Overarching Data Management Ecosystem HELIPORT

FAIR Data Principles in Physical Sciences in NFDI // April 26th, 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**.

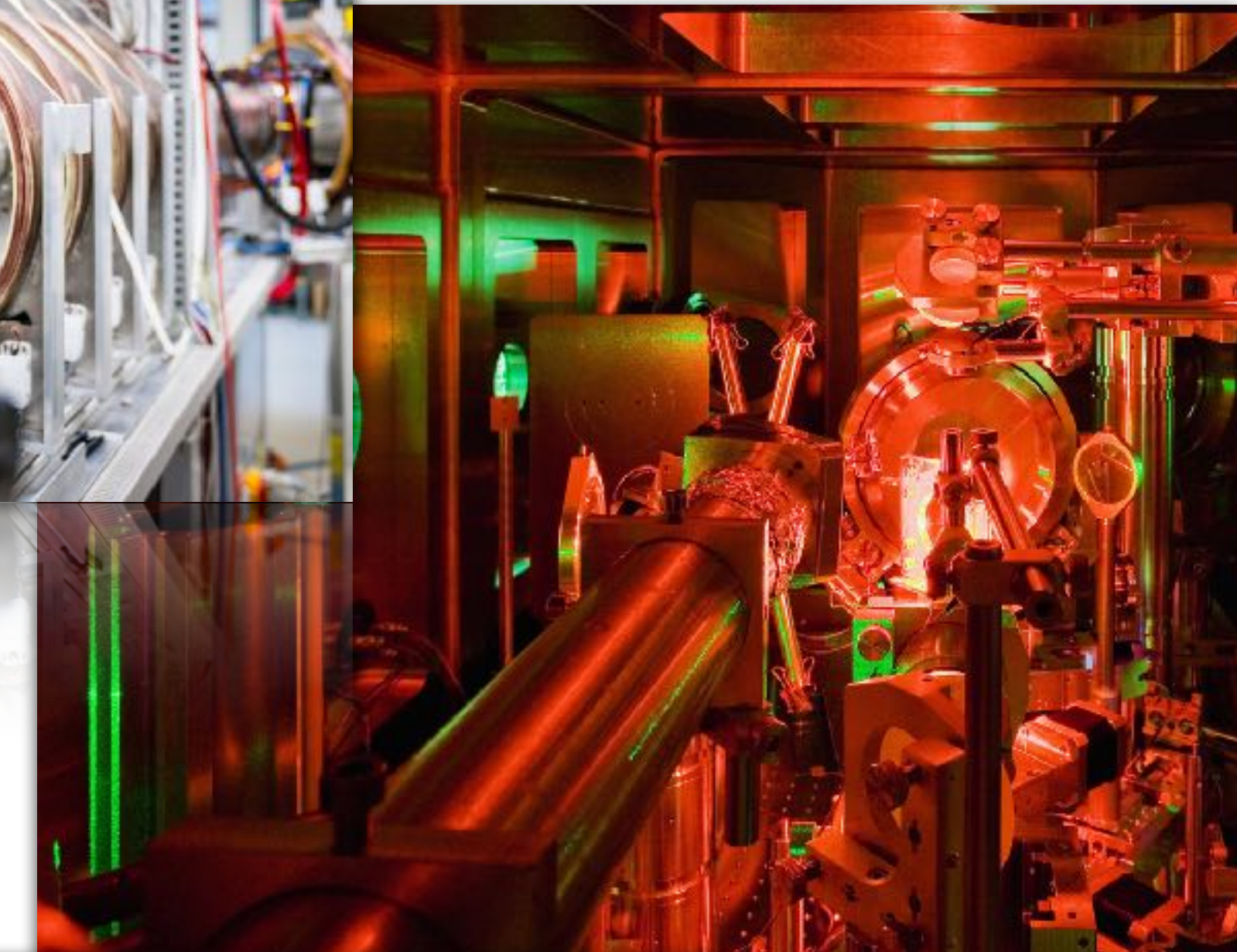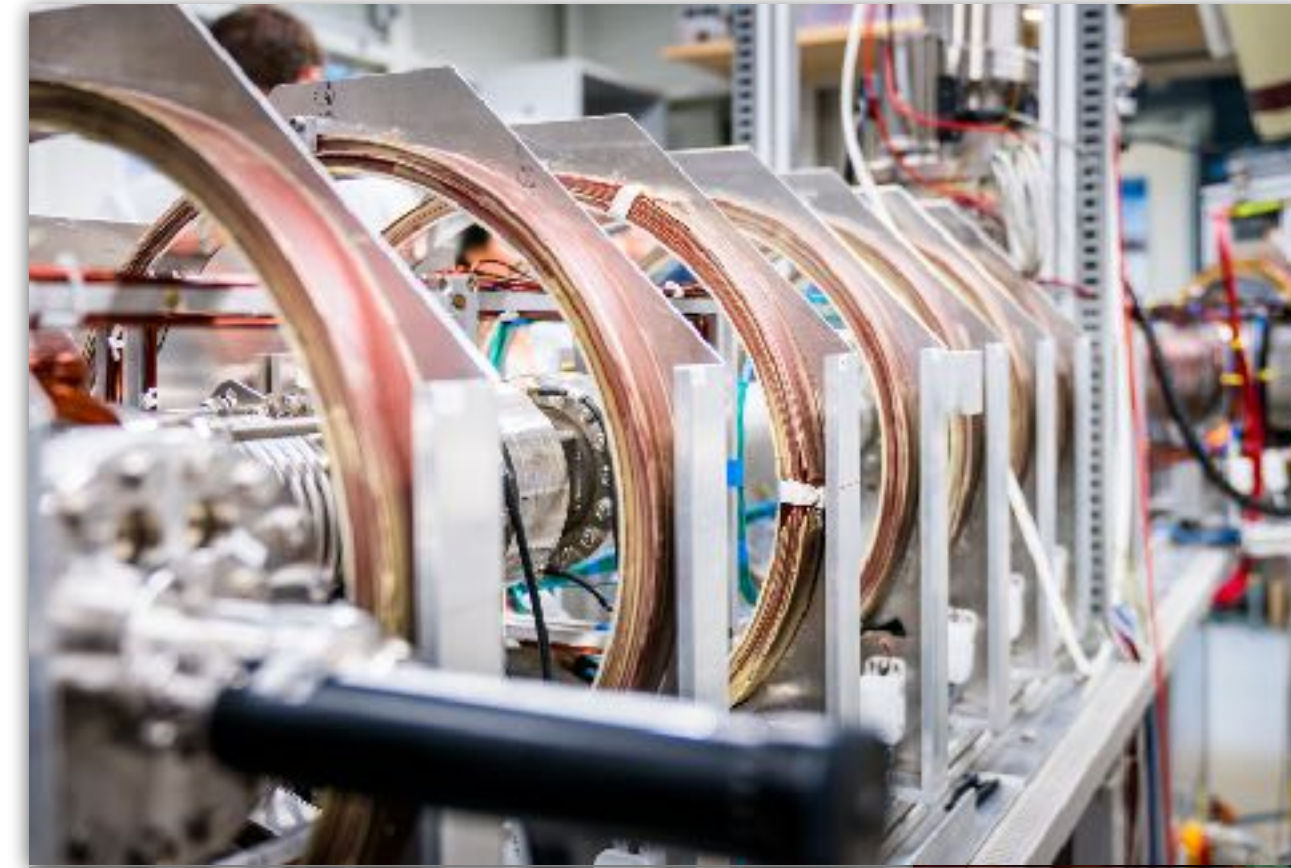## ELBE – Center for High-Power Radiation Sources

— 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

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.
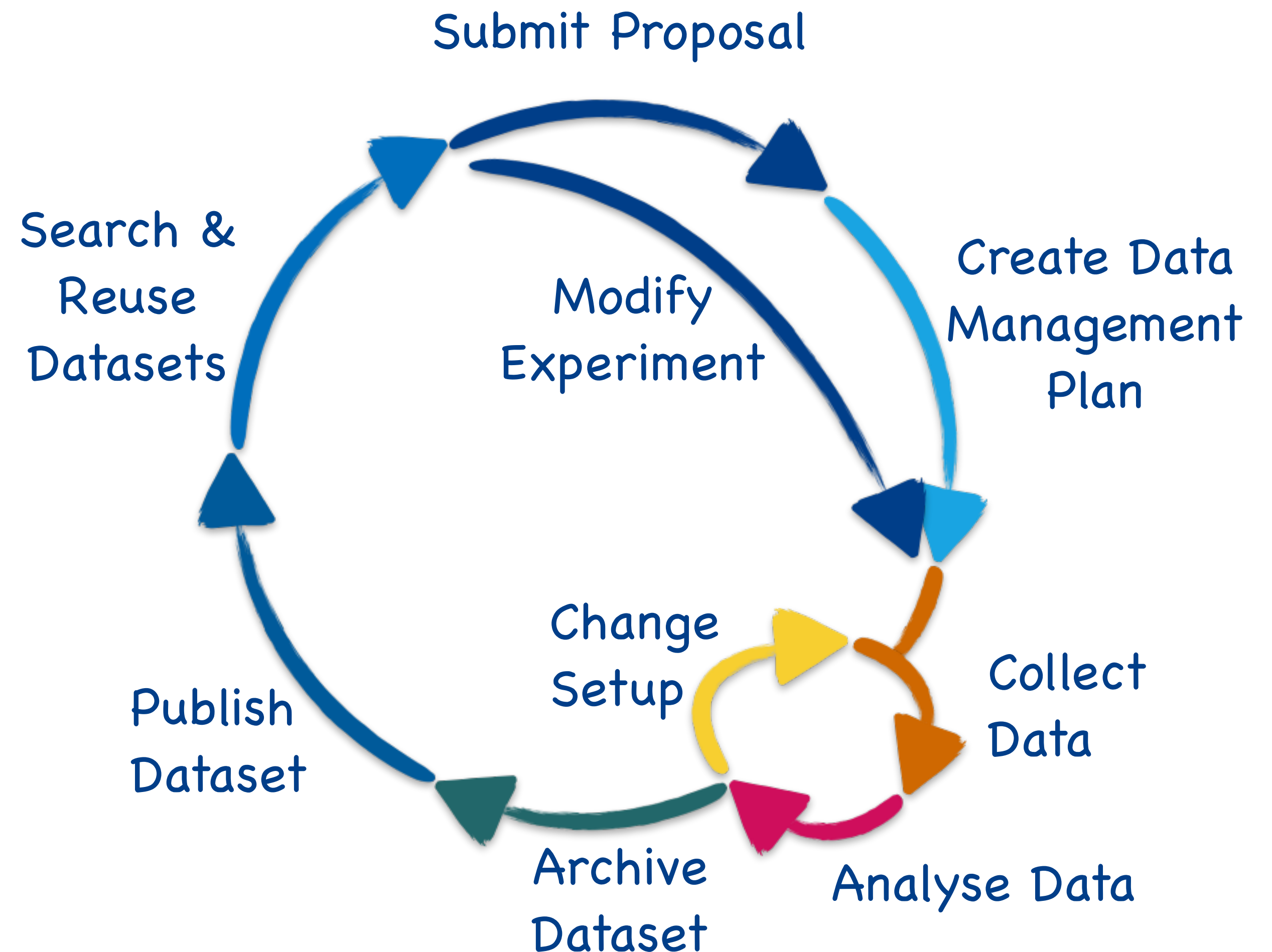
Our Observations and Experiences

