HELIPORT: An overarching Data Management System at HZDR

HELIPORT HELmholtz ScIentific Project WORkflow PlaTform

Stefan E. Müller, Thomas Gruber, Guido Juckeland, Jeffrey Kelling, Oliver Knodel, Mani Lokamani, Martin Voigt, David Pape

Helmholtz-Zentrum Dresden-Rossendorf - Department of Information Services and Computing

DPG Spring Meeting - Karlsruhe,
March 4, 2024

DRESDEN concept

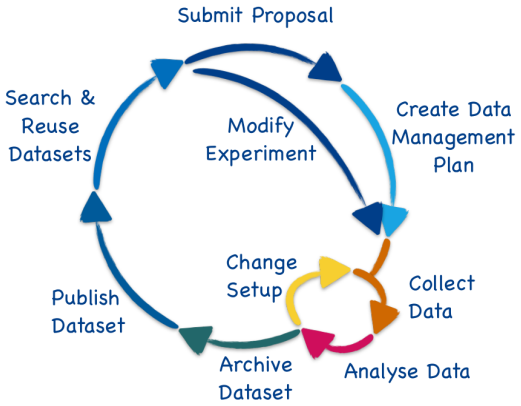
HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

The Helmholtz-Zentrum Dresden-Rossendorf (HZDR)

- About 1470 employees
  - $\sim$ 670 scientists
- Research sites:
  - main site in Dresden-Rossendorf
  - additional sites in Grenoble, Freiberg, Görlitz, Leipzig and Schenefeld
- Research fields:
  - Energy, Health and Matter
- Research facilities
  - ELBE - Center for High-Power Radiation Sources
  - Dresden High Magnetic Field Laboratory (HLD)
  - Ion Beam Center (IBC)

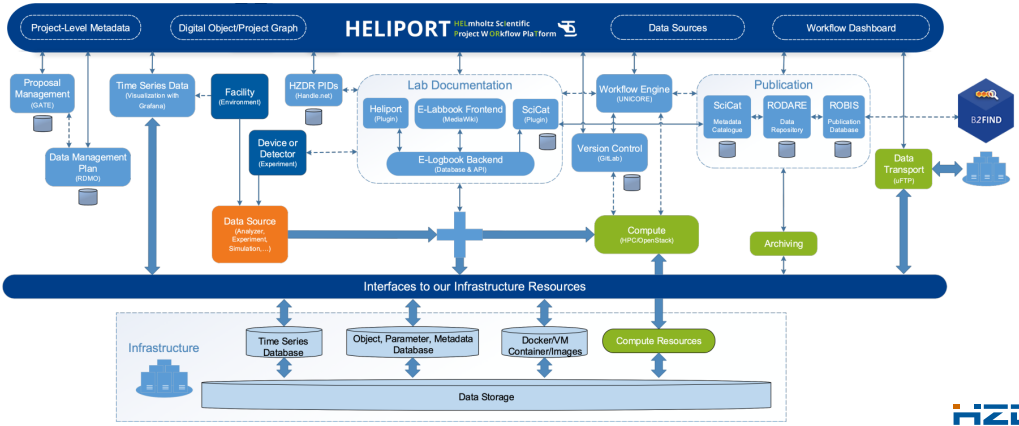
End-to-End Digital Data Lifecycle

- Many tools to support the individual steps of the different research experiments:
  - Electronic lab notebooks
  - Interactive analysis
  - **FAIR publication** of data sets (HZDR's **RODARE** repository)
  - Scientific **workflow** management
  - **Handle** (PID) generation and management
- Uniform and smooth access to and **between** all services and systems is necessary
- Documentation of all the linked resources is essential to create a **comprehensible** and **FAIR** data lifecycle
  - In accordance with the **HZDR Data Policy**



HELIPORT as an overarching guidance system

There is a need to support the entire experiment with reliable **interconnected tools** to enable **FAIR** science. Underlying IT infrastructures are complex, documentation may be missing (lack of time), and often scientists may not know which services are available at facilities and how to use them. An **overarching system** guiding the scientists through the lifecycle of their research project is necessary.



