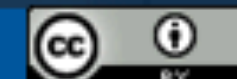


Open Research Project Guidance System: HELIPORT

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



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.

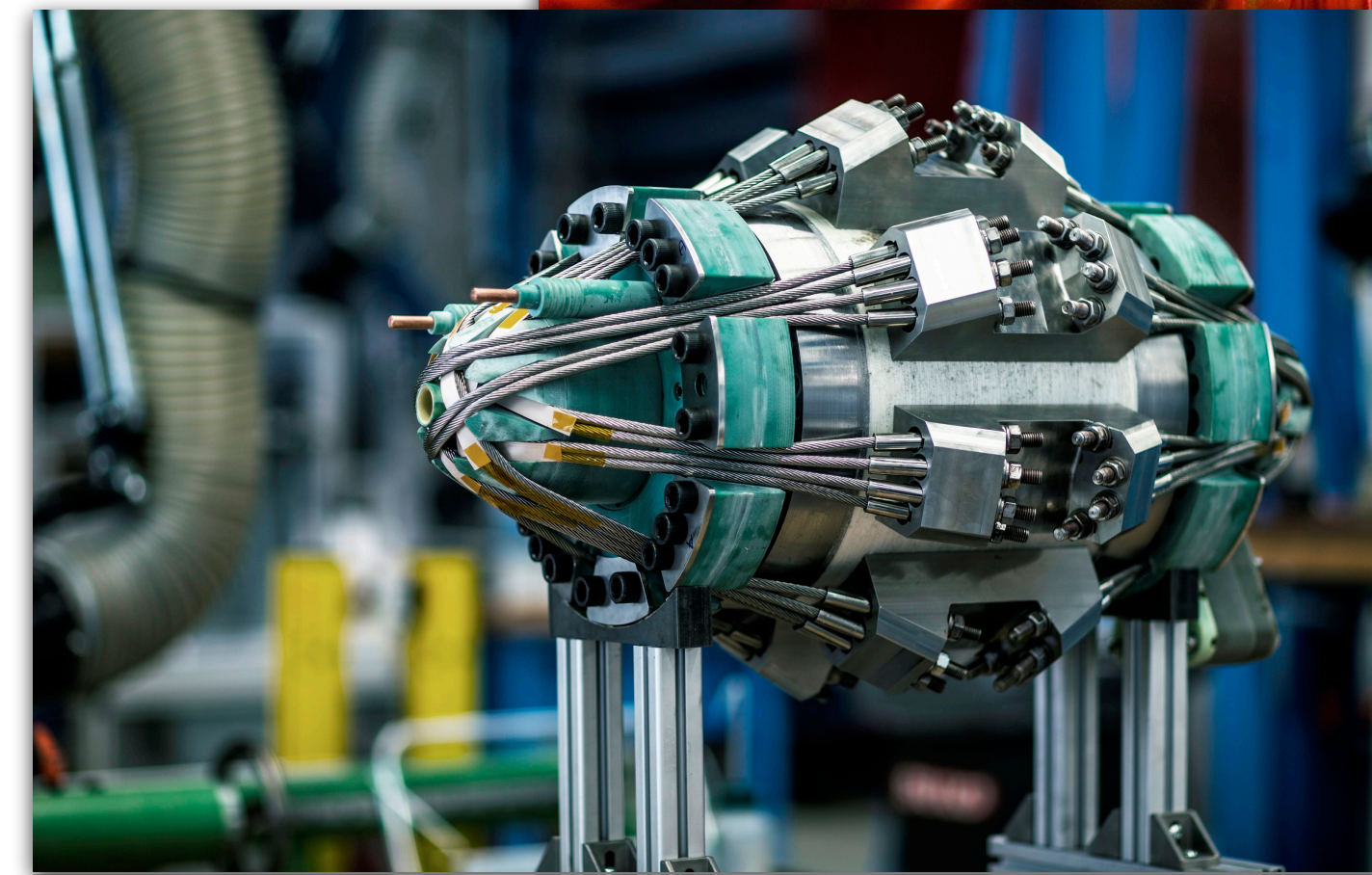
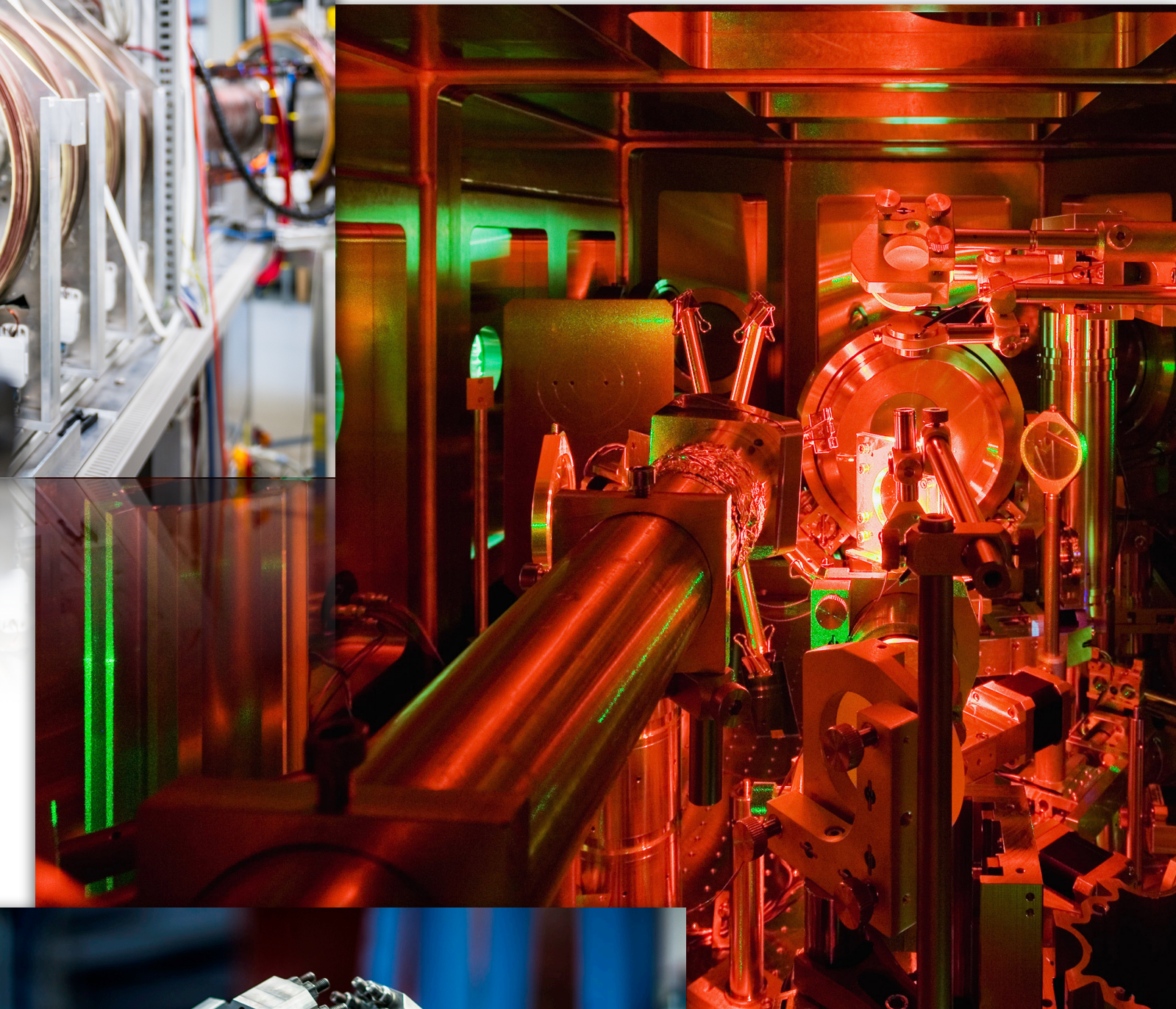
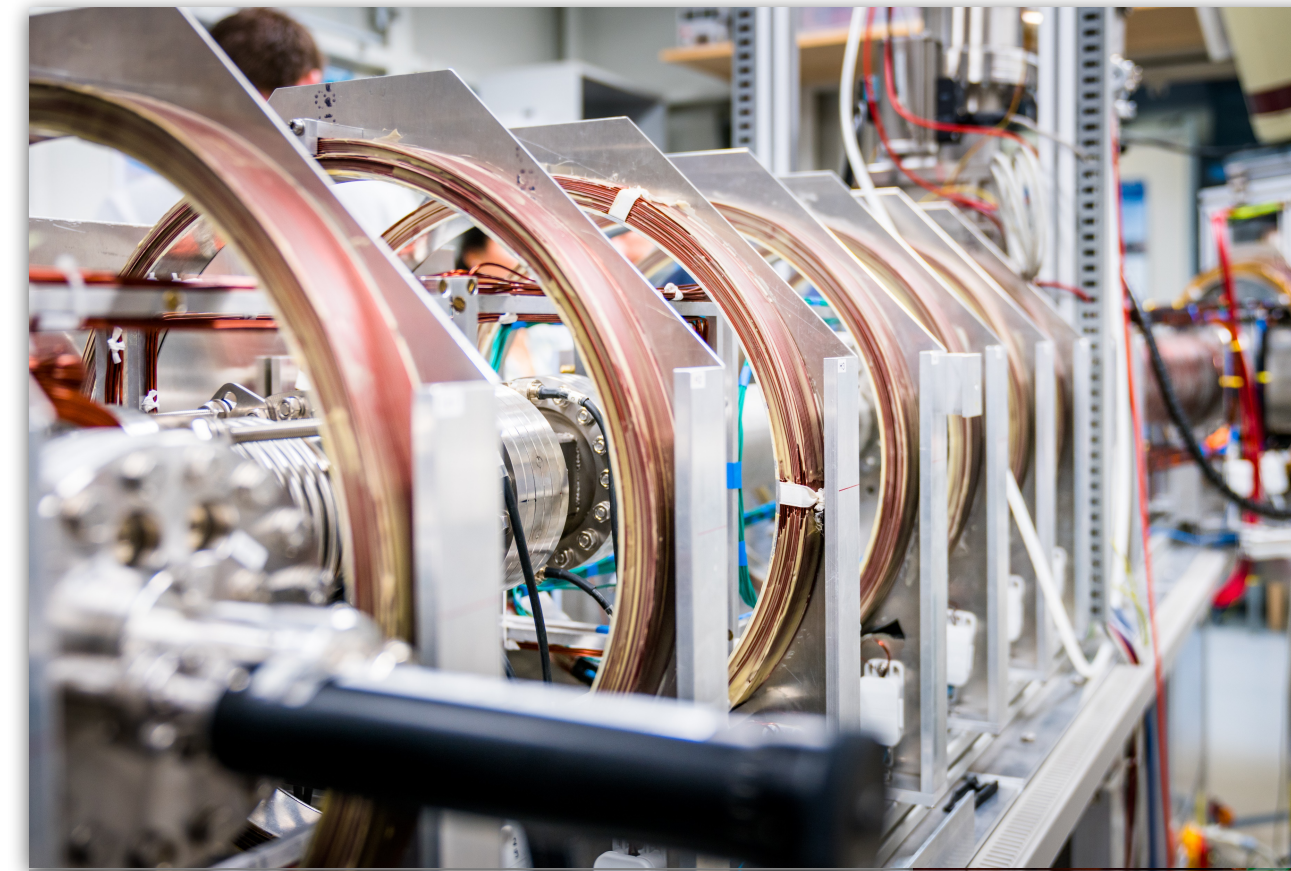
— Positrons, protons, neutrons as well as X-ray and gamma radiation.

Dresden High Magnetic Field Laboratory (HLD)

— Europe's highest pulsed magnetic fields.

