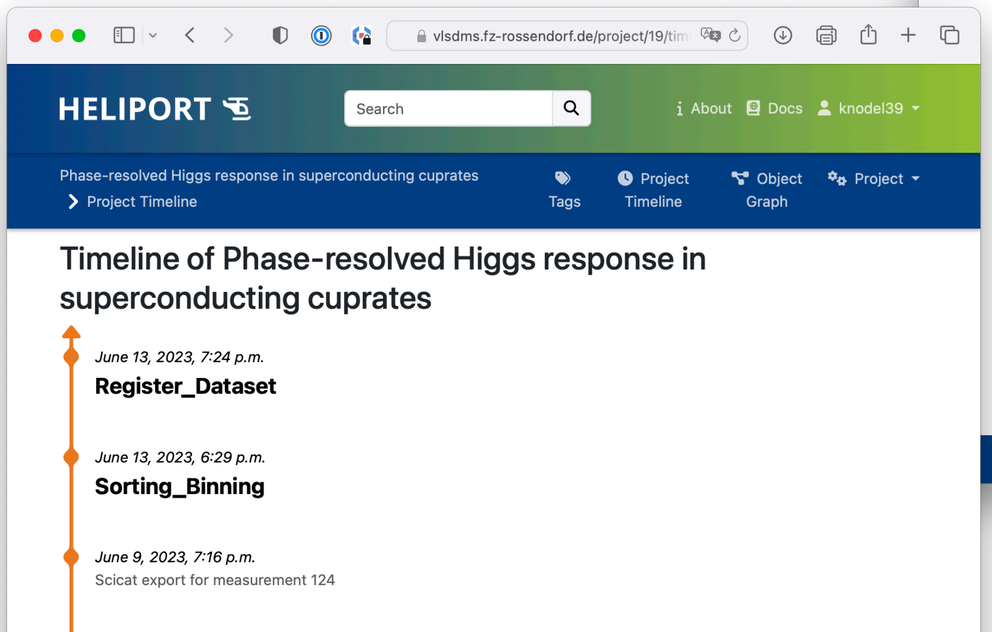
HELIPORT

JÜLICH FORSCHUNGSZENTRUM HI JENA HZDR

VII. 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.



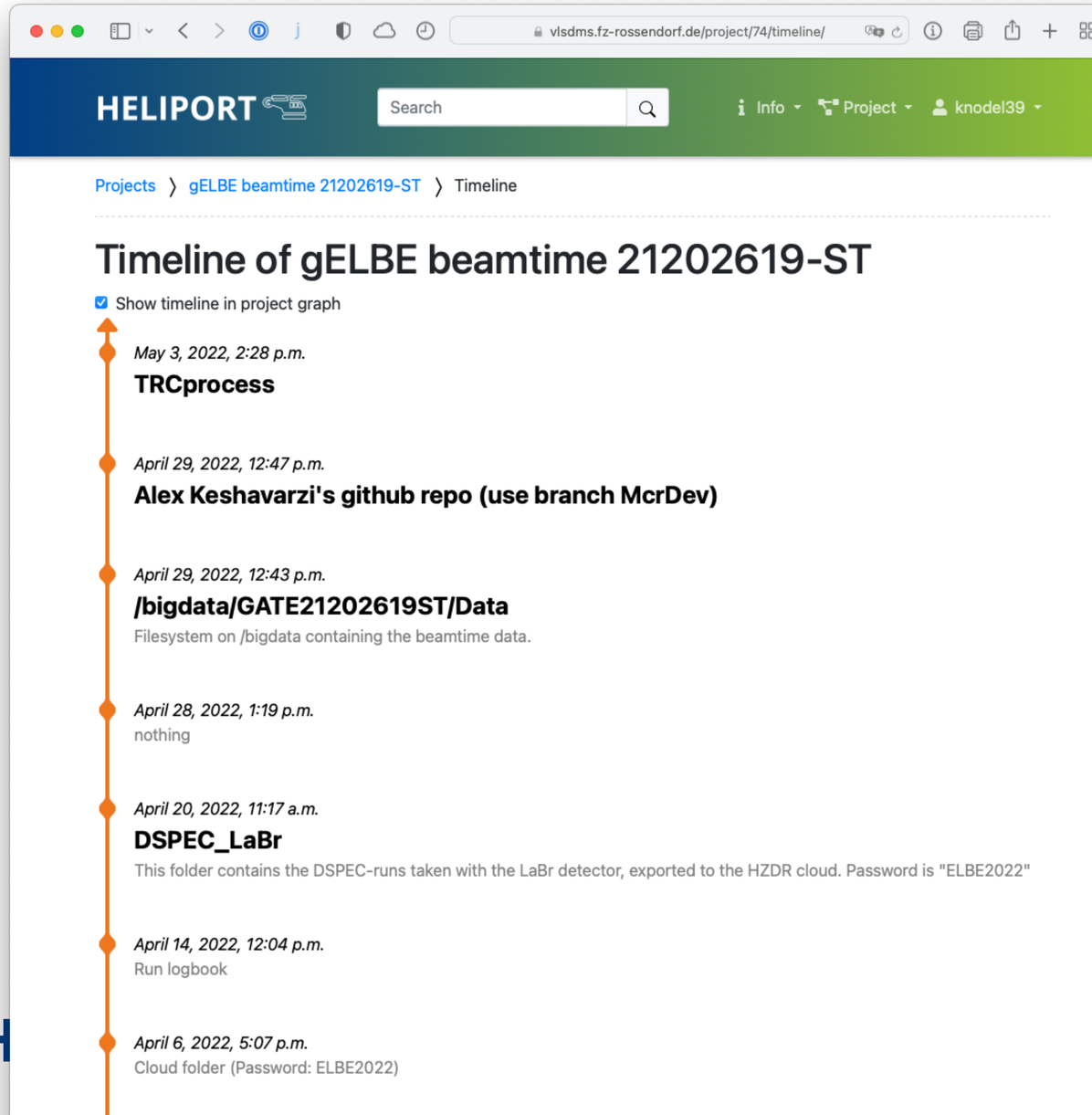


HELIPORT

JÜLICH FORSCHUNGSZENTRUM HI JENA HZDR

Data provenance and Comprehensibility

- For many systems and services we still have to develop necessary plug-ins for the integration into Heliport.
- The versioning of an experiment lifecycle is unavoidable and we are still discussing how we can present the feature in our web frontend:
  - A Git project with all metadata to restore a lifecycle,
  - Or an implementation direct in Heliport?
- Inheritance of projects,
- Different views based on roles (owner, beam line scientist, data curator, …)



Heliport REST API

- The API provides access to our full Heliport infrastructure:
  - Proposal access (GATE),
  - Handle management,
  - CWL execution and monitoring,
  - Project metadata export,
  - Digital Object and
  - Lifecycle management.
- API documentation (ReDOC) available.
- Essential to integrate the Heliport Infrastructure in Experiments.
- Everything can be documented with less user interaction.