The HELIPORT project

"The **HELIPORT** *project aims at developing a platform which accomodates 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

Features:

- Entry point for experiments and scientific projects
- User and group authorisation/management
- Overview of systems and devices involved in a scientific project
- Provision of metadata from proposal management system
- Registration of and access to site-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** Web API
- Authentication via **Helmholtz ID**

HZDR

The HELIPORT project

"The **HELIPORT** *project aims at developing a platform which accomodates 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:

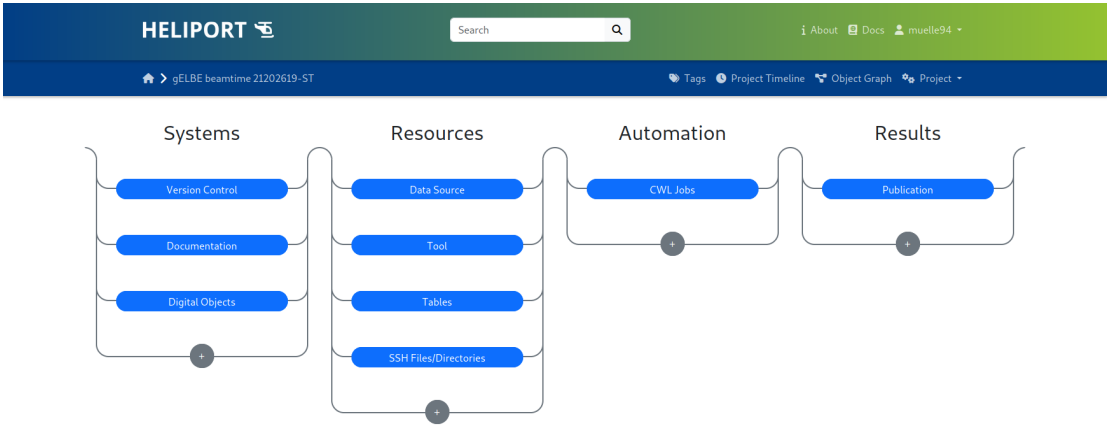
HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

HI JENA Helmholtz Institute Jena

JÜLICH FORSCHUNGSZENTRUM

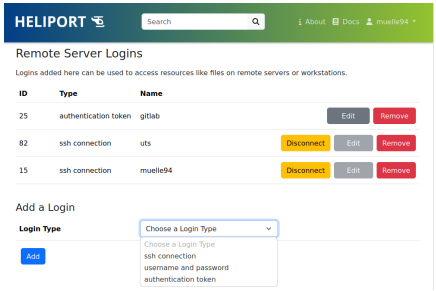
Funded by:

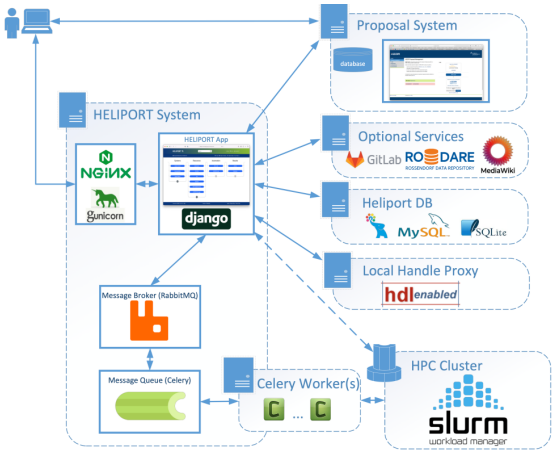
HMC HELMHOLTZ Metadata Collaboration



HELIPORT infrastructure

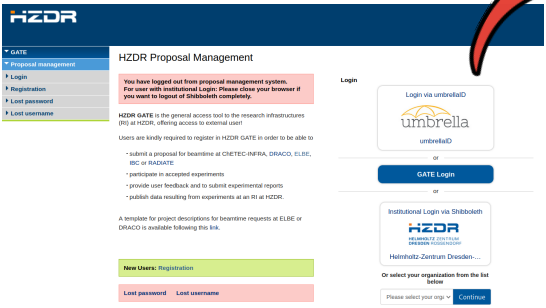
- **HELIPORT** web app is based on **Django**
  - **HELIPORT** communicates with various systems through **Web APIs**
  - Project-level metadata is stored in an SQL database and can be exported in various metadata schemes
- Computational workflows are managed in **HELIPORT** and executed on HPC clusters using **slurm** or **UNICORE**