— 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!
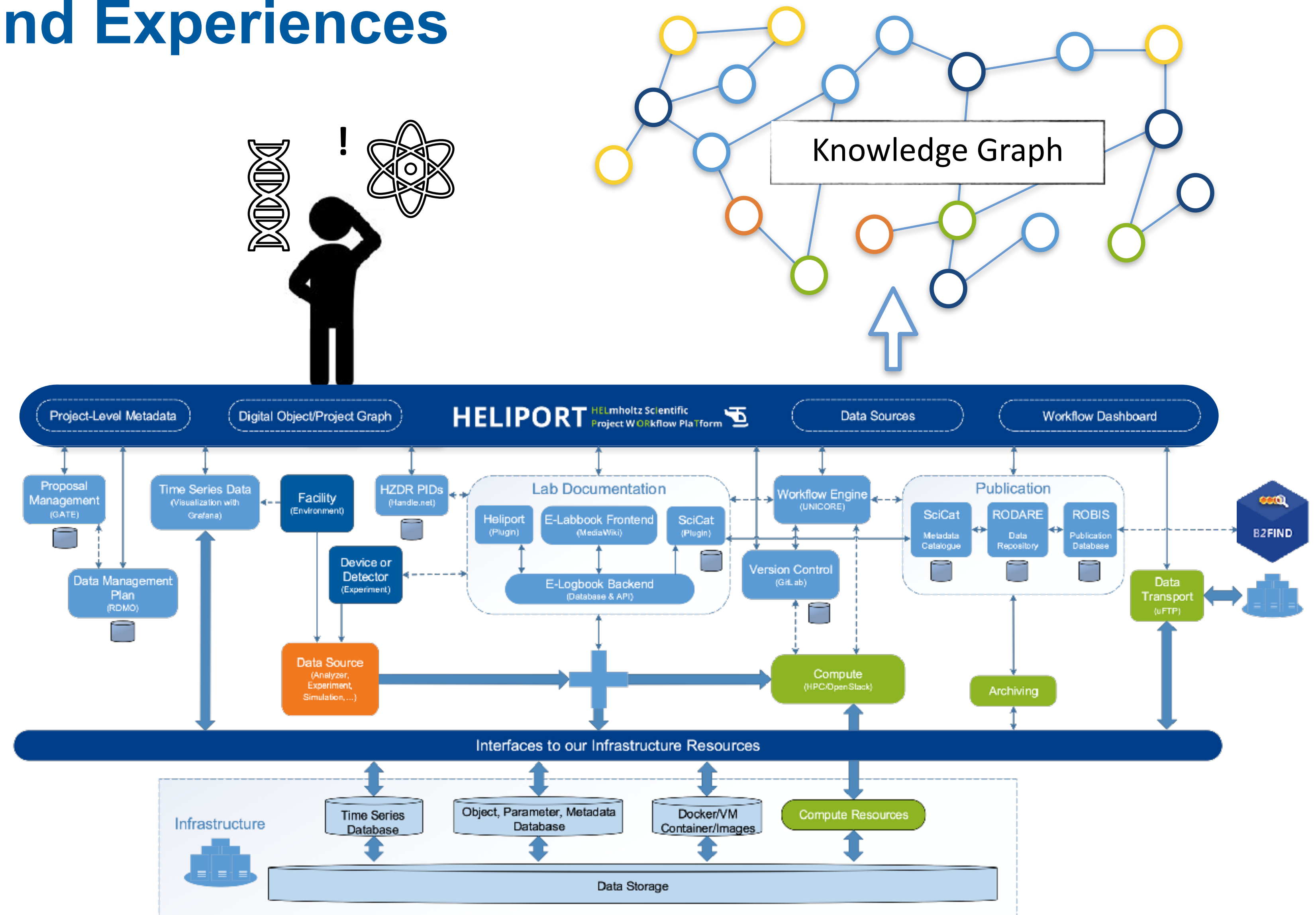
The Idea for a HMC Project: Common entry point for experiments
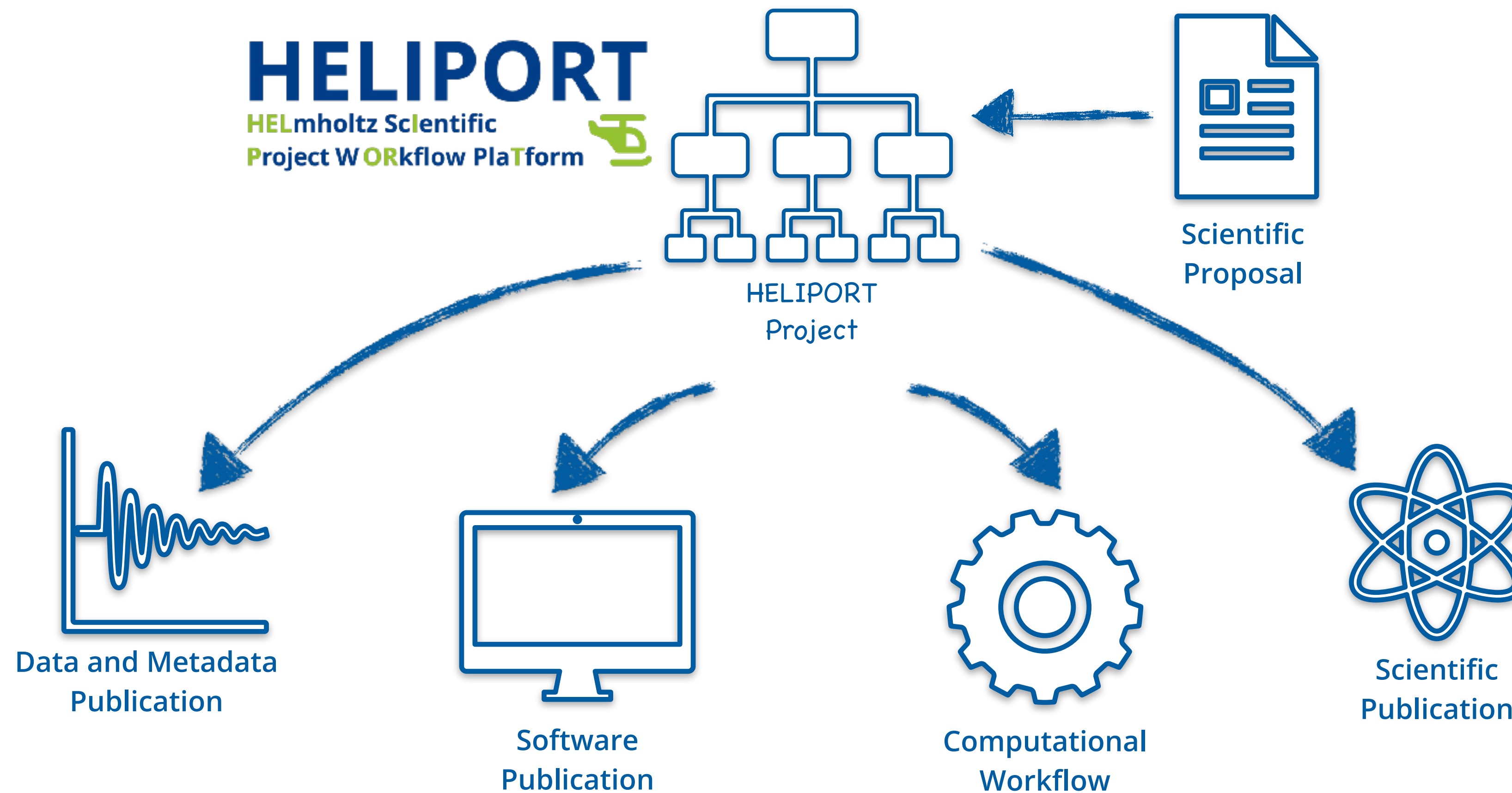
The HMC Proposal HELIPORT

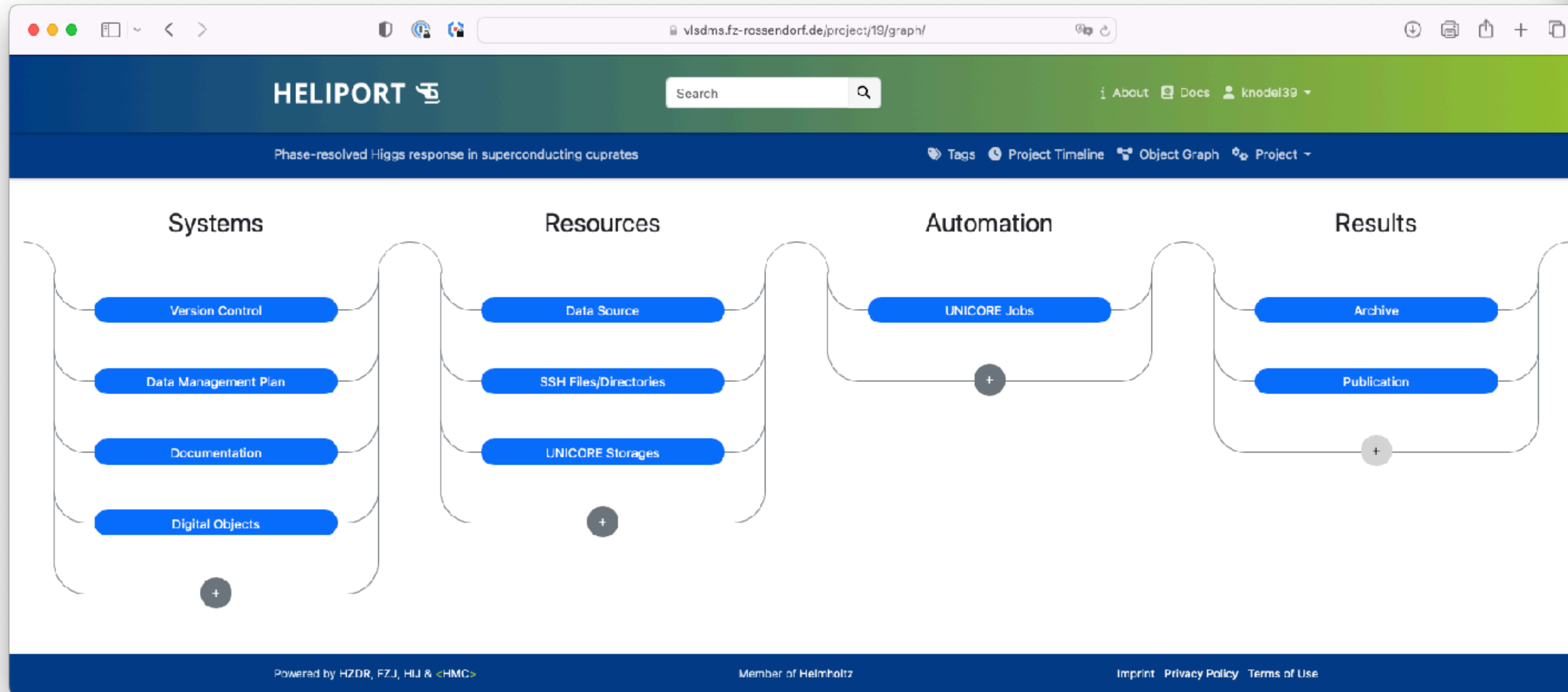**HEL**mholtz **ScI**entific **P**roject **W**ORkflow Pla**T**form