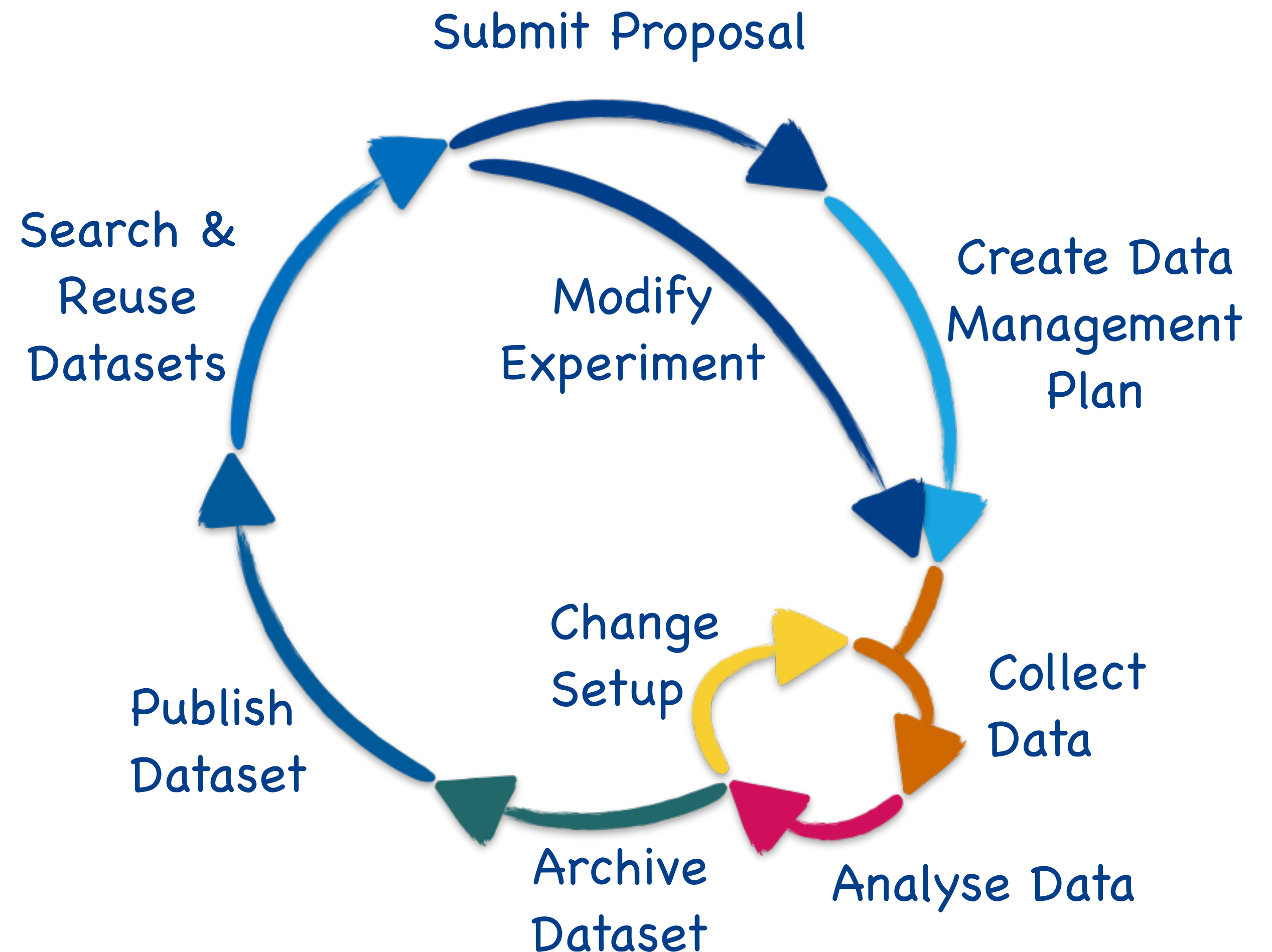
Ion Beam Center (IBC)

— Nanoscale surface analysis and modification.



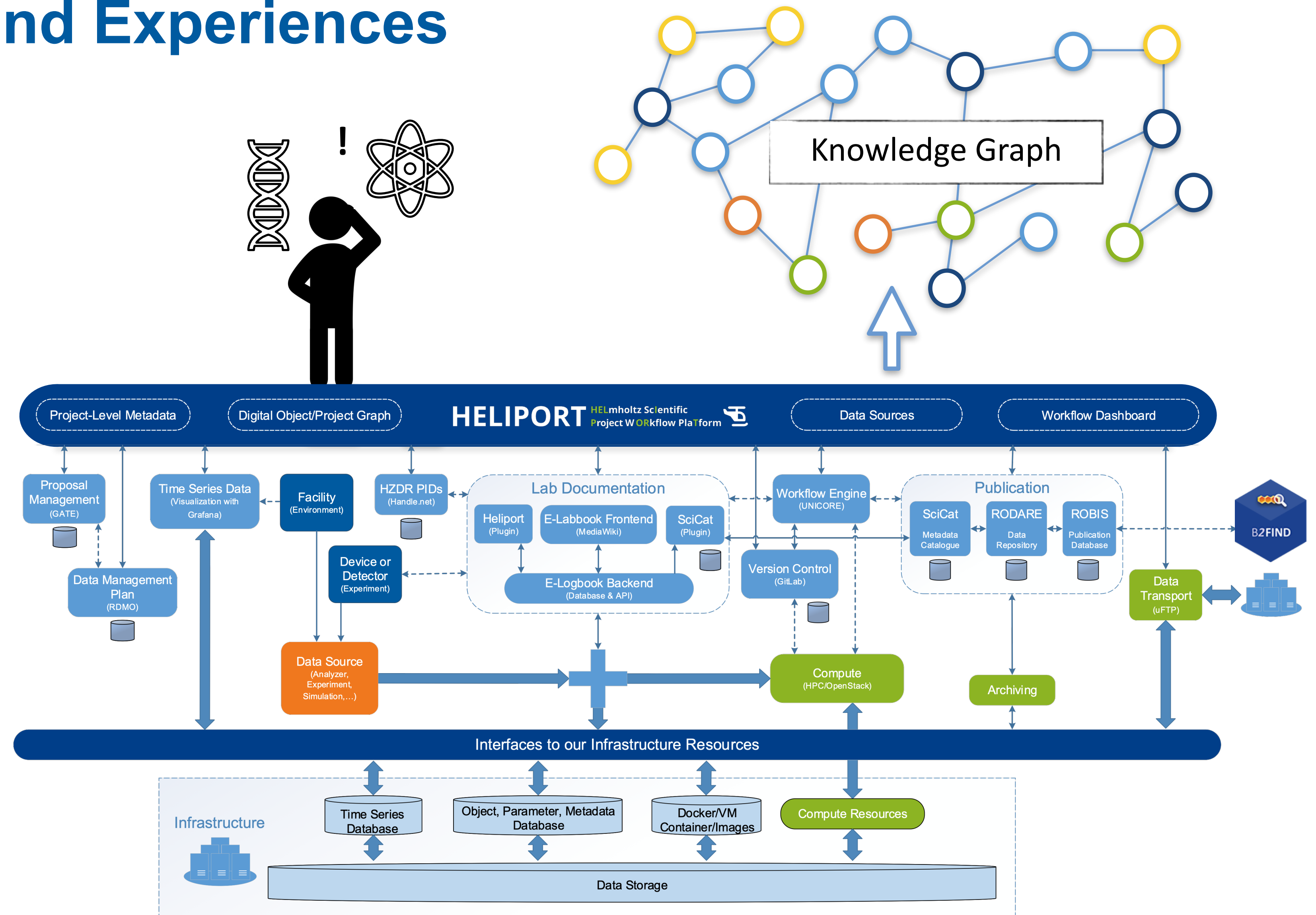
Our Challenge: An End-to-End Digital Data Lifecycle

- We support many steps of our different research experiment (matter, energy and health) with 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!



HELIPORT HELMholtz Scientific Project WOrkflow PlaTform



“ 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 09:14:34.296524+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://ddd",
      "rdfs:label": "ddd"
    }
  ],
  "heliport:has_Archive": [
    {
      "heliport:archive_id": 7,
      "datacite:uri": "https://ddd",
      "heliport:documentation_system": "MediaWiki",
      "datacite:hasDescription": "ddd"
    }
  ],
  "heliport:has_DataSource": [
    {
      "heliport:data_source_id": 11,
      "datacite:uri": "http://ddd",
      "heliport:use_computer": null,
      "rdfs:label": "ddd",
      "datacite:hasDescription": ""
    }
  ]
}

```

Metadata crosswalk to
schema.org
ResearchProject

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)?

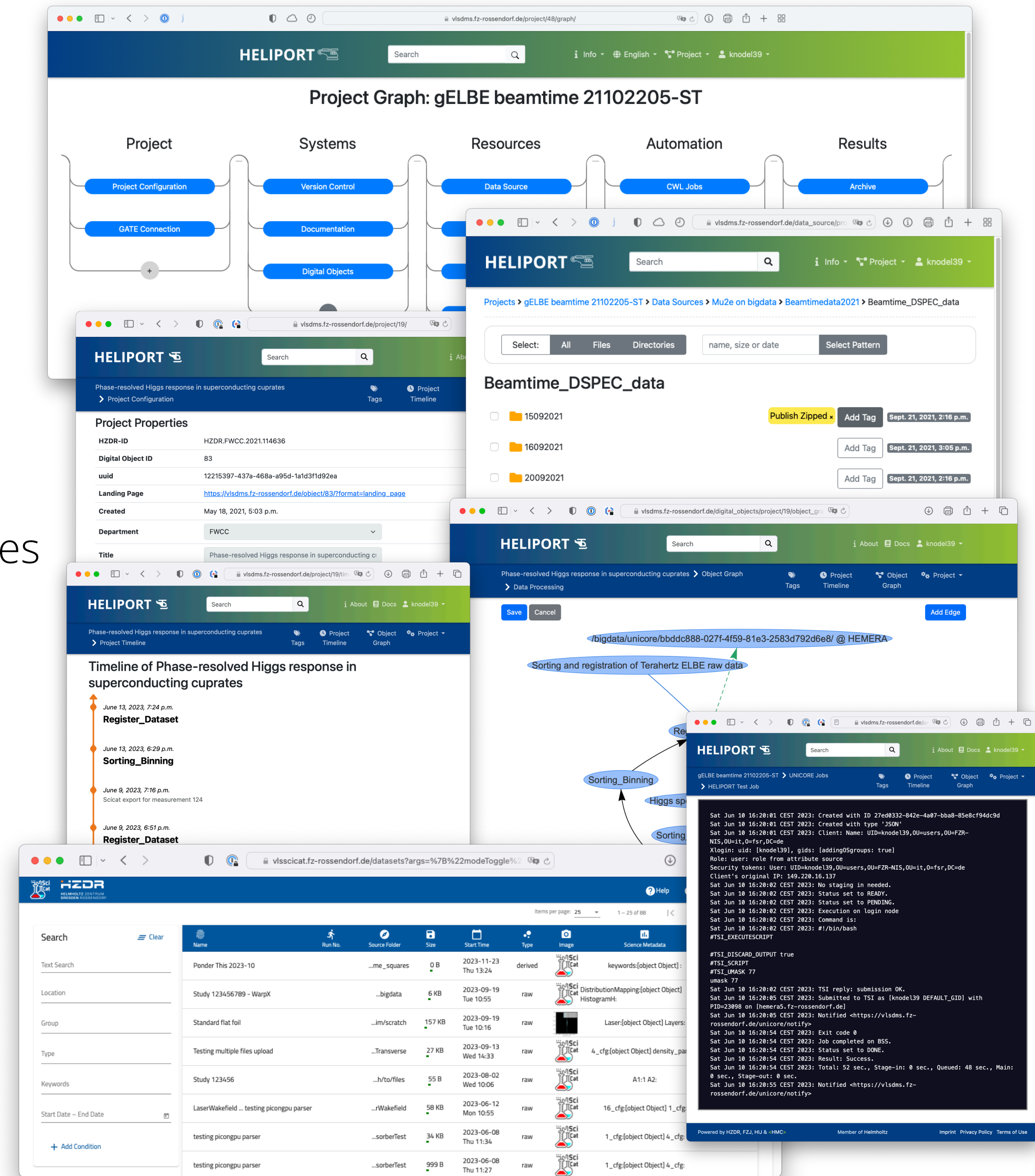
How can we bring **new team members** or external scientists into our project lifecycle and associated services/tools?