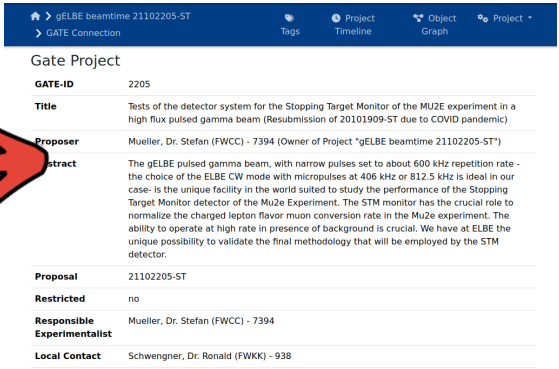




HELIPORT interface to Proposal Management System

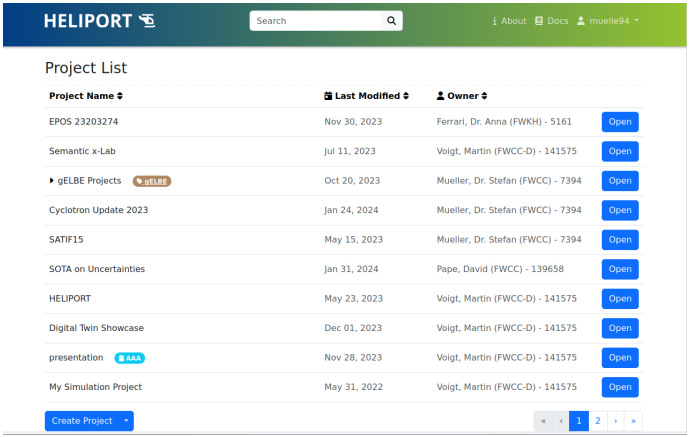
- Automated transfer of project metadata from beamtime proposal management system into **HELIPORT**
  - Title, Authors, Description
  - Beamtime schedule
  - Research facility used





Project list

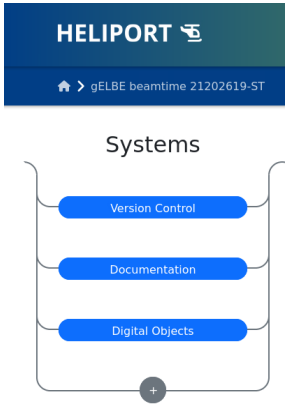
- The owner of a project is typically the corresponding beamline scientist, the project proposer acts as a manager and can add additional project members
- Tags and sub-projects including inheritance are possible in the project list



Systems: Documentation and Code Repositories

The “Systems” section is typically used to refer to all internal and external systems or services which are used:

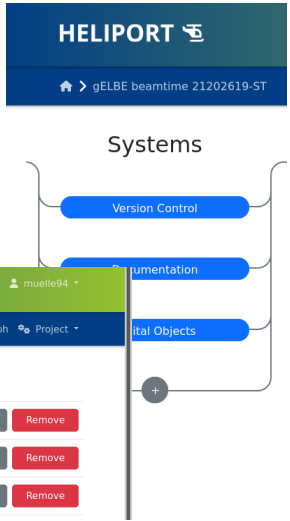
- Electronic Lab Notebooks (Mediawiki, Hedgedoc, Google-Docs,...)
- GitLab, Github, Workflowhub, ...
- Authentication via pre-defined Login-method (ssh, token, username and password)



Systems: Documentation and Code Repositories

The “Systems” section is typically used to refer to all internal and external systems or services which are used:

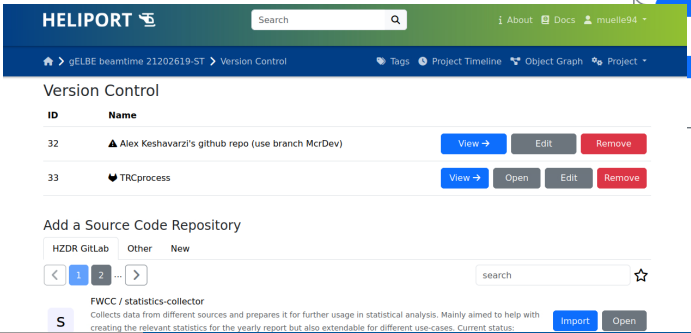
- Electronic Lab Notebooks (Mediawiki, Hedgedoc, Google-Docs,...)
- GitLab, Github, Workflowhub, ...
- Authentication via pre-defined Login-method (ssh, token, username and password)