❝ 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.
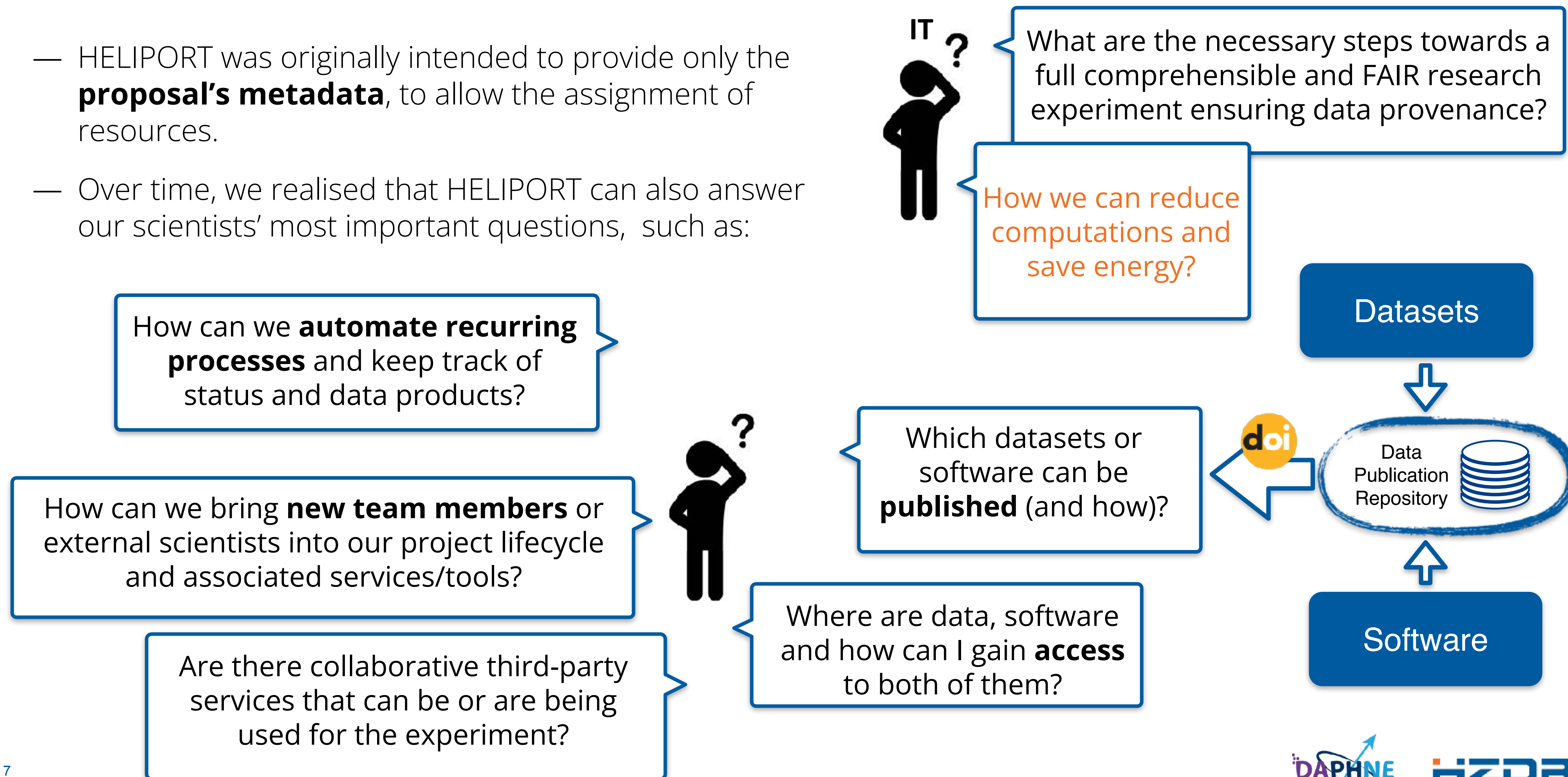
Project Members:

HZDR — HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

HI JENA — Helmholtz Institute Jena

JÜLICH — FORSCHUNGSZENTRUM

Funded by:

\<HMC\> HELMHOLTZ Metadata Collaboration

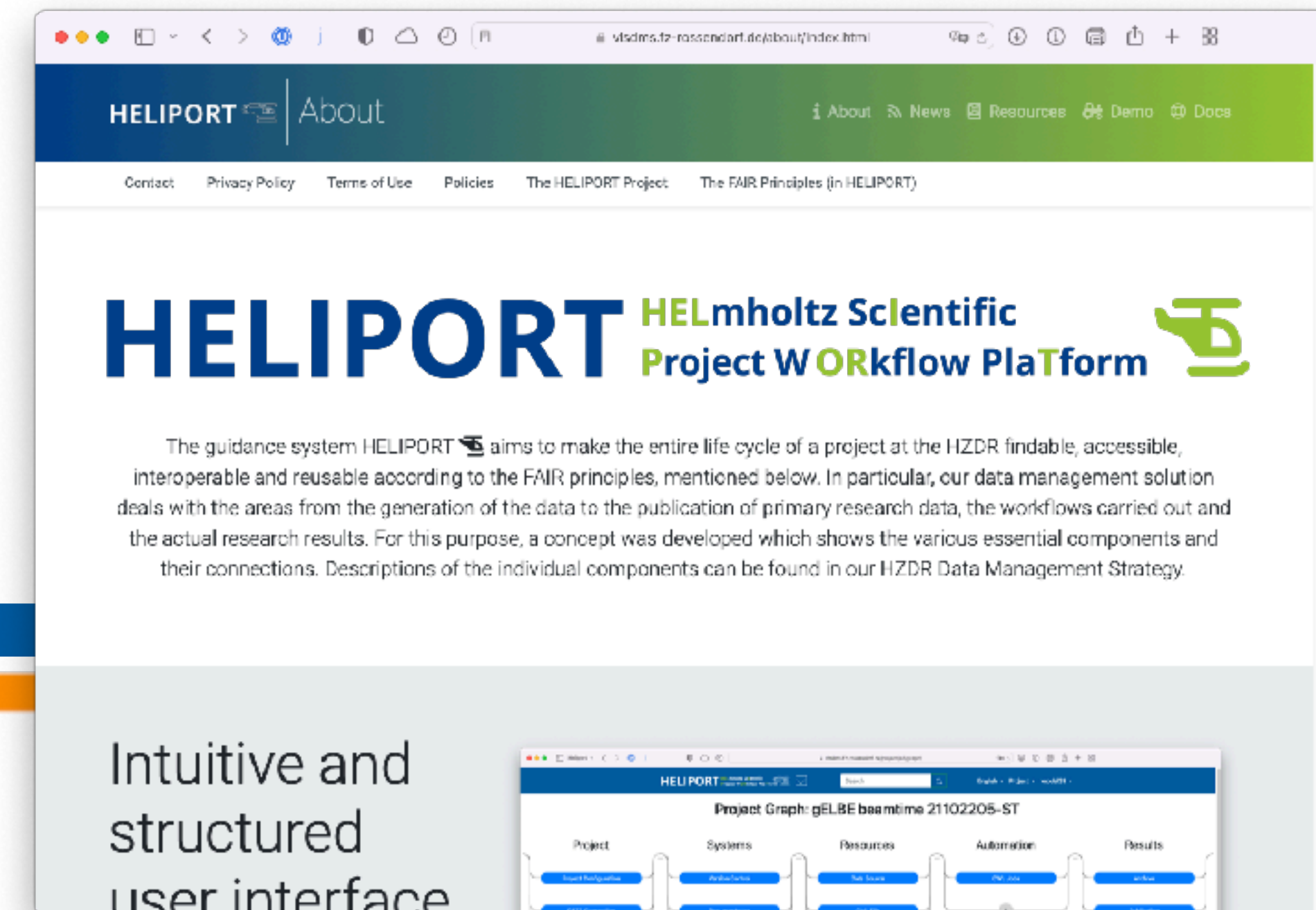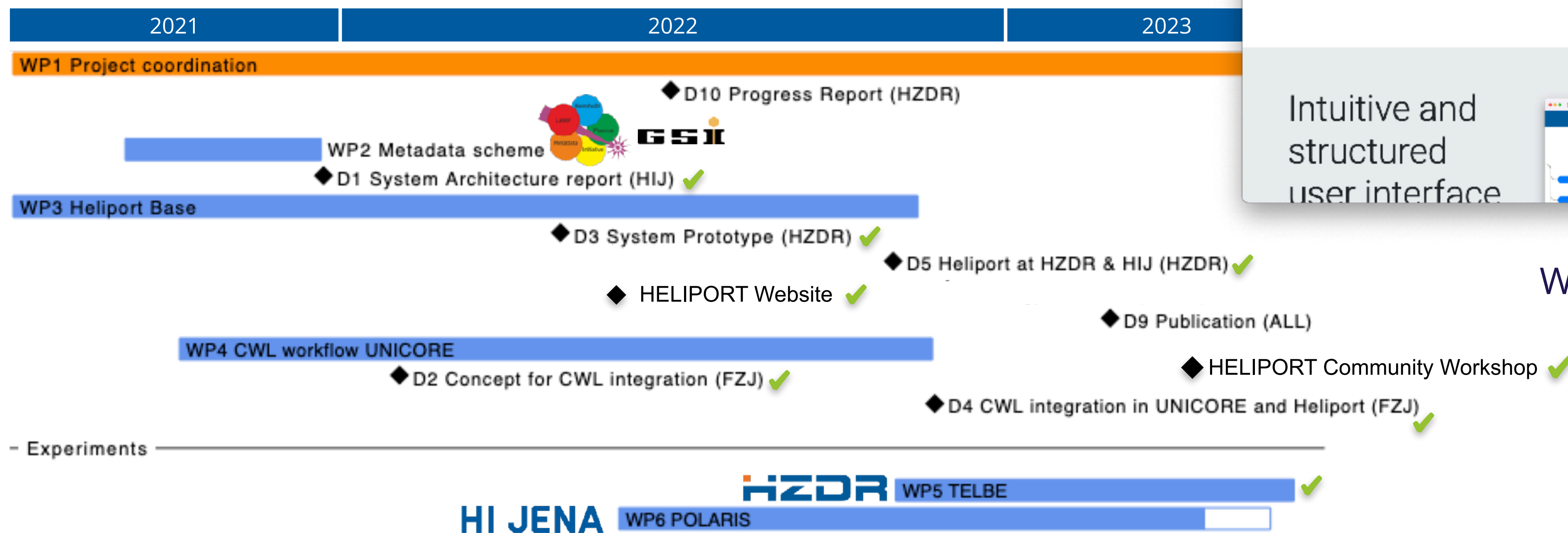Metadata crosswalk to **schema.org** **ResearchProject**

Motivation for HELIPORT — User Perspective

— 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:

IT ?