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



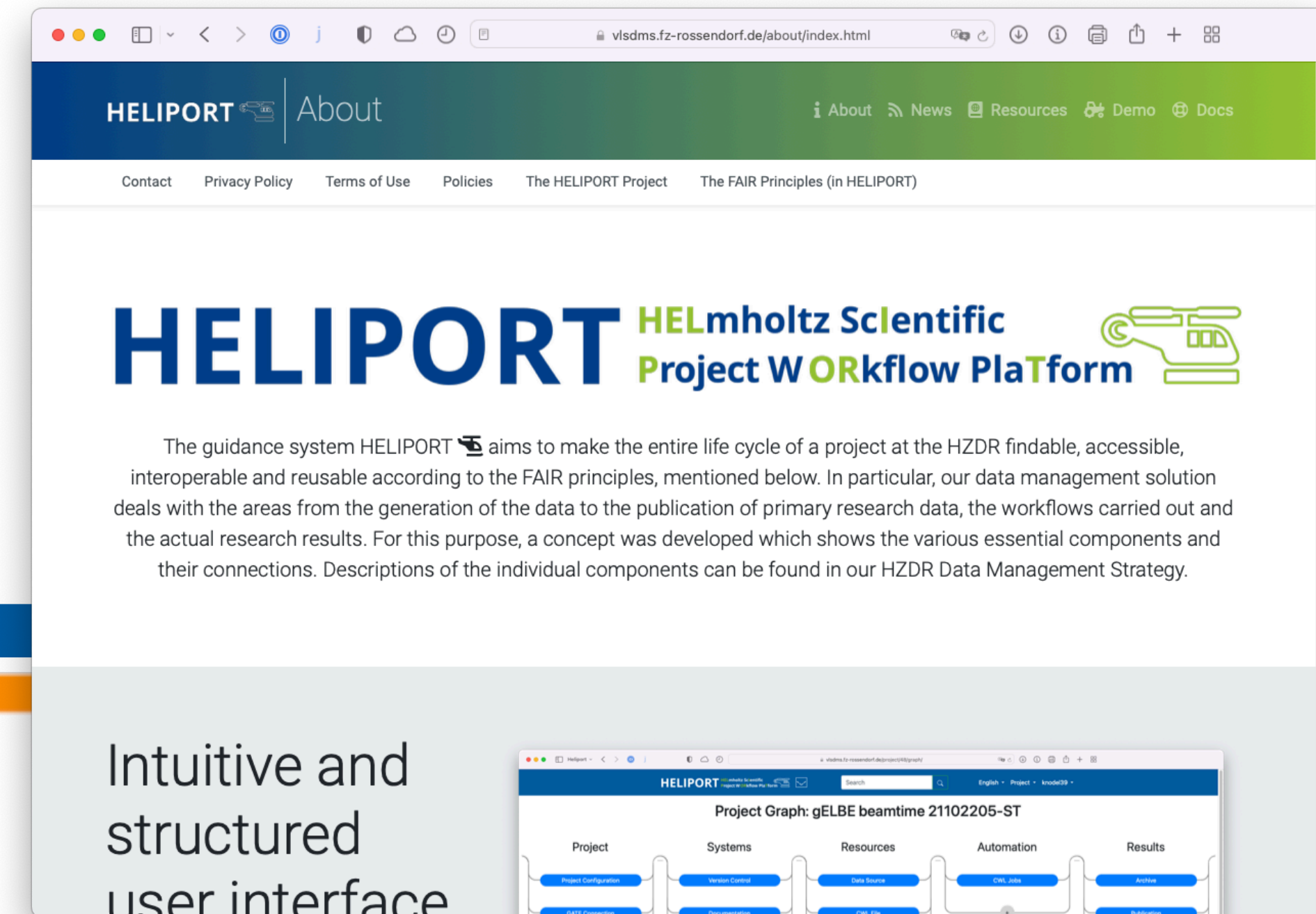
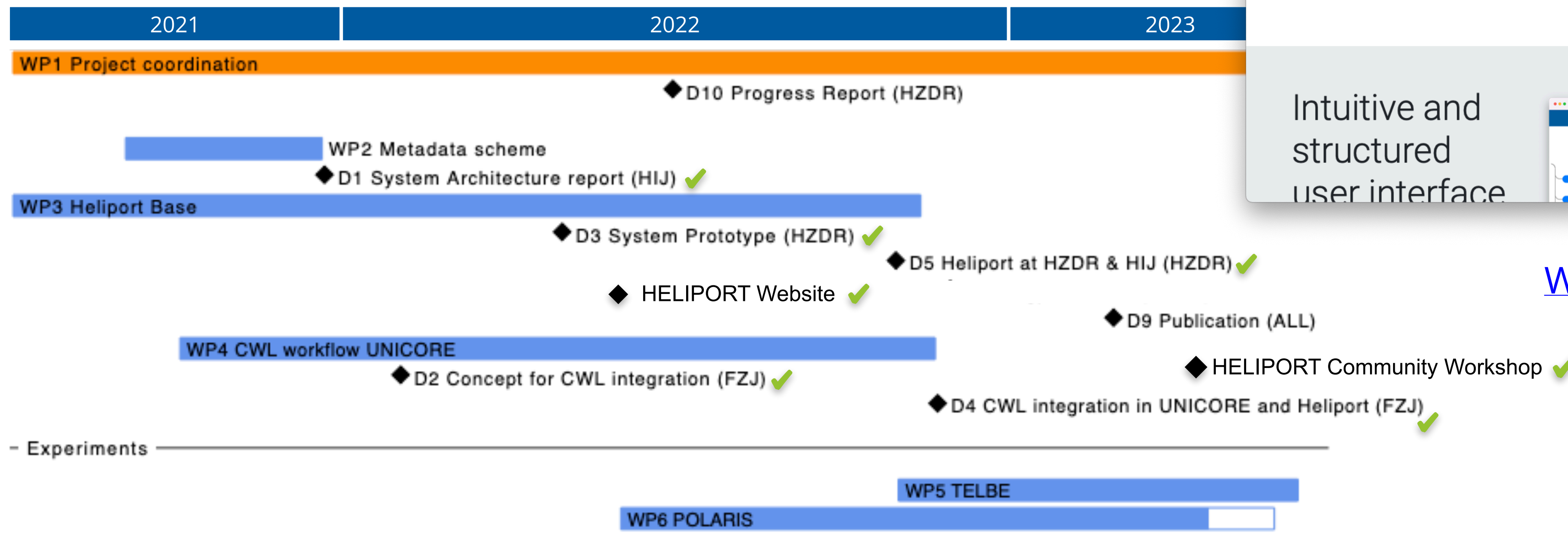
HELIPORT Features

- 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



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:



Intuitive and structured user interface

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



Heliport (Project) Timeline

First Draft: Project Plan (August 2020)

- Project and user management
- Configurable stages
- **REST API** for proposal information
- CWL visualization prototype

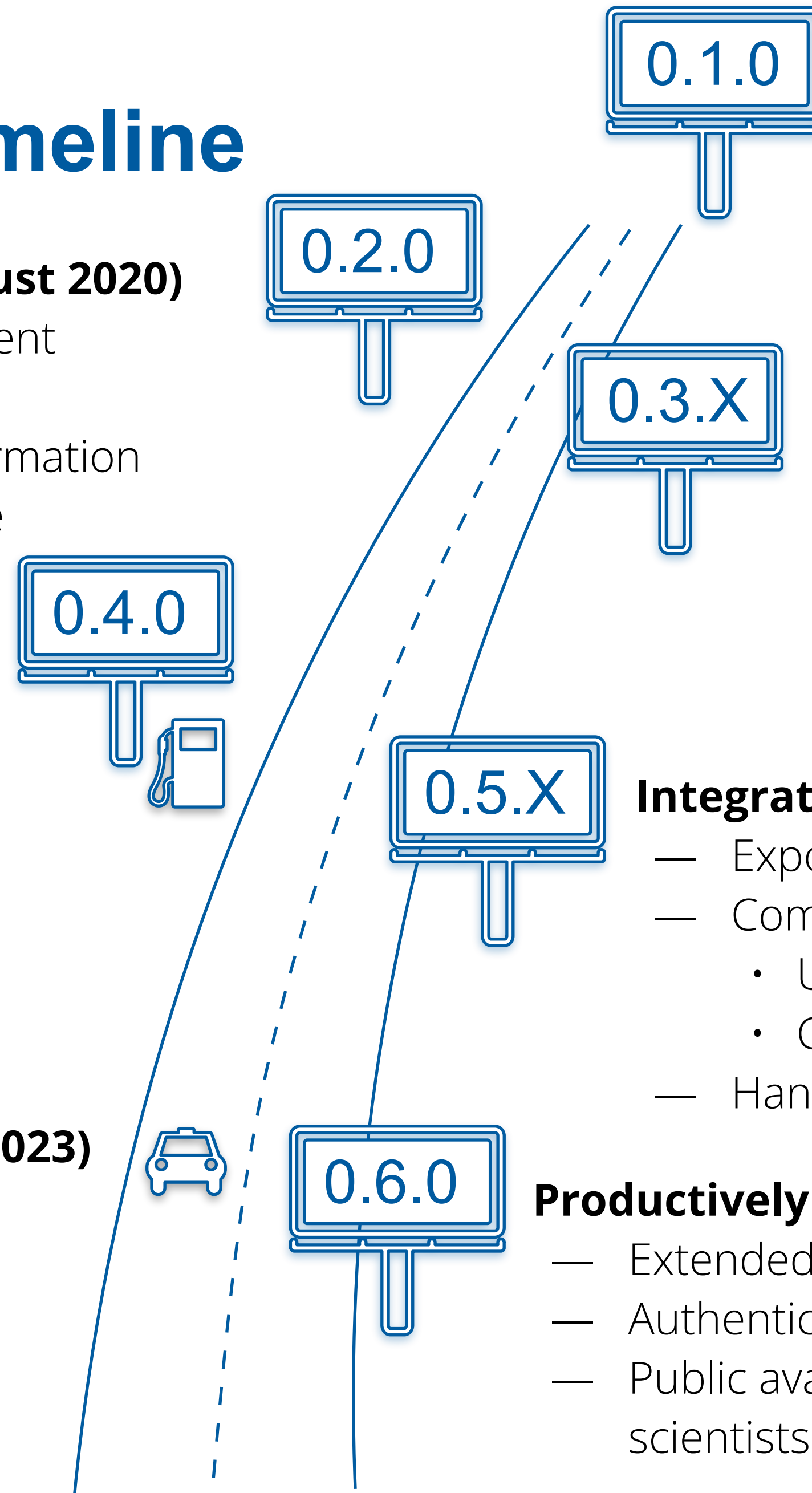
Modular Structure (July 2021)

- Official start of the HMC founded Heliport project:



- Redesign to provide modular and highly configurable system

HELIPORT Community Workshop (July 2023)



Initial Version (June 2020)

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

Improved Project Plan (December 2020)

- Configurable stages and modules
- Infrastructure and database updates