Systems: Documentation and Code Repositories

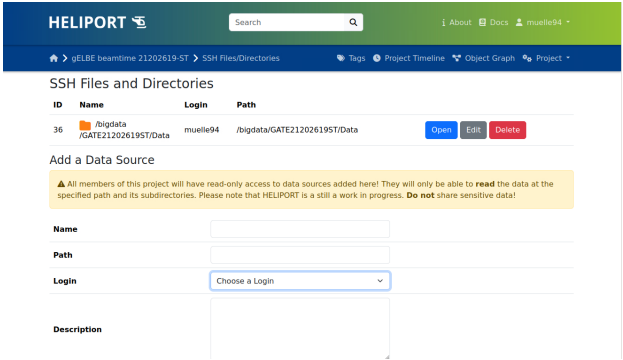
The “Systems” section is typically used to refer to all internal and external systems or services which are used:

- Electronic Lab Notebooks (Mediawiki, Hedgedoc, Google-Docs,...)
- GitLab, Github, Workflowhub, ...
- Authentication via pre-defined Login-method (ssh, token, username and password)



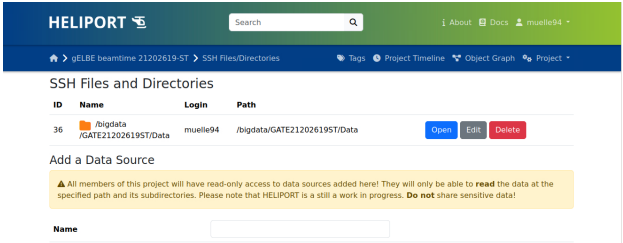
Data resources

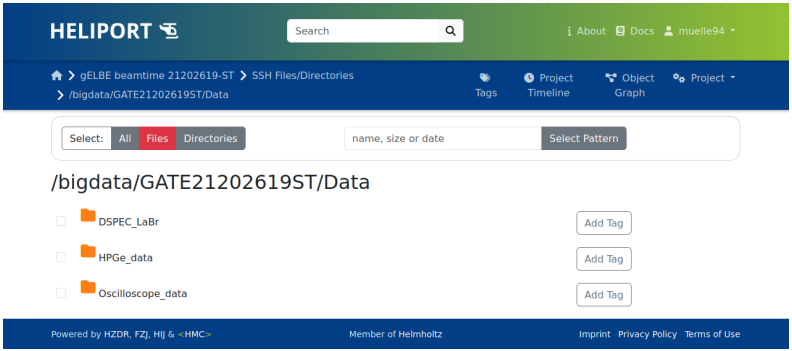
- Folders and files in site-internal filesystems can be registered in **HELIPORT** as **data source**
- Each **project member** has read-only access to the files and folders using the stored login credentials of the **HELIPORT** project
- The provenance of the data sets generated from an experiment is entirely comprehensible



Data resources

- Folders and files in site-internal filesystems can be registered in **HELIPORT** as **data source**
- Each **project member** has read-only access to the files and folders using the stored login credentials of the **HELIPORT** project
- The provenance of the data sets generated from an experiment is entirely comprehensible