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

How we can reduce computations and save energy?

Datasets

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

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

Data Publication Repository

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?

Software

Are there collaborative third-party services that can be or are being used for the experiment?

DAPHNE 4NFDI

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

2021	2022	2023

WP1 Project coordination

◆ D10 Progress Report (HZDR)

WP2 Metadata scheme

◆ D1 System Architecture report (HIJ) ✓

WP3 Heliport Base

◆ D3 System Prototype (HZDR) ✓

◆ D5 Heliport at HZDR & HIJ (HZDR) ✓

◆ HELIPORT Website ✓

◆ D9 Publication (ALL)

WP4 CWL workflow UNICORE

◆ D2 Concept for CWL integration (FZJ) ✓

◆ HELIPORT Community Workshop ✓

◆ D4 CWL integration in UNICORE and Heliport (FZJ) ✓

– Experiments

WP5 TELBE ✓

HI JENA WP6 POLARIS

Heliport (Project) Timeline

0.1.0

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

0.2.0

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

0.3.X

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

0.4.0

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

<HMC> | HELMHOLTZ METADATA COLLABORATION

— Redesign to provide modular and highly configurable system

0.5.X

Integration of various Apps and Features (Mid 2022)
— Export for (different) metadata schemas
— Computational/scientific workflow execution
  · UNICORE support
  · Computing job management and monitoring
— Handle management with public landing pages
— **HELIPORT Website**: heliport.hzdr.de

HELIPORT Community Workshop (July 2023)

GFZ Helmholtz-Zentrum POTSDAM   HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

JÜLICH FORSCHUNGSZENTRUM   GSI   HI JENA

0.6.0

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

DAPHNE 4NFDI   HZDR

Community Website: heliport.hzdr.de

— Central access point to our resources at **heliport.hzdr.de** with:

  · Overall information on HELIPORT,

  · Documentation,

  · Ressources (Poster, Presentation, Paper),

  · News and Events section and

— Our HELIPORT instances is a first step towards an overall community portal (maybe heliport.helmholtz.cloud?).
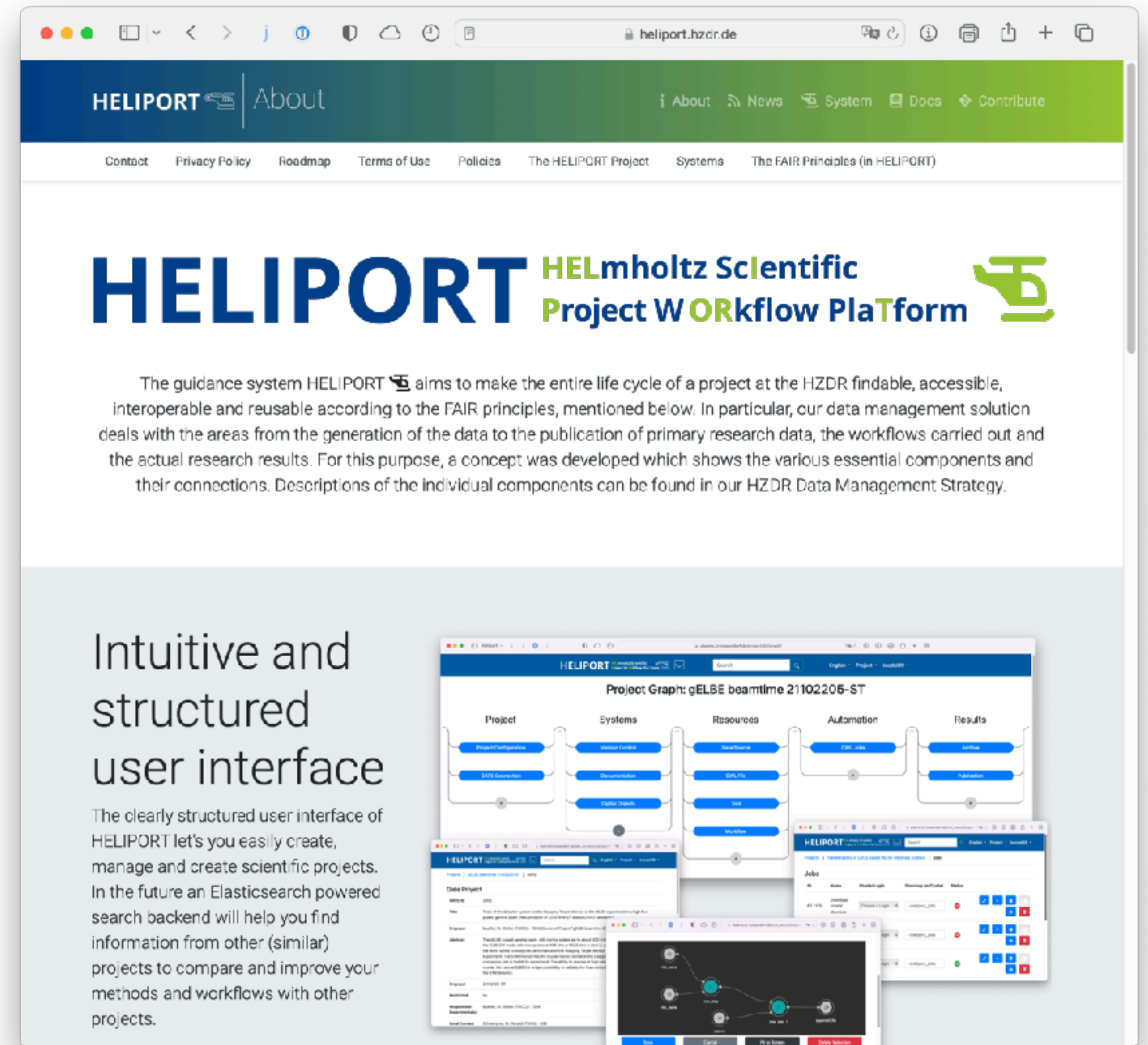
HELIPORT Features

— 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)

— Digital object and handle management with graph visualisation

— Timeline representing changes

— HELIPORT REST API

— Authentication via Helmholtz ID

— ToDos:
  - Integration of reproducible computational workflows
  - HPC cluster access (slurm, UNICORE)
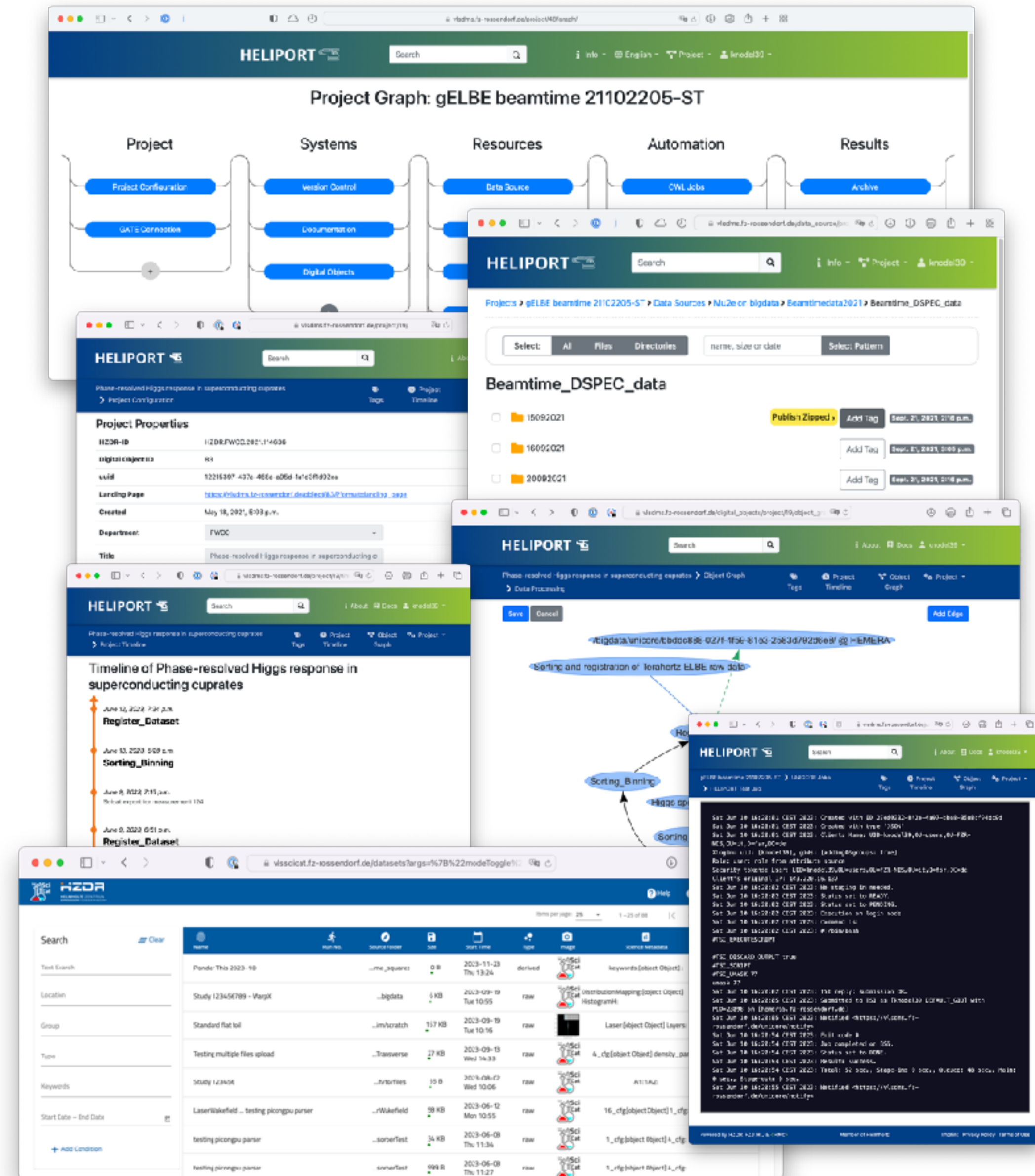
Example: TELBE Beamtime