- Daily proposal database update
- Advanced logging and monitoring



Integration of various Apps and Features

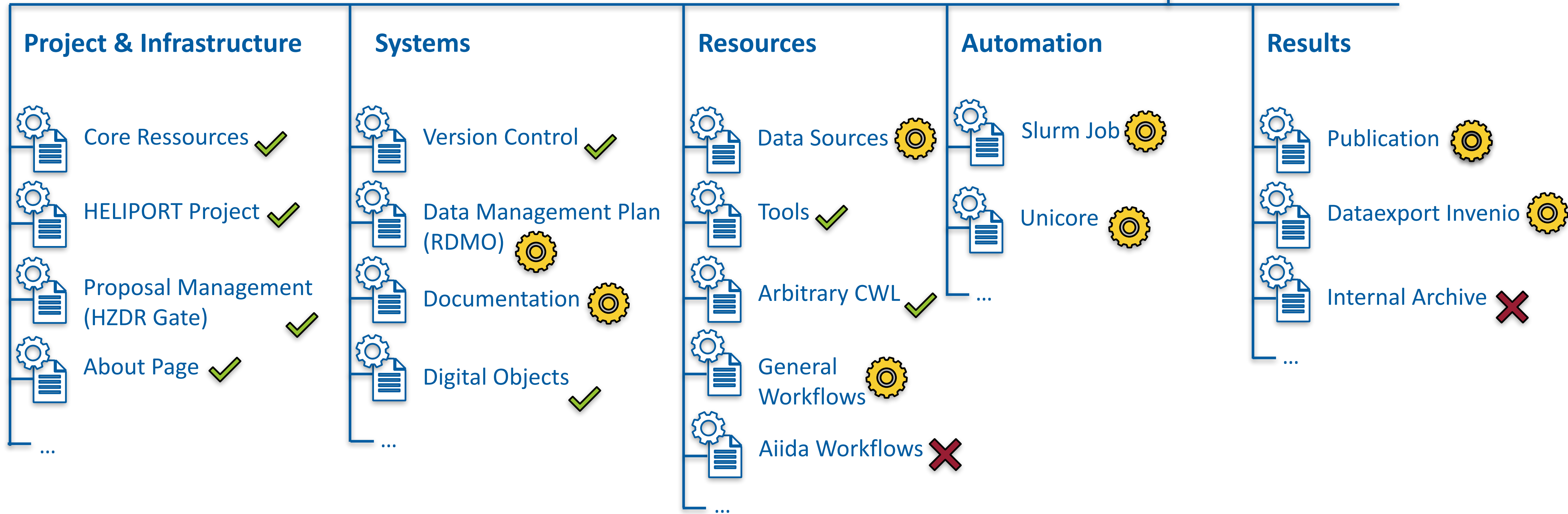
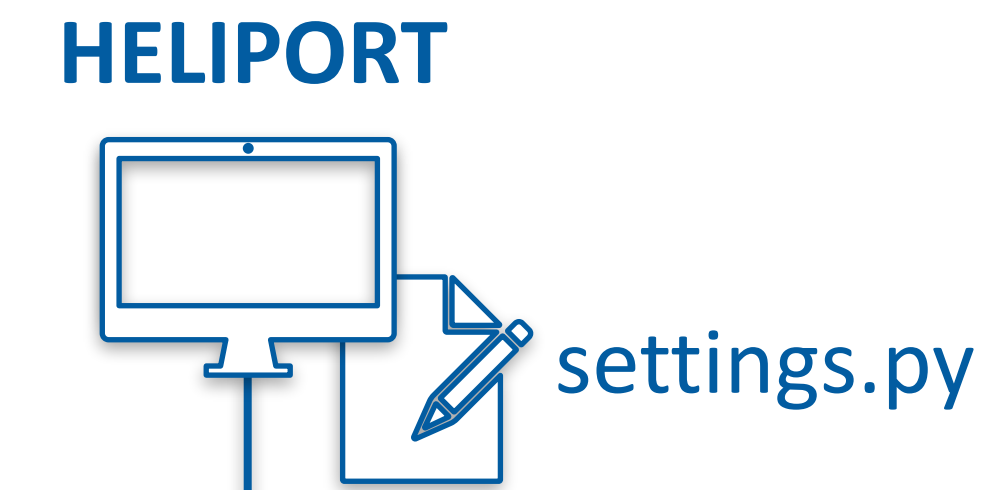
- Export for (different) metadata schemas
- Computational/scientific workflow execution
 - UNICORE support ()
 - Computing job management and monitoring
- Handle management with public landing pages

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

Modular HELIPOINT Design

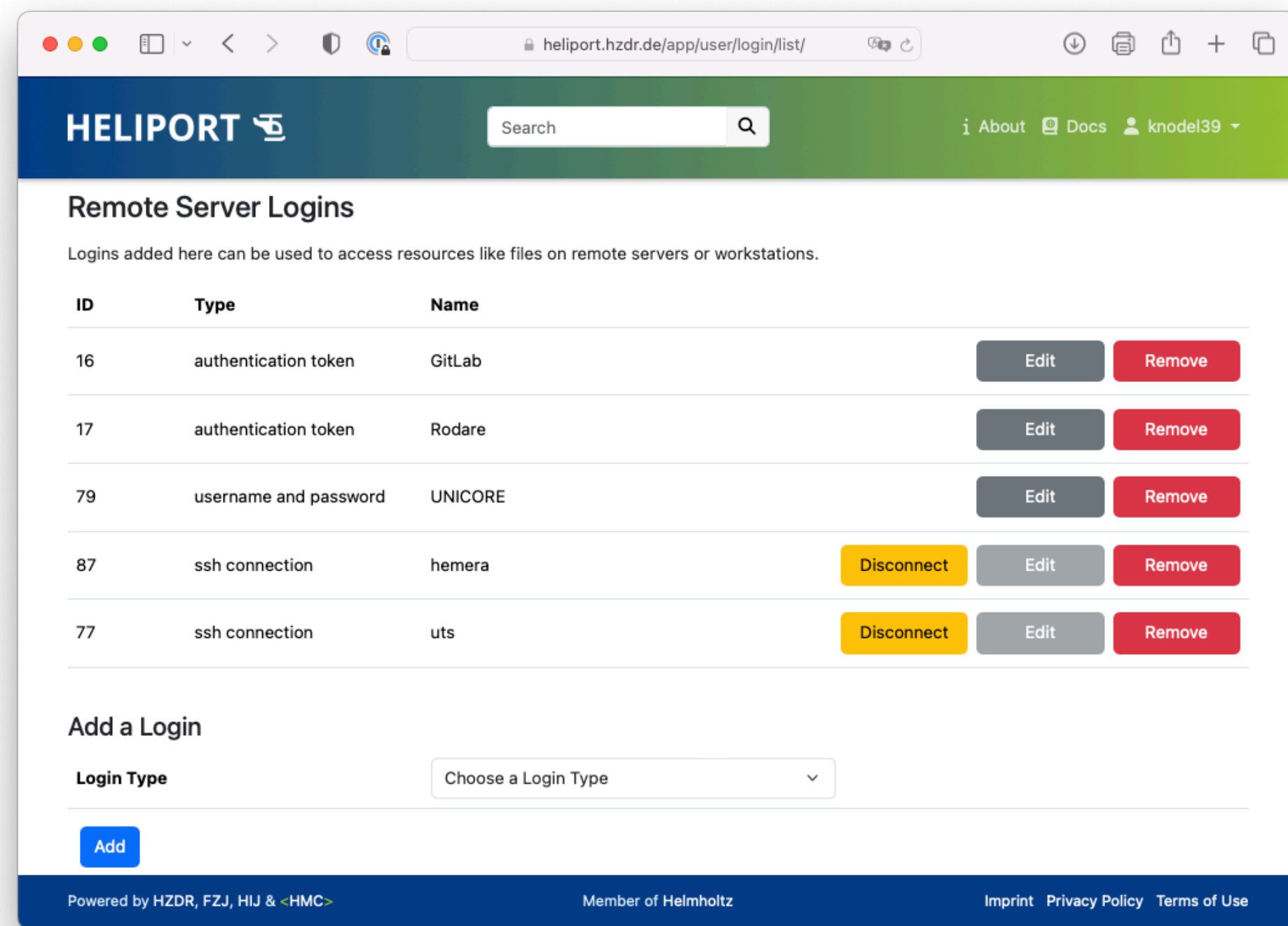
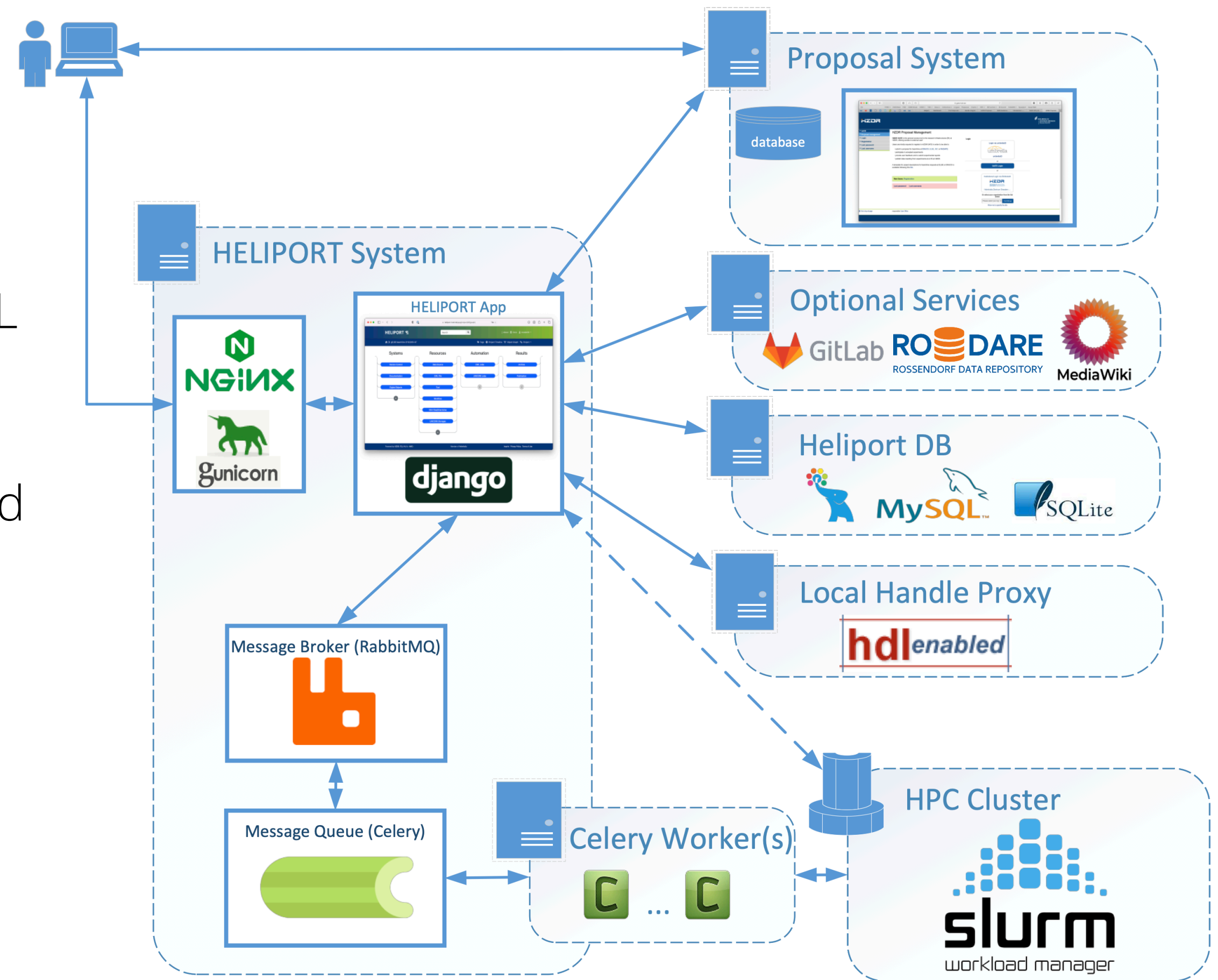
- Plugin architecture based on Django
- All plugins offer a REST API, usable in internal workflows,
- Source code available under GNU GPLv3  and  DOI [10.14278/rodare.947](https://doi.org/10.14278/rodare.947)



 Available
  In development
  Planned

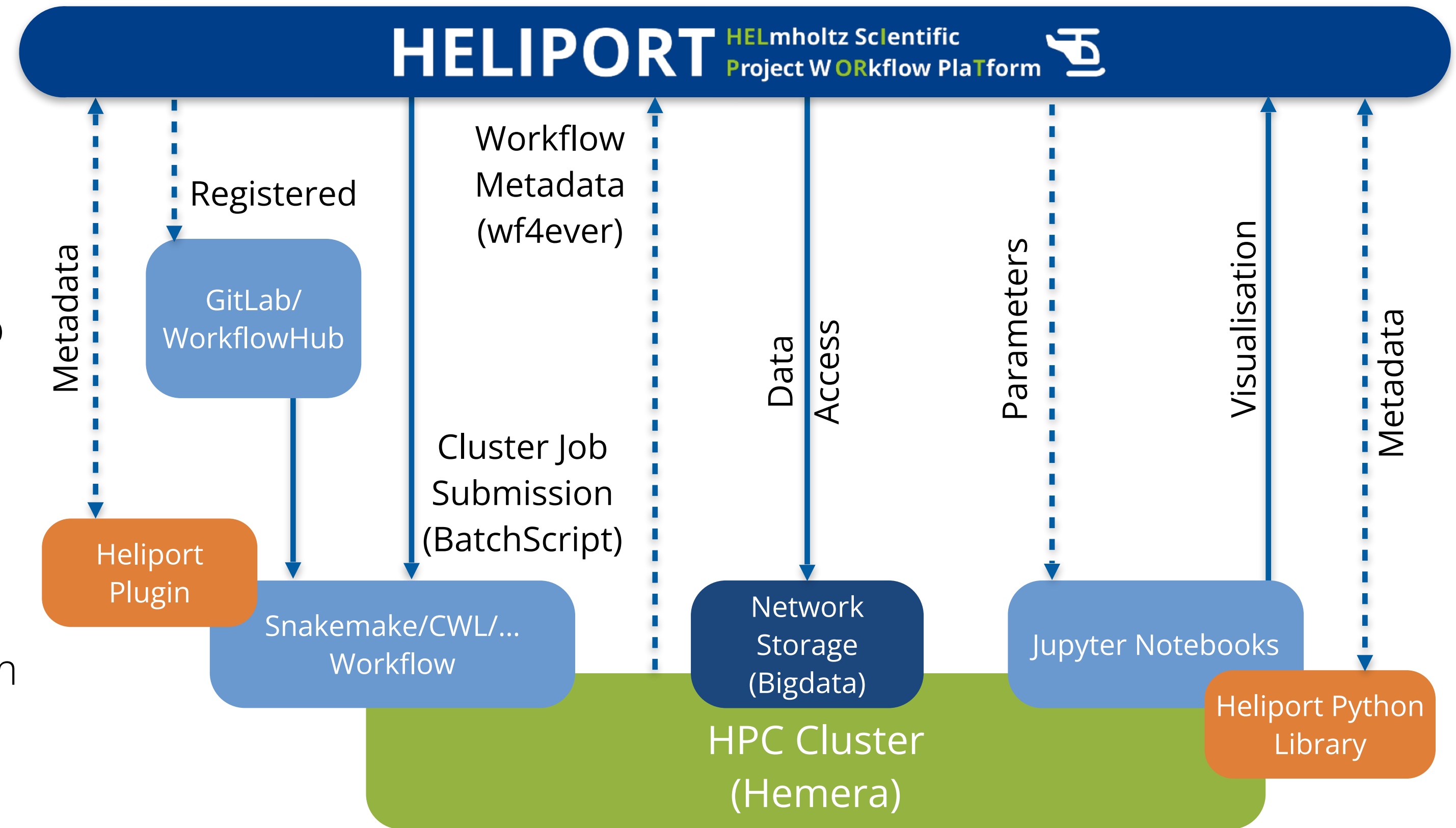
HELIPORT Infrastructure

- The HELIPORT web app is based on Django:
 - Heliport communicates with various systems 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.