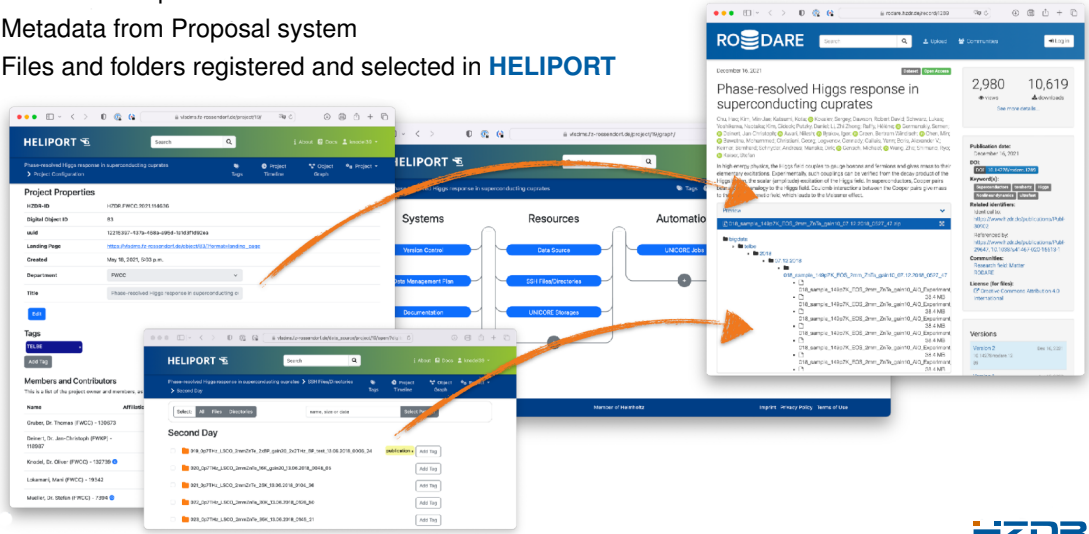




Integration in Overall Publication Workflow

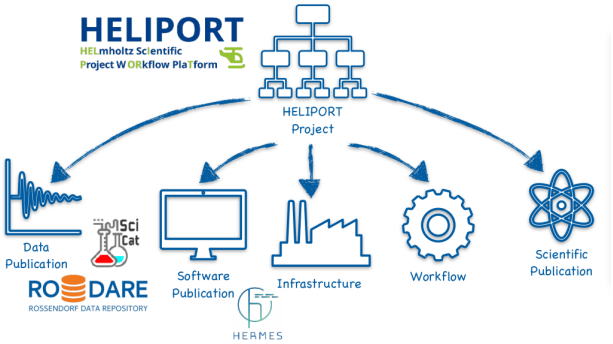
Automated data publication with:

- Metadata from Proposal system
- Files and folders registered and selected in **HELIPORT**



Conclusions

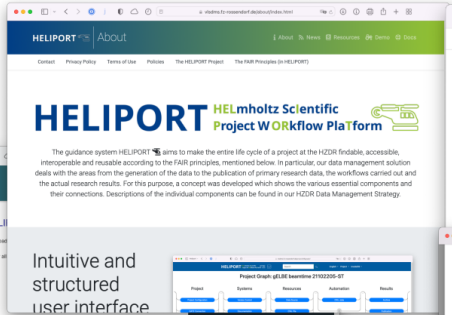
- The **HELIPORT** system allows to describe and collect metadata from services and systems involved in a scientific experiment from the initial proposal to the final publication and eventual data reuse
- This is very important to provide **FAIR** and **comprehensible** research projects
- Metadata is shared between services and systems by dedicated **interfaces** (APIs)

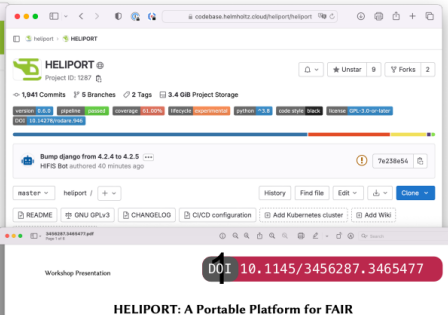


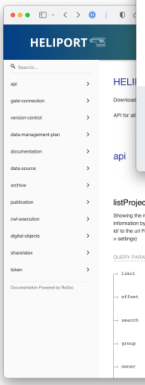
Resources

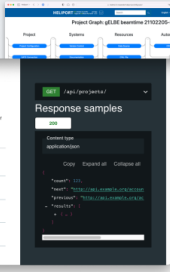
Website: heliport.hzdr.de

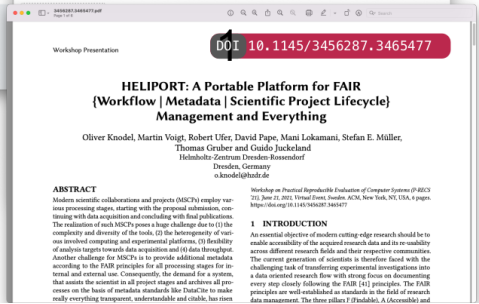
Repository: codebase.helmholtz.cloud/heliport











API doc: heliport.hzdr.de/app/redoc/