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 our HPC Cluster with visualisation in HELIPORT
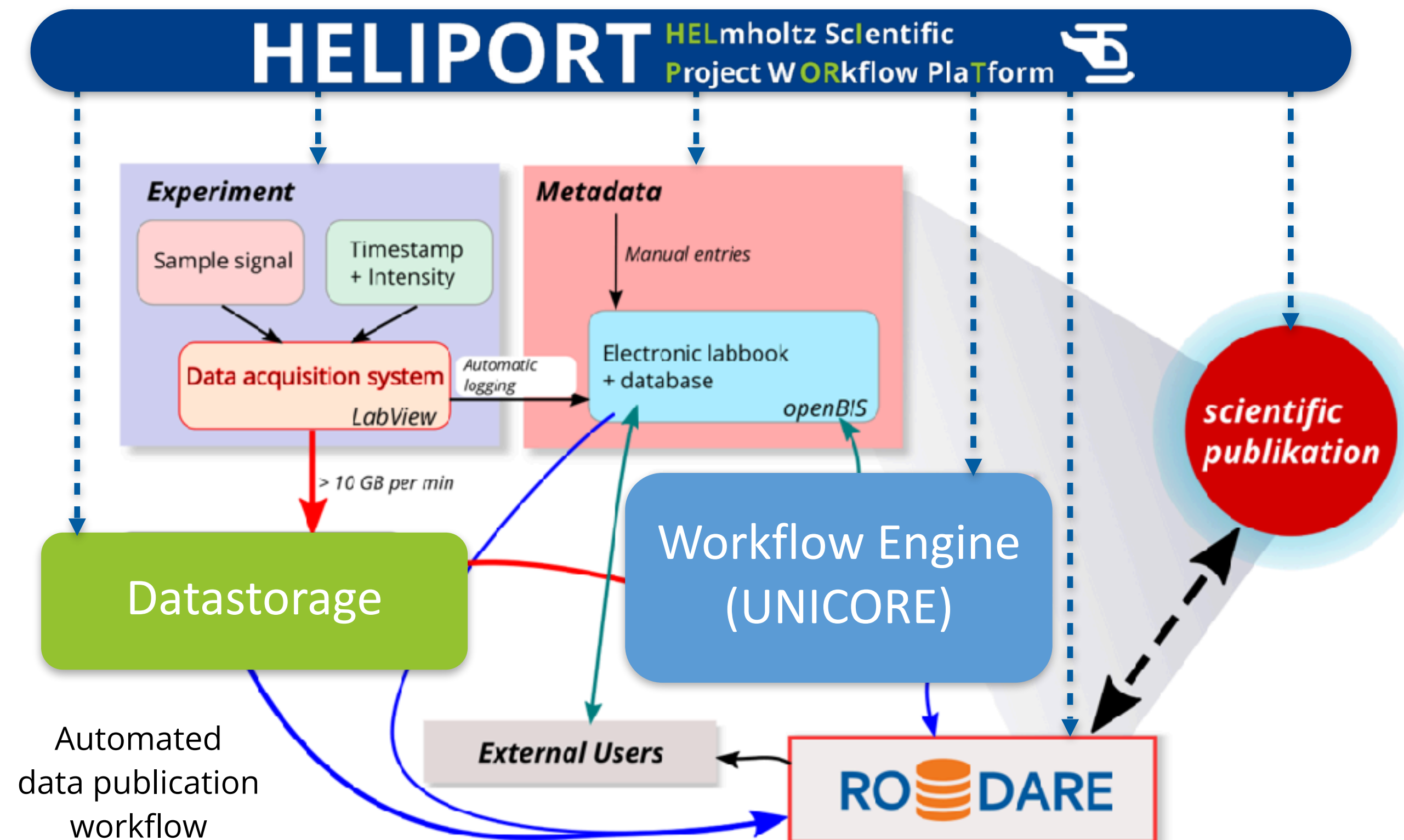
Mapping of the TELBE Resources to HELIPORT
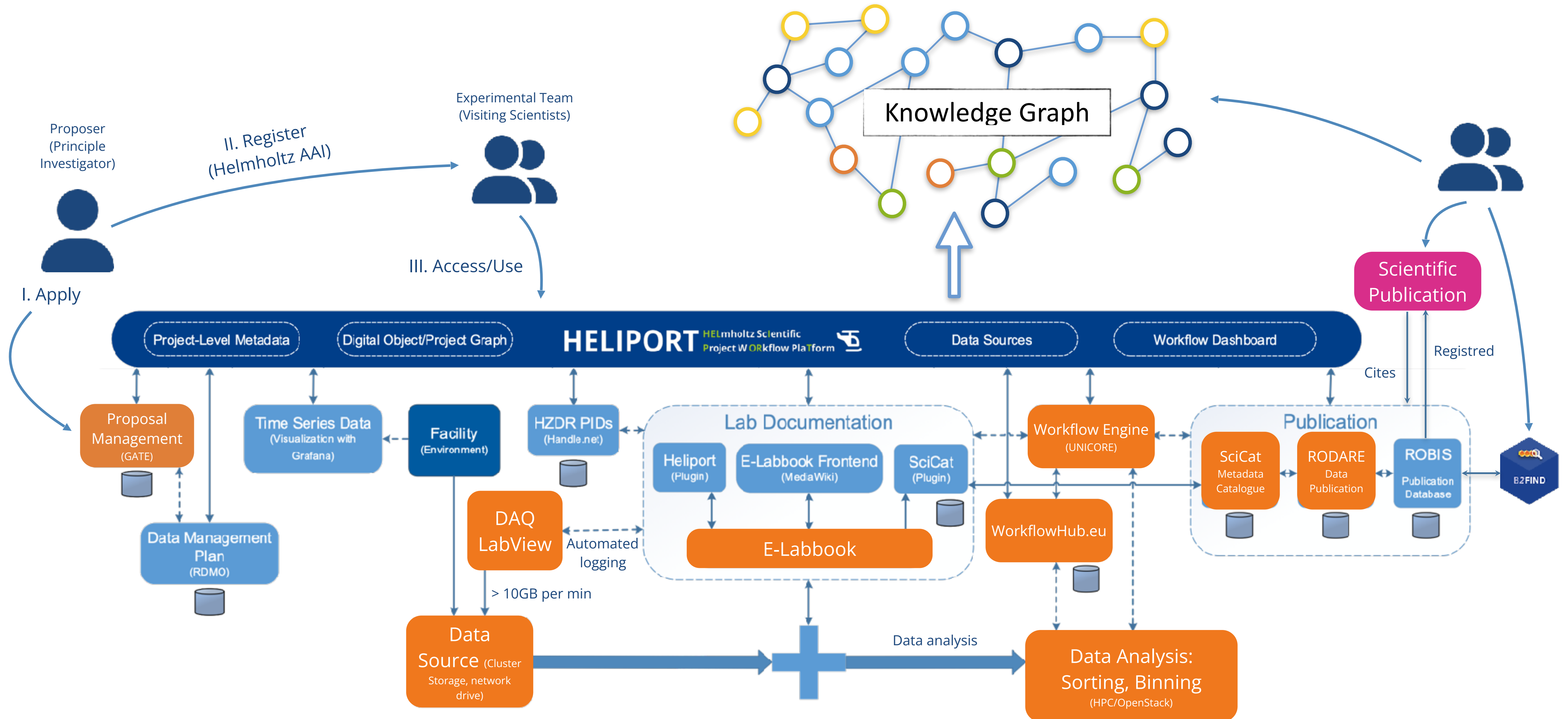
I. Proposal Submission