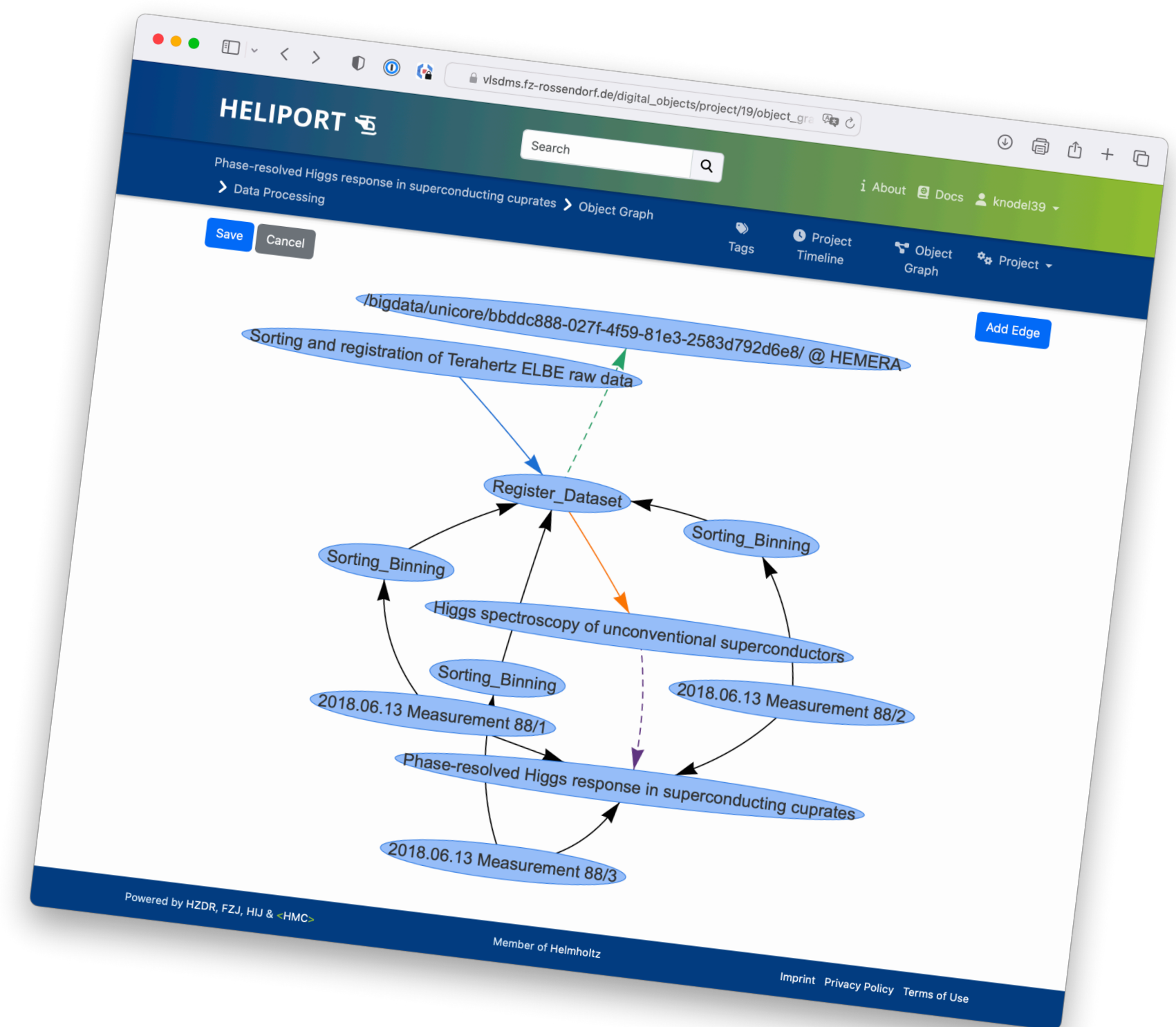
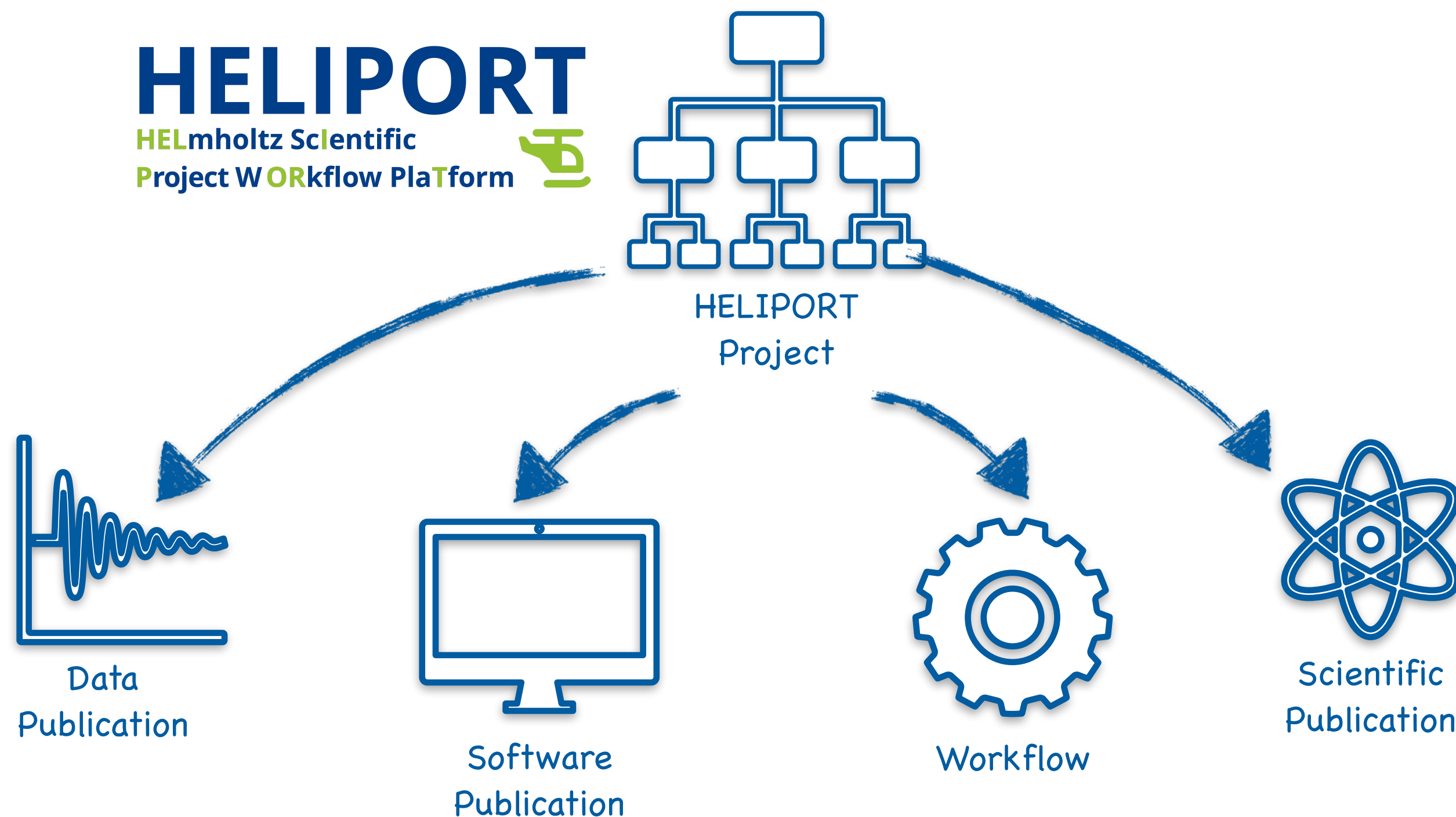
Workflow Architecture (in development)

- 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](https://codebase.helmholtz.cloud/heliport/heliport)



HELIPORT HELMholtz Scientific Project WORKflow PlaTform

The guidance system HELIPORT aims to make the entire life cycle of a project at the HZDR findable, accessible, interoperable and reusable according to the FAIR principles, mentioned below. In particular, our data management solution deals with the areas from the generation of the data to the publication of primary research data, the workflows carried out and the actual research results. For this purpose, a concept was developed which shows the various essential components and their connections. Descriptions of the individual components can be found in our HZDR Data Management Strategy.

HELIPORT Project ID: 1287

1,941 Commits 5 Branches 2 Tags 3.4 GIB Project Storage

version 0.6.0 pipeline passed coverage 61.00% lifecycle experimental python ^3.8 code style black license GPL-3.0-or-later
DOI 10.14278/rodare.946

Bump django from 4.2.4 to 4.2.5
HIFIS Bot authored 40 minutes ago

master heliport / +

History Find file Edit Clone

README GNU GPLv3 CHANGELOG CI/CD configuration Add Kubernetes cluster Add Wiki

HELIPORT

Search...