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

— Title, Authors, Description,

— Beamtime schedule,

— Large-scale facility used,

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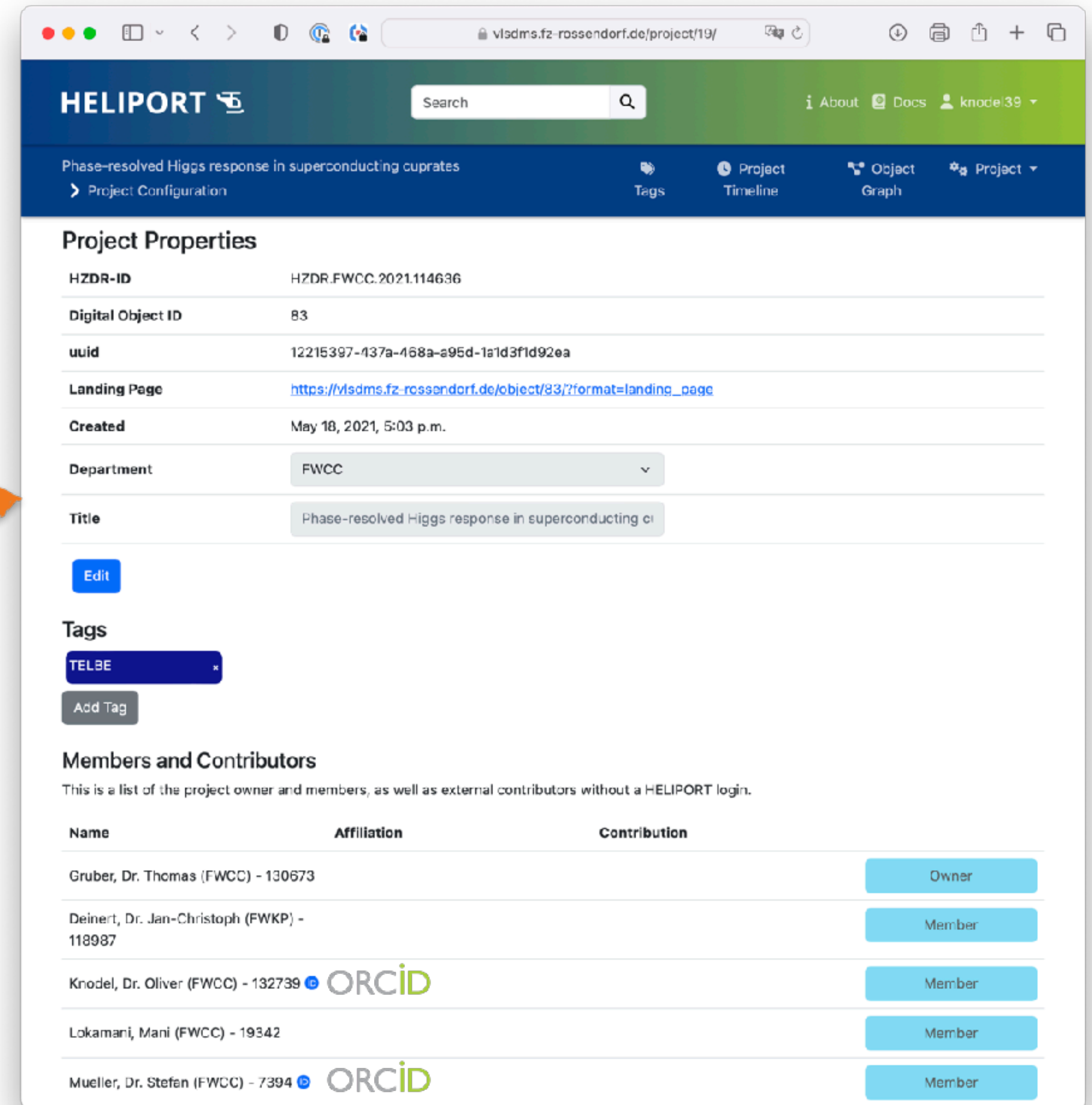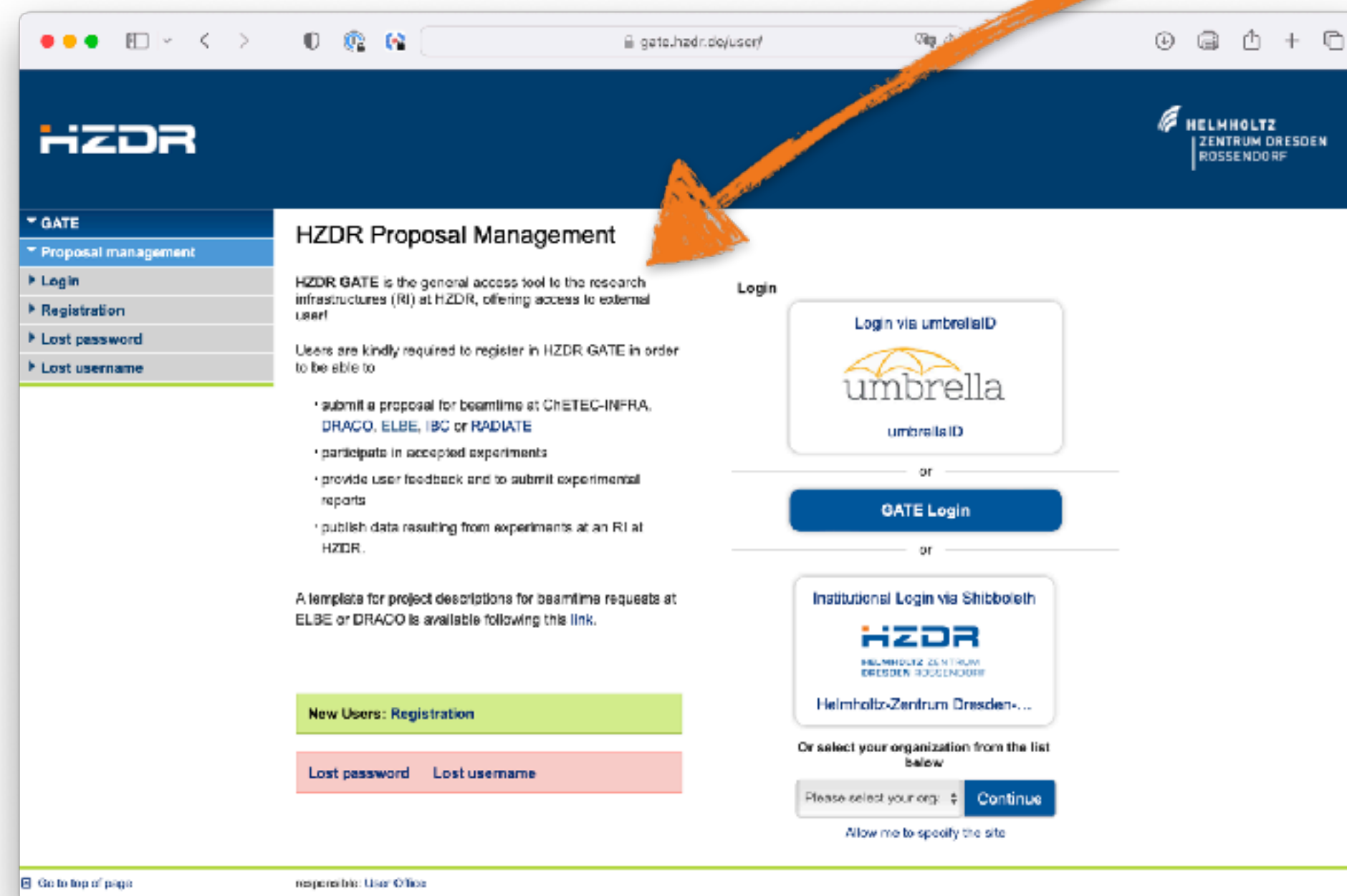
II. Project List and Dashboard

— 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 viewed by any project member directly in the HELIPORT web frontend.
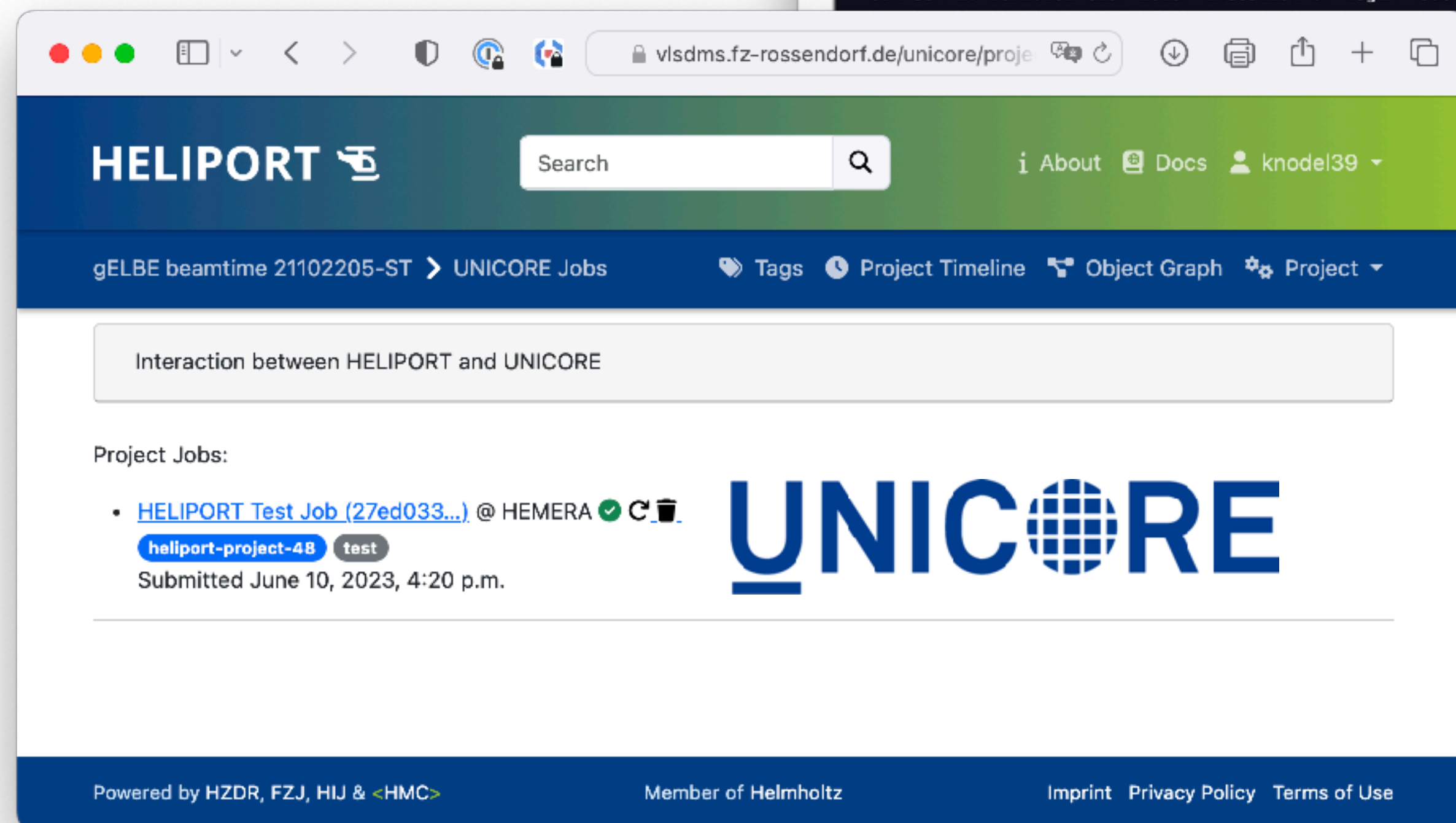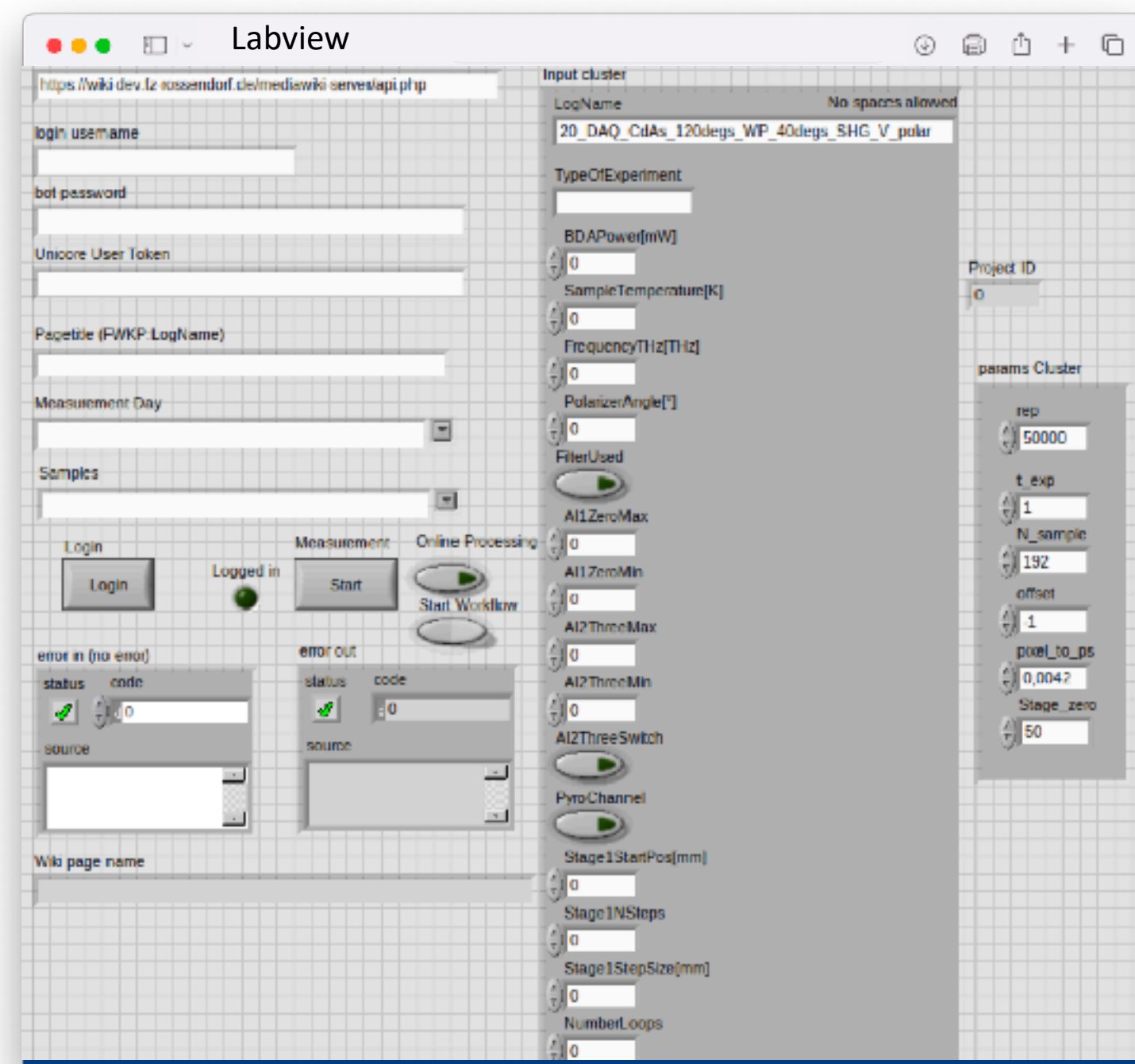
V. Data Sources

— 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.
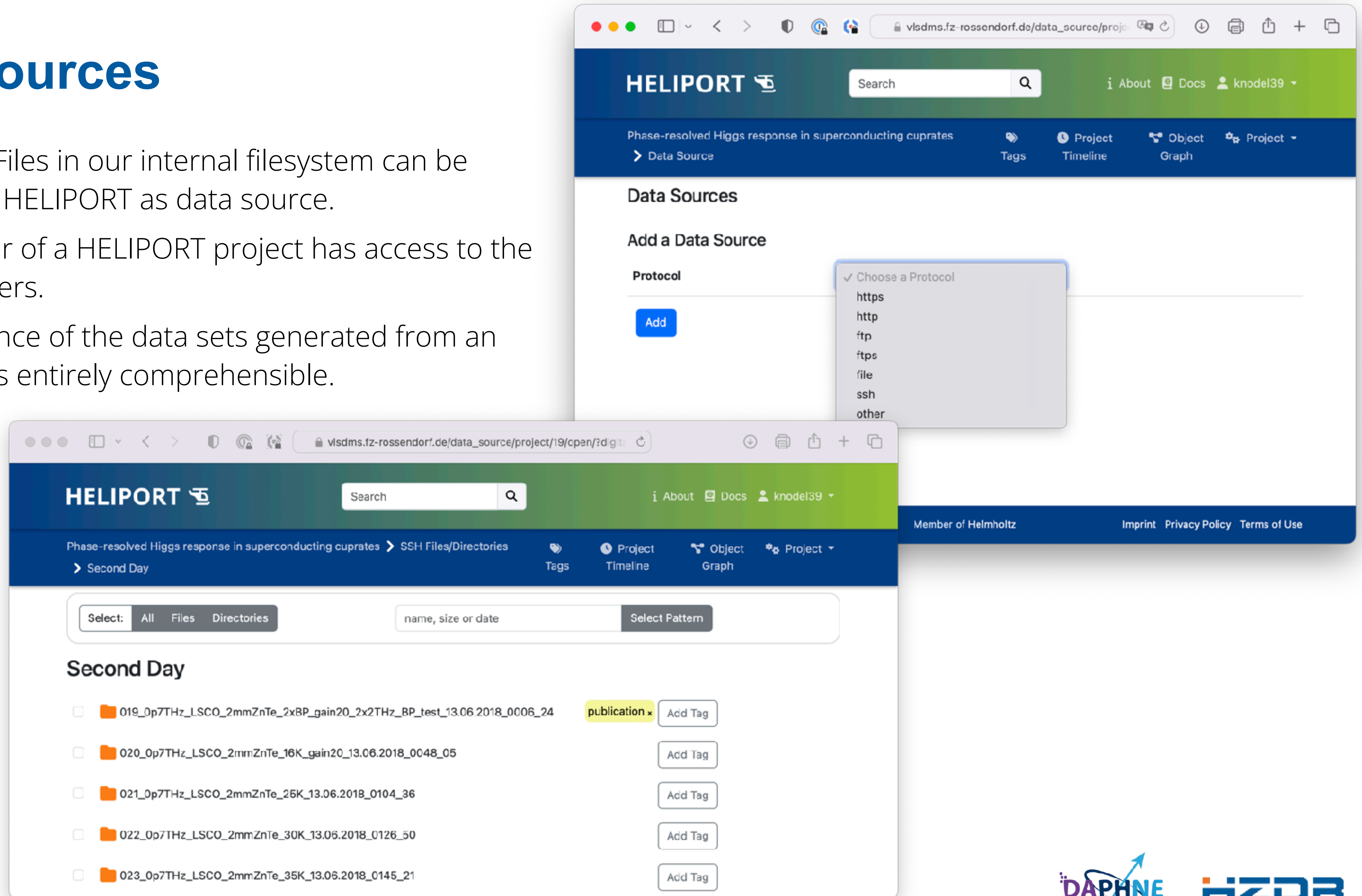
VI. Integration in an Overall Data Publication Workflow

Automated data publication with:

— Metadata from Proposal System,

— Files and folders registered and selected in HELIPORT.

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.
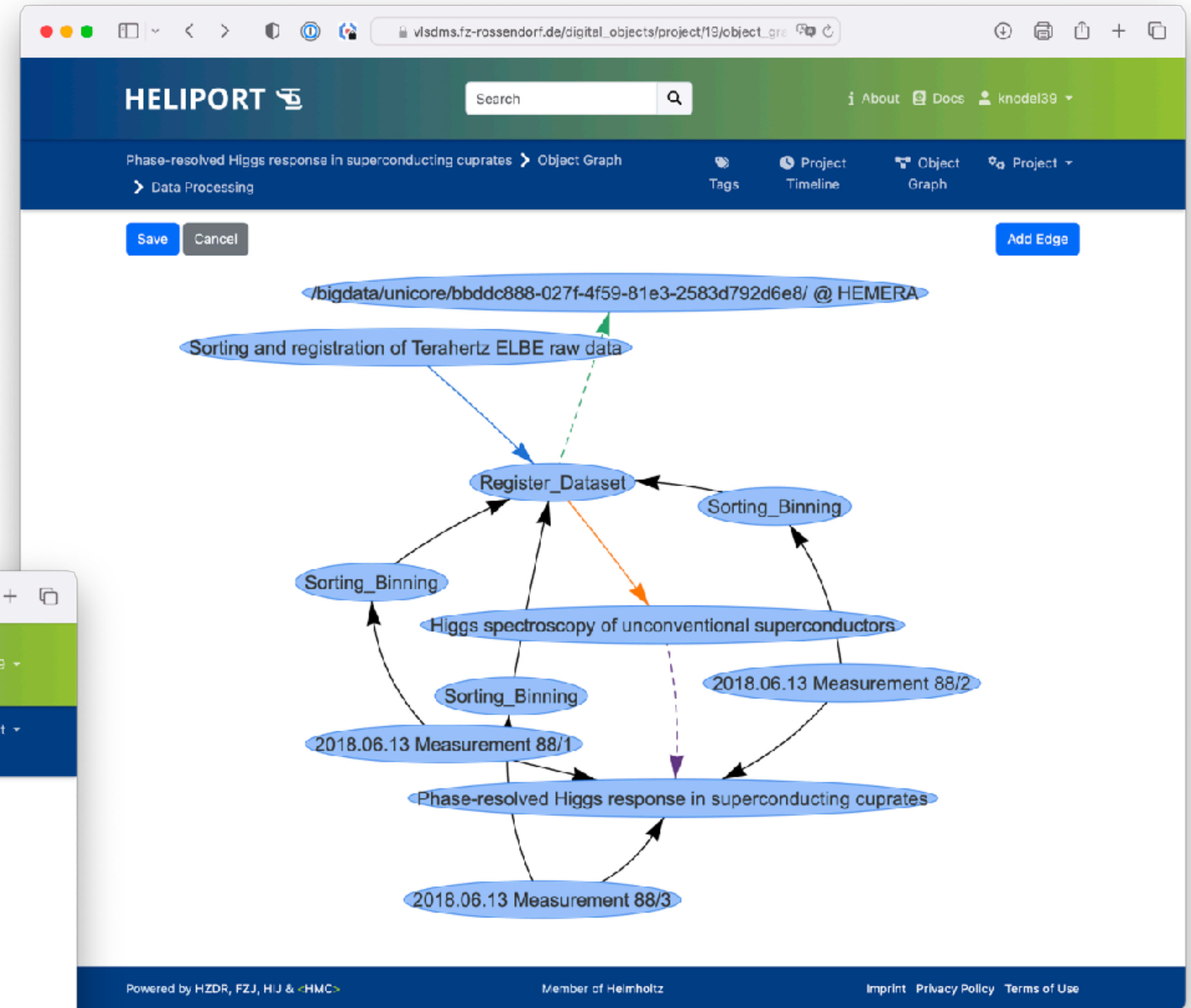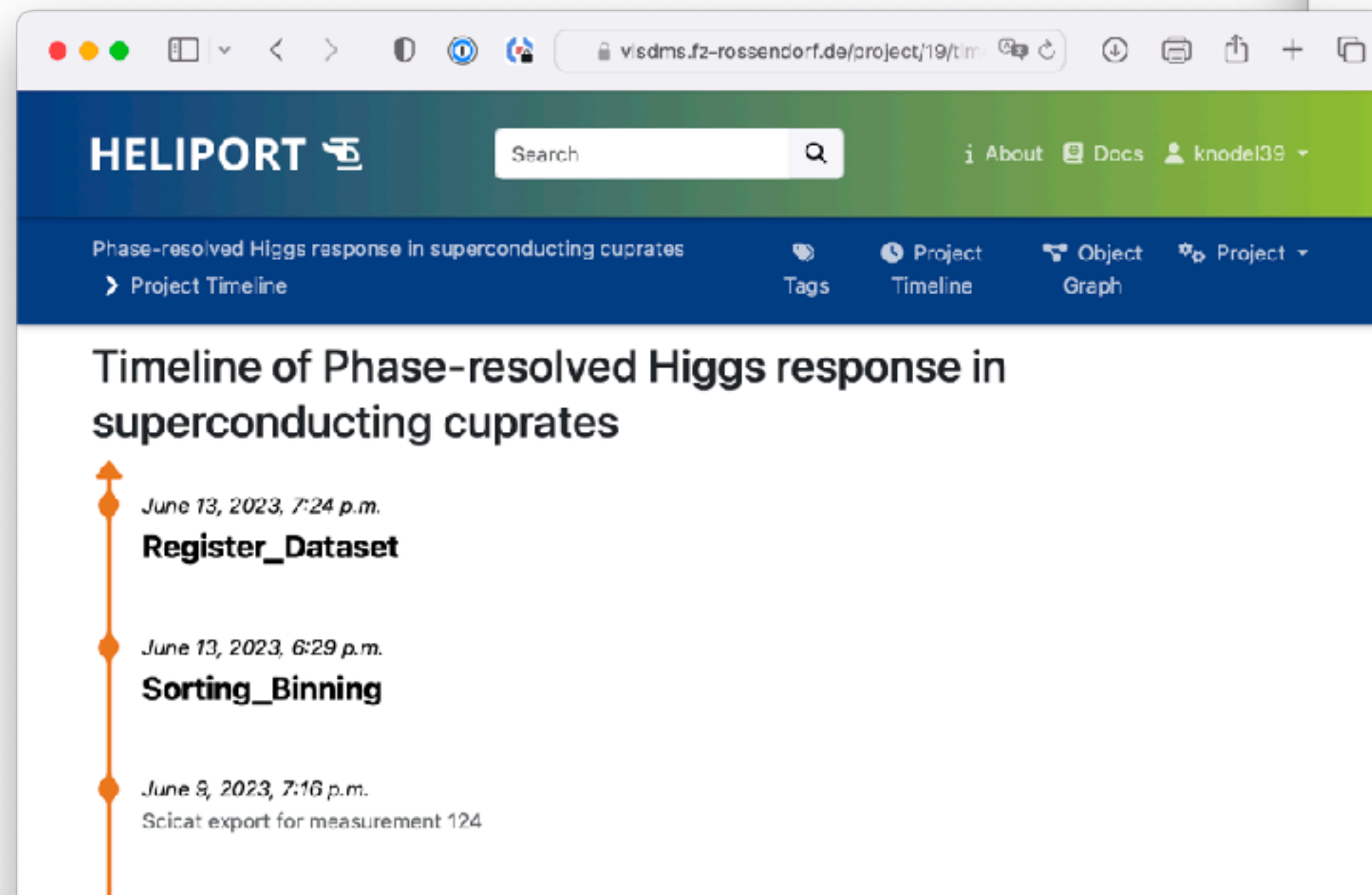
Conclusions