- api
- gate-connection
- version-control
- data-management-plan
- documentation
- data-source
- archive
- publication
- cwl-execution
- digital-objects
- sharelatex
- token

Documentation Powered by ReDoc

Intuitive and structured user interface



GET /api/projects/

Response samples

200

Content type: application/json

```

{
  "count": 123,
  "next": "http://api.example.org/accoun",
  "previous": "http://api.example.org/ac",
  "results": [
    + { - }
  ]
}
  
```

Workshop Presentation

DOI 10.1145/3456287.3465477

HELIPORT: A Portable Platform for FAIR {Workflow | Metadata | Scientific Project Lifecycle} Management and Everything

Oliver Knodel, Martin Voigt, Robert Ufer, David Pape, Mani Lokamani, Stefan E. Müller, Thomas Gruber and Guido Juckeland
Helmholtz-Zentrum Dresden-Rossendorf
Dresden, Germany
o.knodel@hzdr.de

ABSTRACT
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

Workshop on Practical Reproducible Evaluation of Computer Systems (P-RECS '21), June 21, 2021, Virtual Event, Sweden. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3456287.3465477>

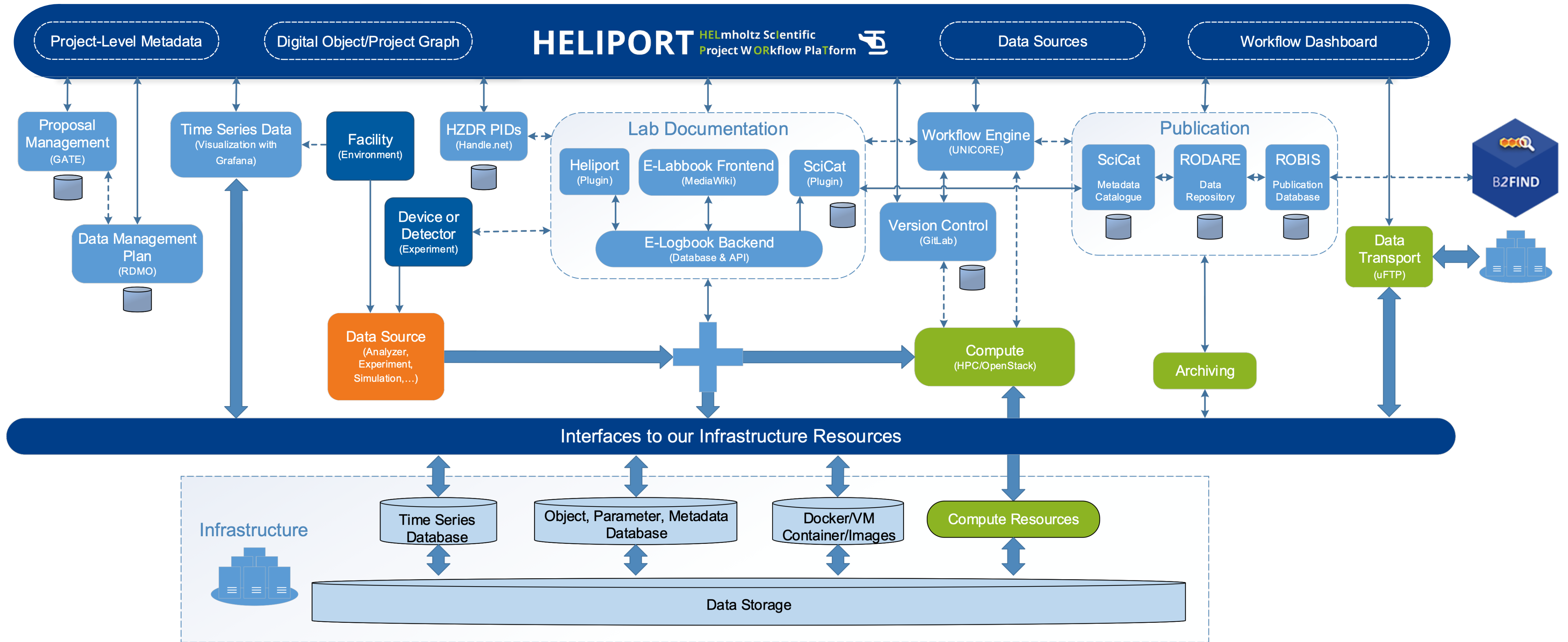
1 INTRODUCTION
An essential objective of modern cutting-edge research should be to enable accessibility of the acquired research data and its re-usability across different research fields and their respective communities. The current generation of scientists is therefore faced with the challenging task 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

API Doc: heliport.hzdr.de/redoc/



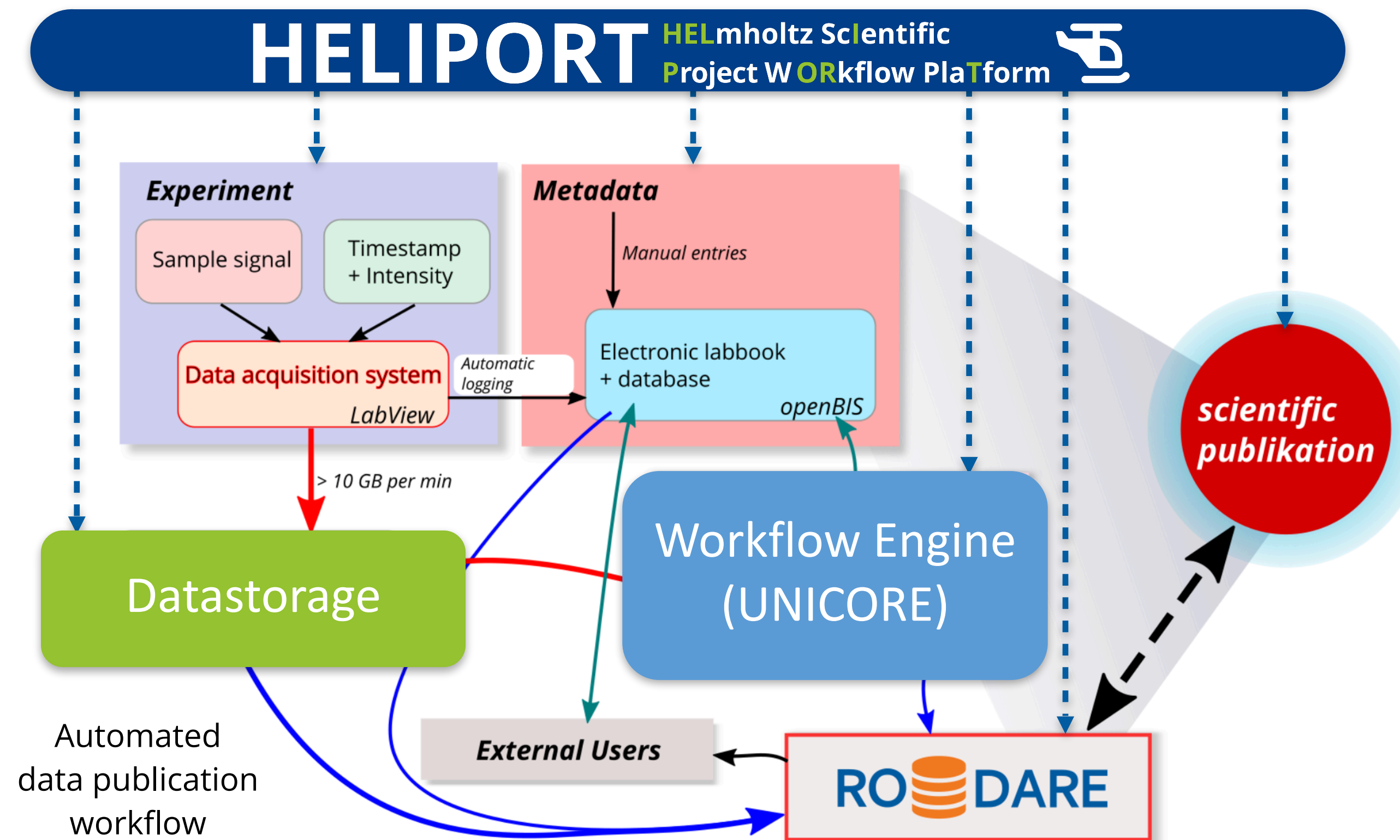
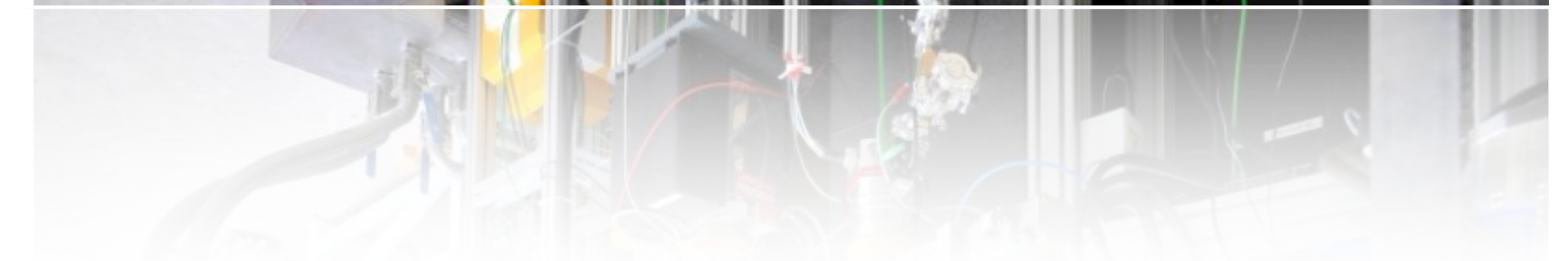
Appendix

HZDR Research (Infrastructure) Landscape

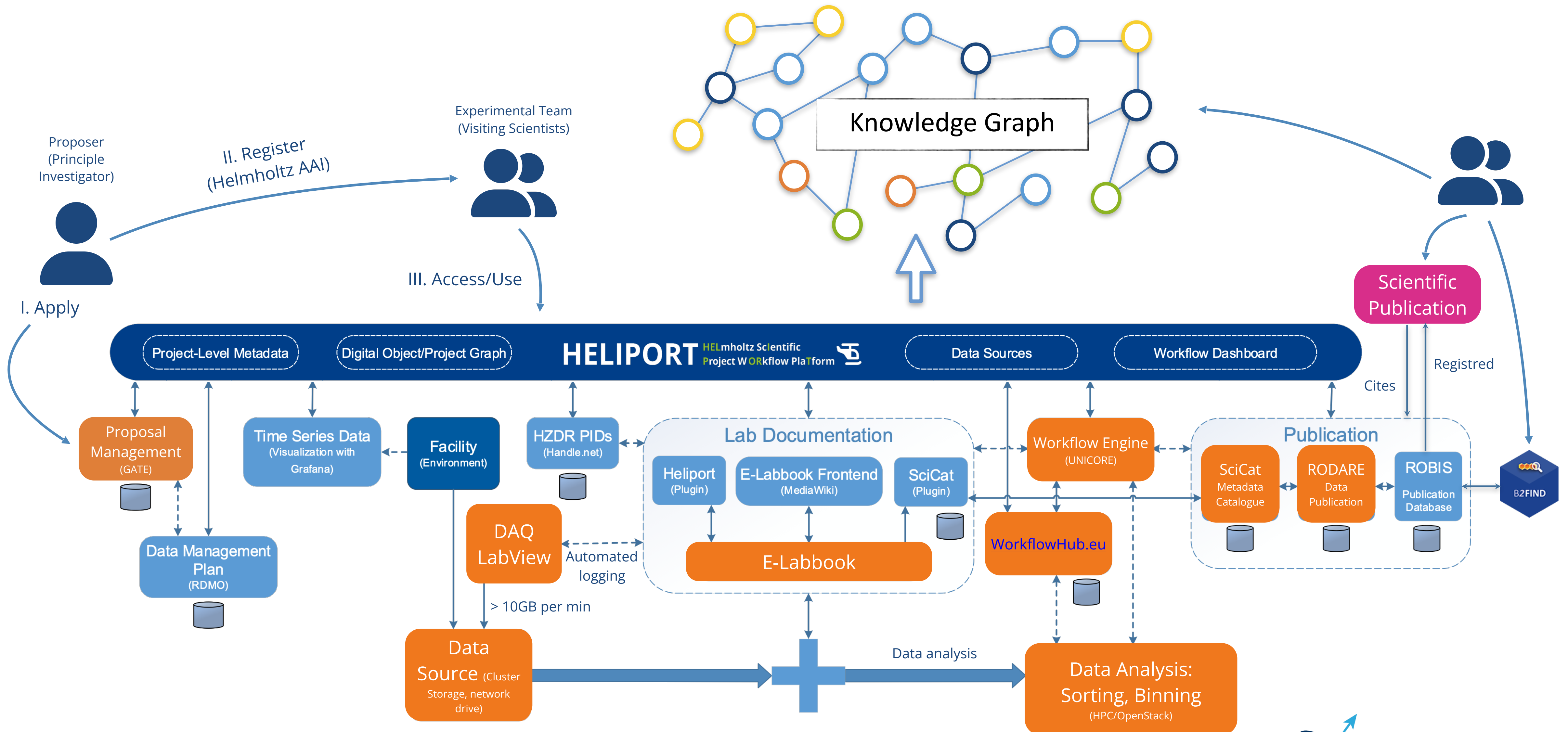


TELBE Data Flow

- 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

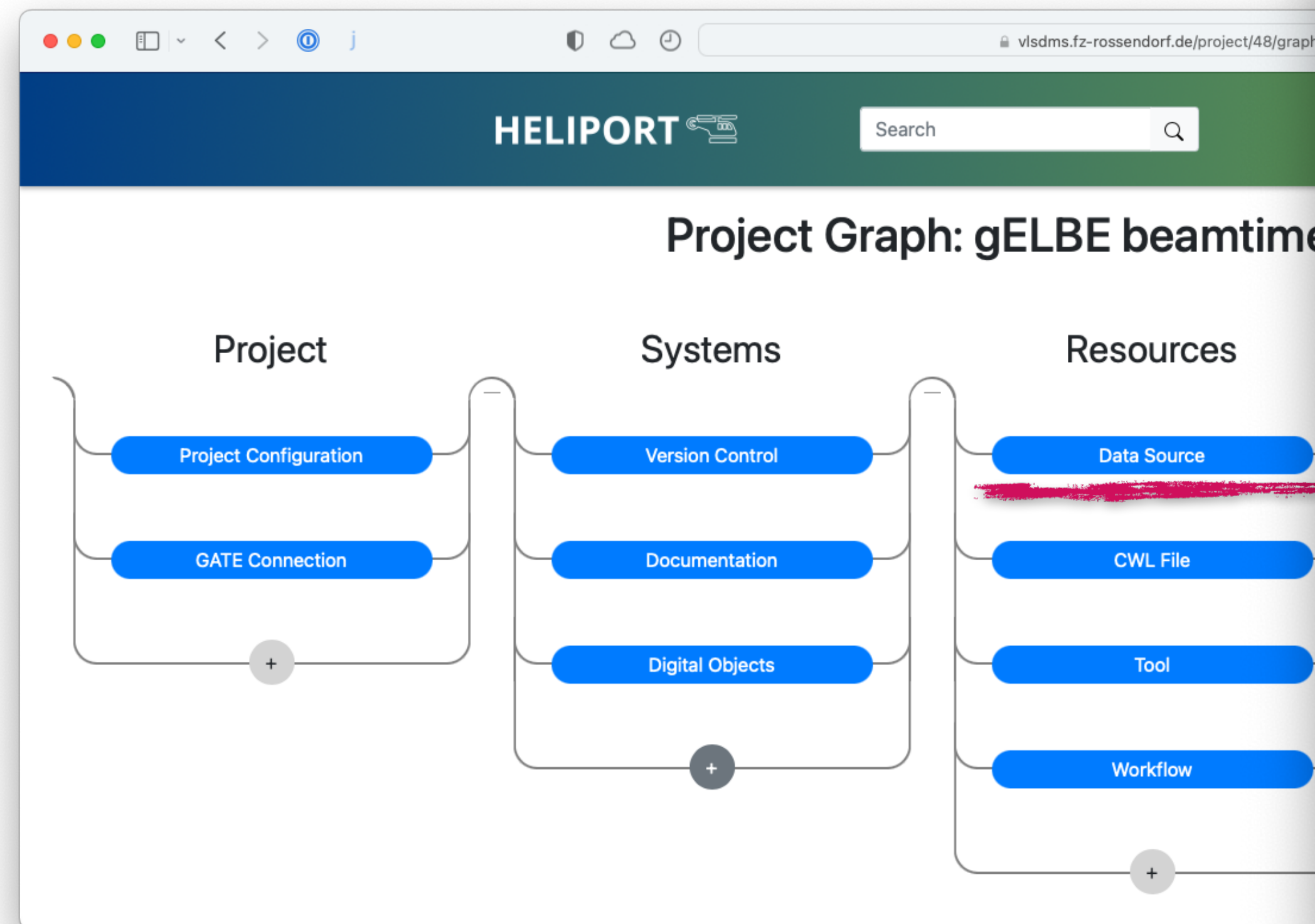


Mapping of the TELBE Resources to HELIPORT



Data Sources can be Registered...