— 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.

— Data sources and workflows are essential to keep track of everything what happened during an experiment.
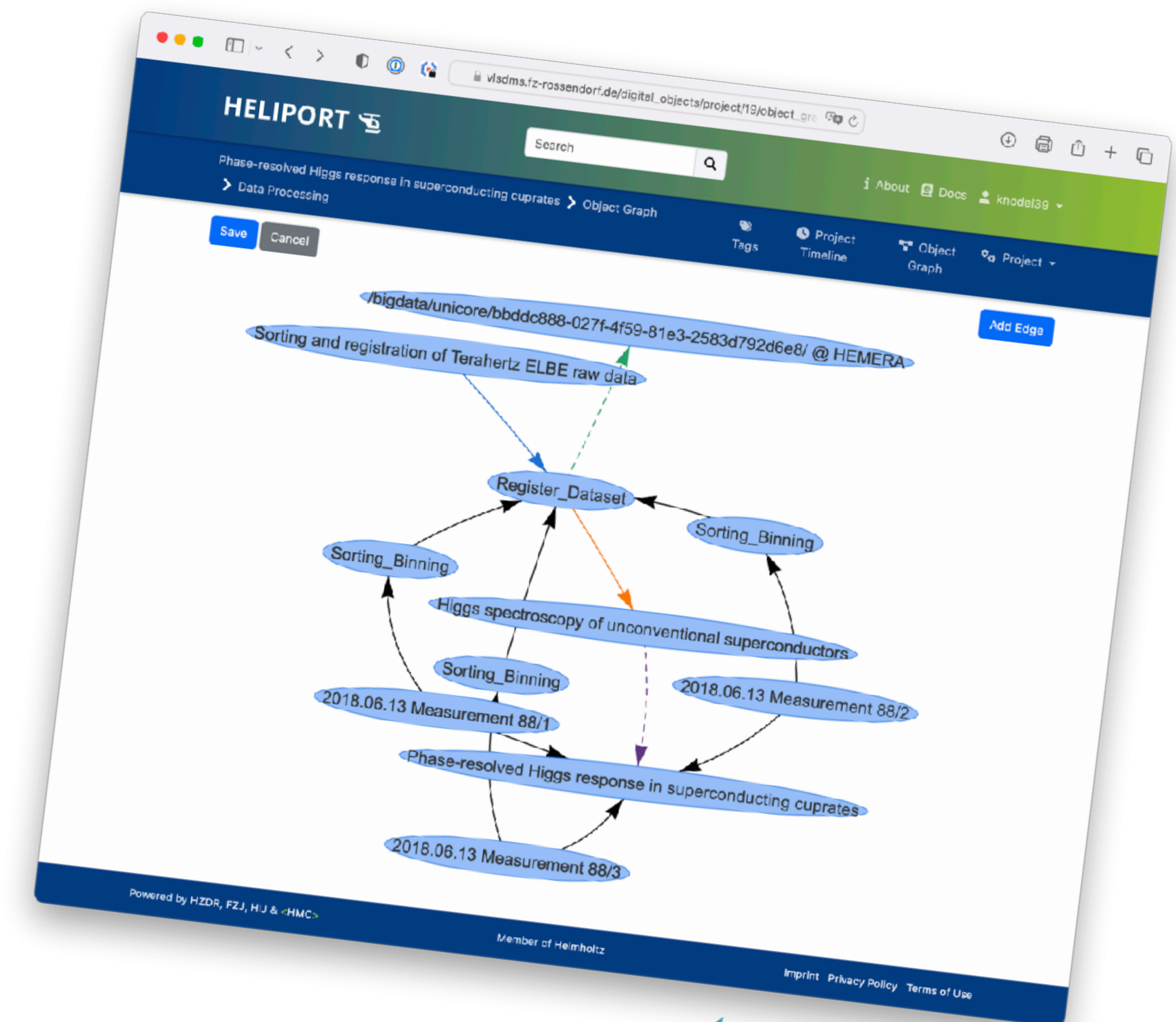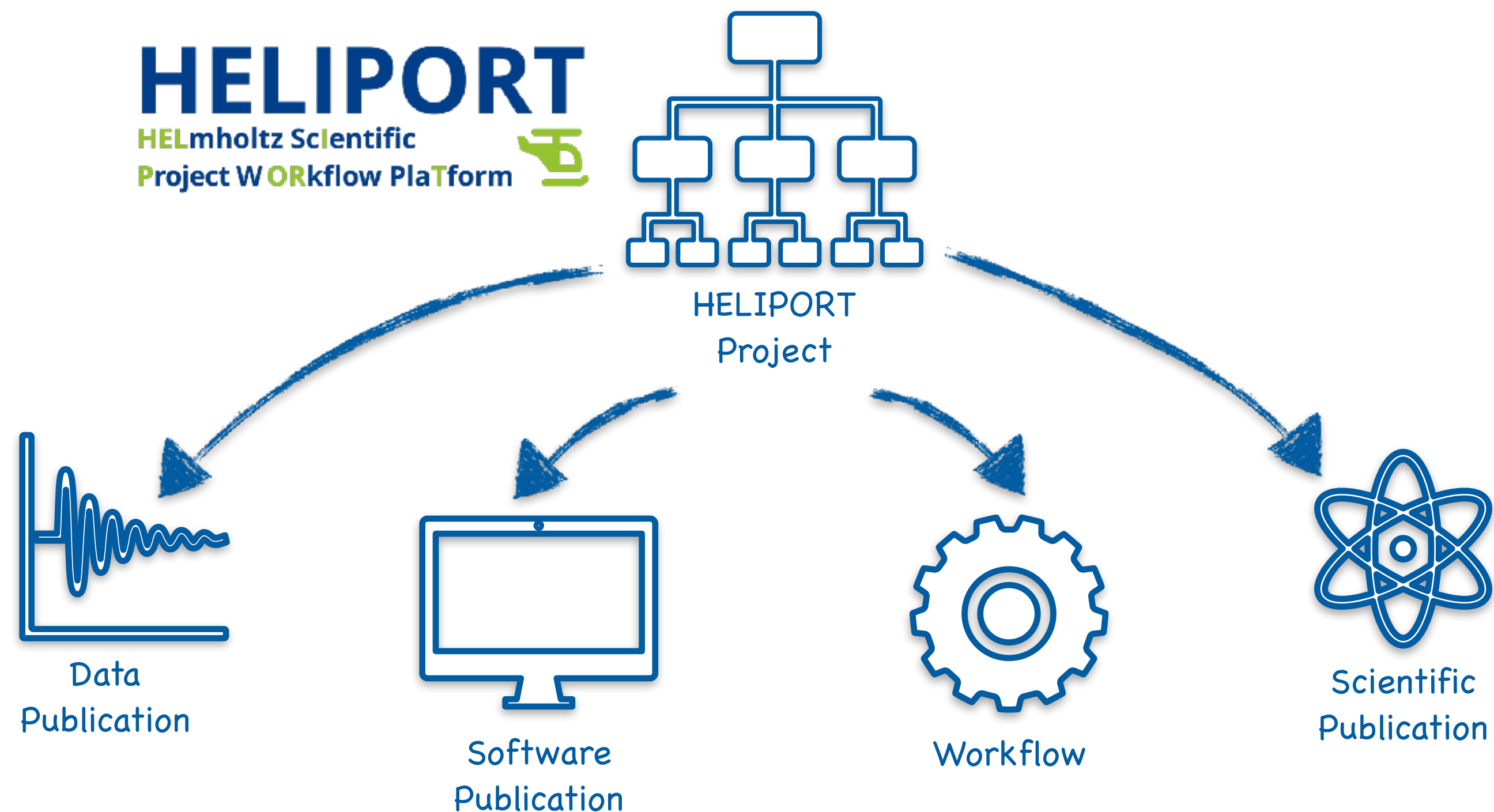
Resources

Website: heliport.hzdr.de

Repository: codebase.helmholtz.cloud/heliport

API Doc: heliport.hzdr.de/redoc/