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

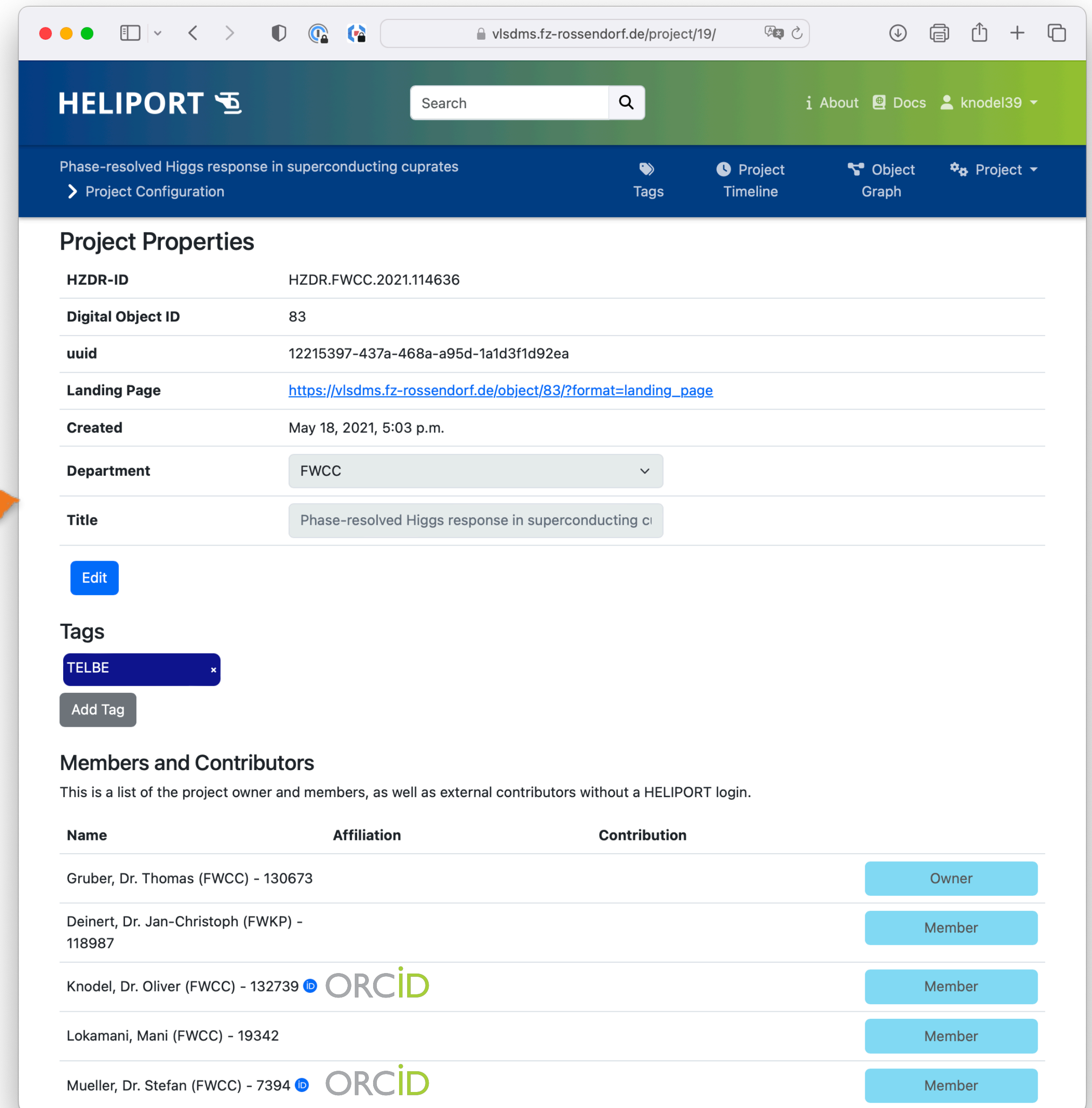
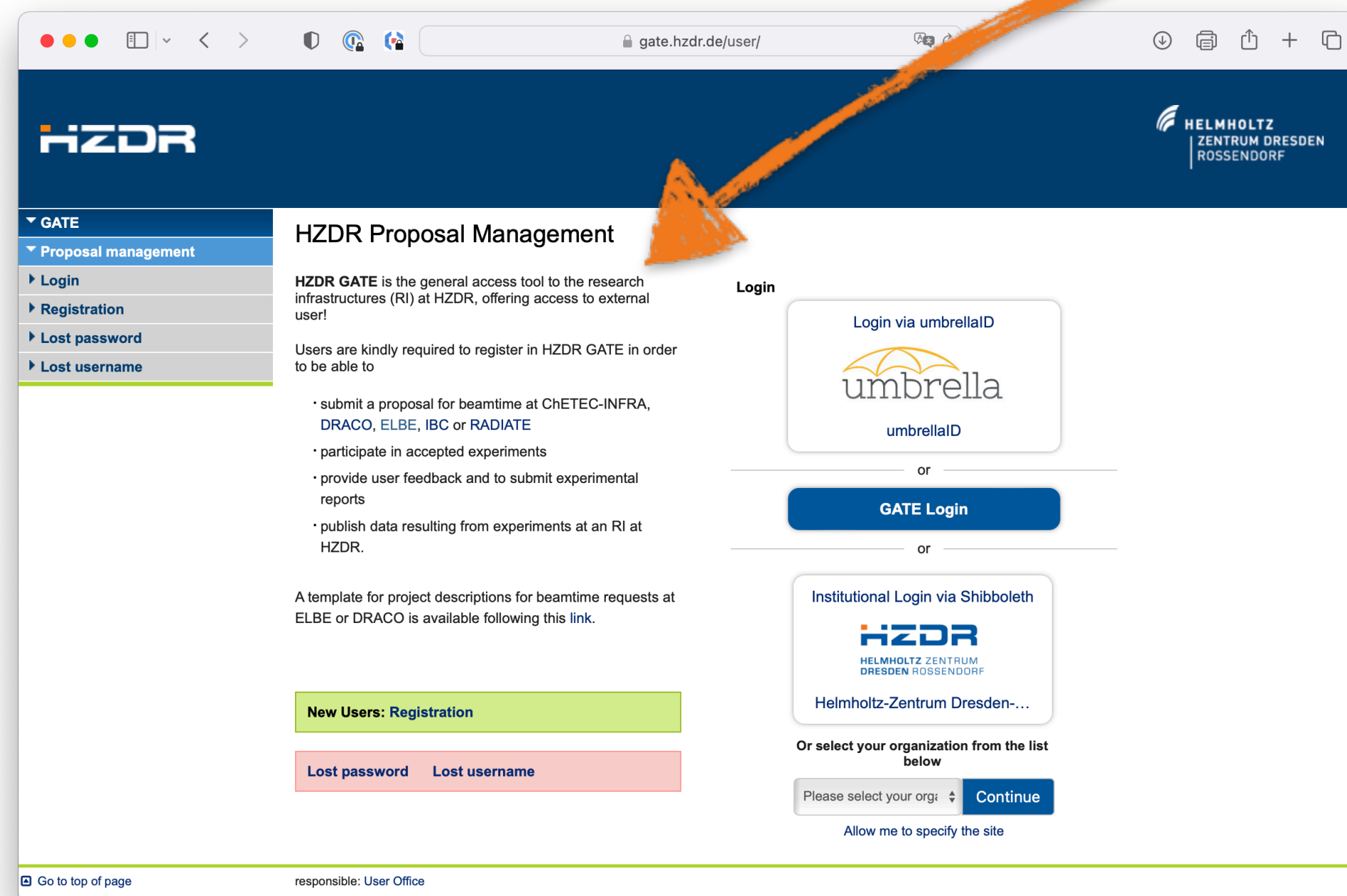


The screenshot shows the HELIPORT web interface. The breadcrumb trail is: Projects > gELBE beamtime 21102205-ST > Data Sources > Mu2e on bigdata > Beamtimedata2021 > Beamtime_DSPEC_data. The interface includes a search bar, a 'Select:' dropdown with options 'All', 'Files', and 'Directories', and a search filter 'name, size or date'. Below this, the 'Beamtime_DSPEC_data' section lists several folders and files with checkboxes, 'Publish' buttons, and 'Add Tag' buttons. The folders are 15092021, 16092021, and 20092021. The files are ELBE TI Current-data-2021-09-17 16_52_59.xlsx, ELBE_2021-09-15.png, and ELBE_2021-09-16.png. The 'Data Source' resource in the diagram is highlighted with a red brush.

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, Owner, and an Open button. The projects listed include Semantic x-Lab, gELBE Projects (with sub-projects like gELBE beamtime 21102205-ST), Example parent project, ML Ops Project, SOTA on Uncertainties, Phase-resolved Higgs response in superconducting cuprates, Digital Twin Showcase, Beamtime Dashboard Test, and Rodare Data Publication Project. A 'Create Project' button and pagination controls are at the bottom.

Project Name	Last Modified	Owner	Open
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 features a workflow diagram with four main stages: Systems, Resources, Automation, and Results. The Systems stage includes Version Control, Data Management Plan, Documentation, and Digital Objects. The Resources stage includes Data Source, SSH Files/Directories, and UNICORE Storages. The Automation stage includes UNICORE Jobs. The Results stage includes Archive and Publication. Plus signs indicate expandable sections.

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, ...

The screenshot shows the WorkflowHub interface for a workflow titled "Sorting and registration of Terahertz ELBE raw data" (Version 1). The page includes a search bar, navigation tabs for Overview, Files, and Related items, and a detailed description of the workflow. It also lists the creator (Thomas Gruber), license (Creative Commons Attribution 4.0), and version history.

The screenshot shows the HELIPORT documentation page. The header includes the HELIPORT logo, a search bar, and navigation links for About, Docs, and user profile. The main content area is titled "Documentation" and contains a table with columns for ID, Description, and System. Below the table is a section for "Add a Documentation".

ID	Description	System
57	Project documentation in Mediawiki	MediaWiki

The screenshot shows the HELIPORT Version Control page. It features a search bar and navigation links. The main content area is titled "Version Control" and contains a table with columns for ID and Name. Below the table is a section for "Add a Source Code Repository" with options for HZDR GitLab, Other, and New. It also includes a section for "Select where you want to create your new repository" with buttons for HZDR GitLab repository and GitHub repository.

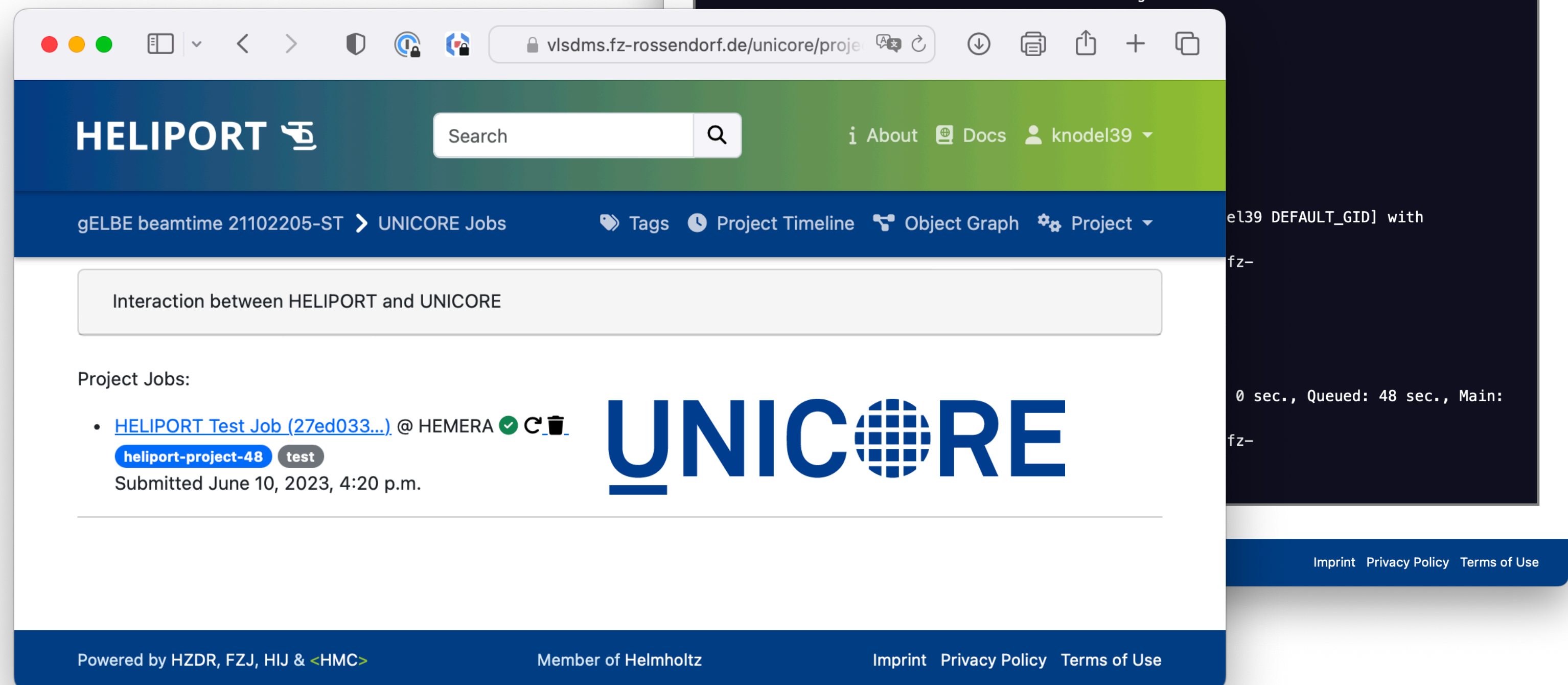
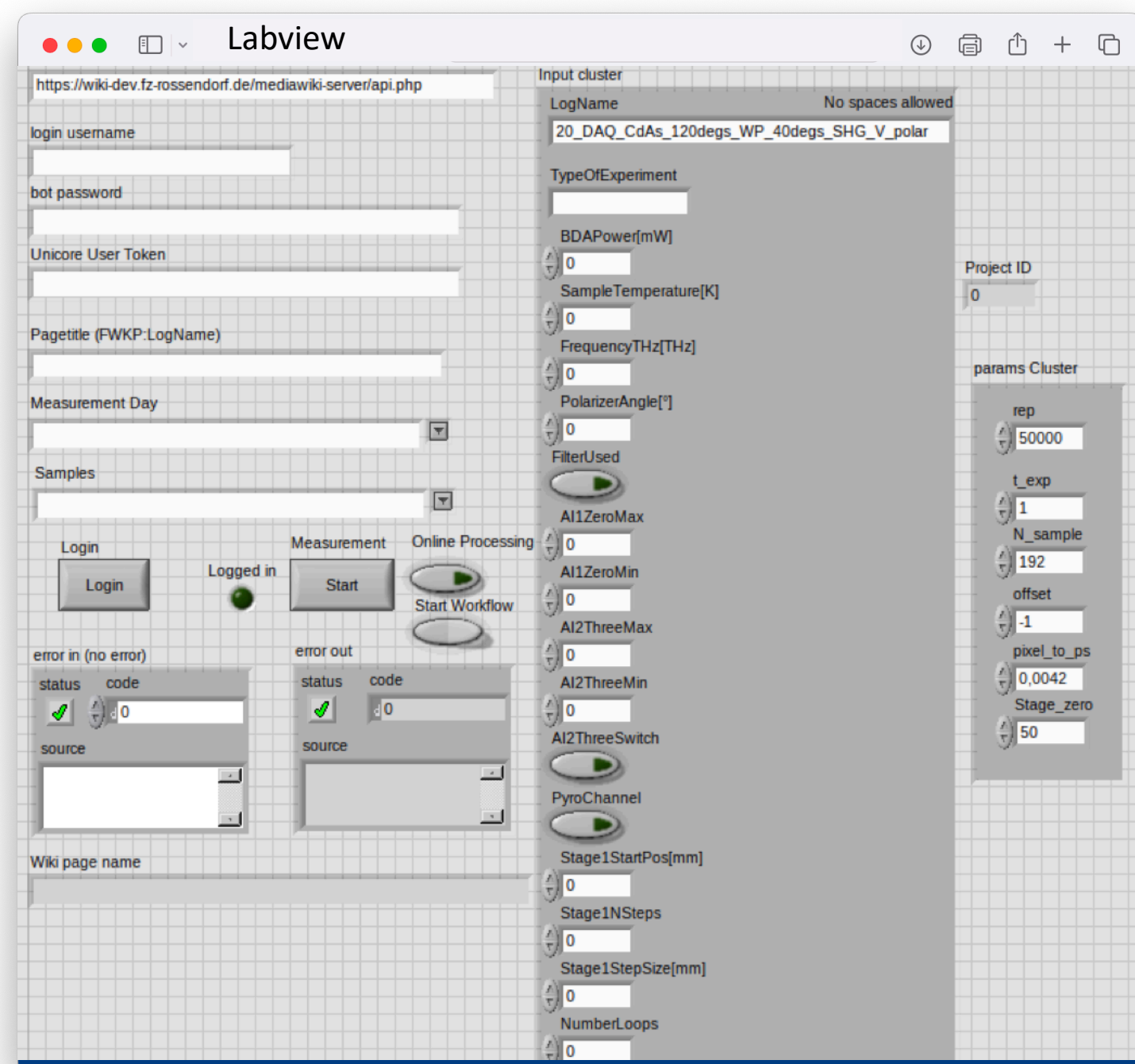
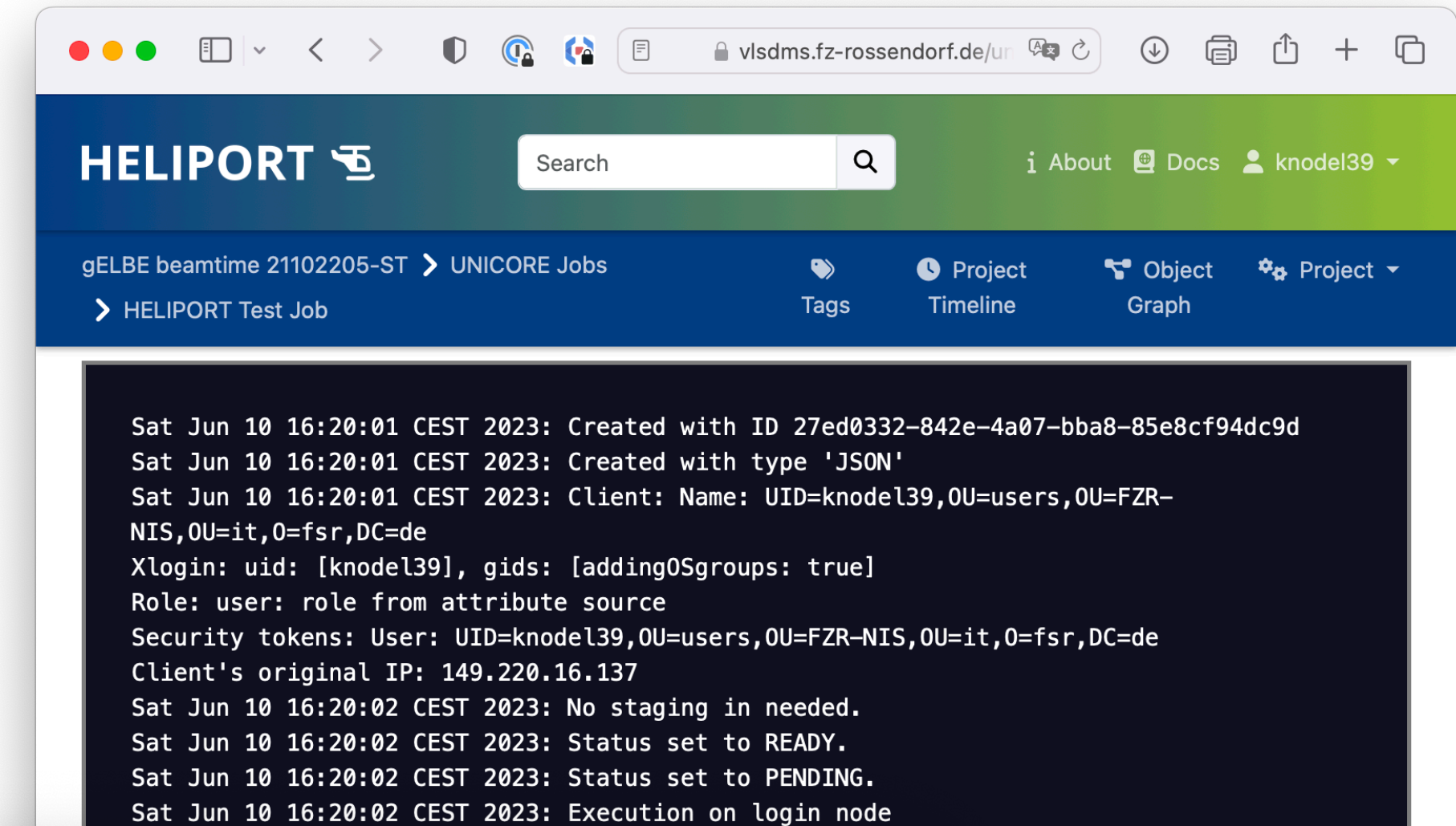
ID	Name
7	Telbe sorting script

The screenshot shows a Mediawiki page for a dataset titled "FWKP:22 DAQ CdAs 120degs WP 45degs SHG V polar 01". The page includes a navigation menu, a "Dataset" section with three plots, and a table of data files. The plots show the direct plot, power linescale plot, and power logscale plot of the dataset.

Data Files	File:FWKP:22 DAQ CdAs 120degs WP 45degs SHG V polar all loops.dat
Workflowhub URL	https://workflowhub.eu/workflows/459/ro_crate?version=1
Workflowhub Version	1
Repetition Rate	50000 Hz
time on single step measurement	1

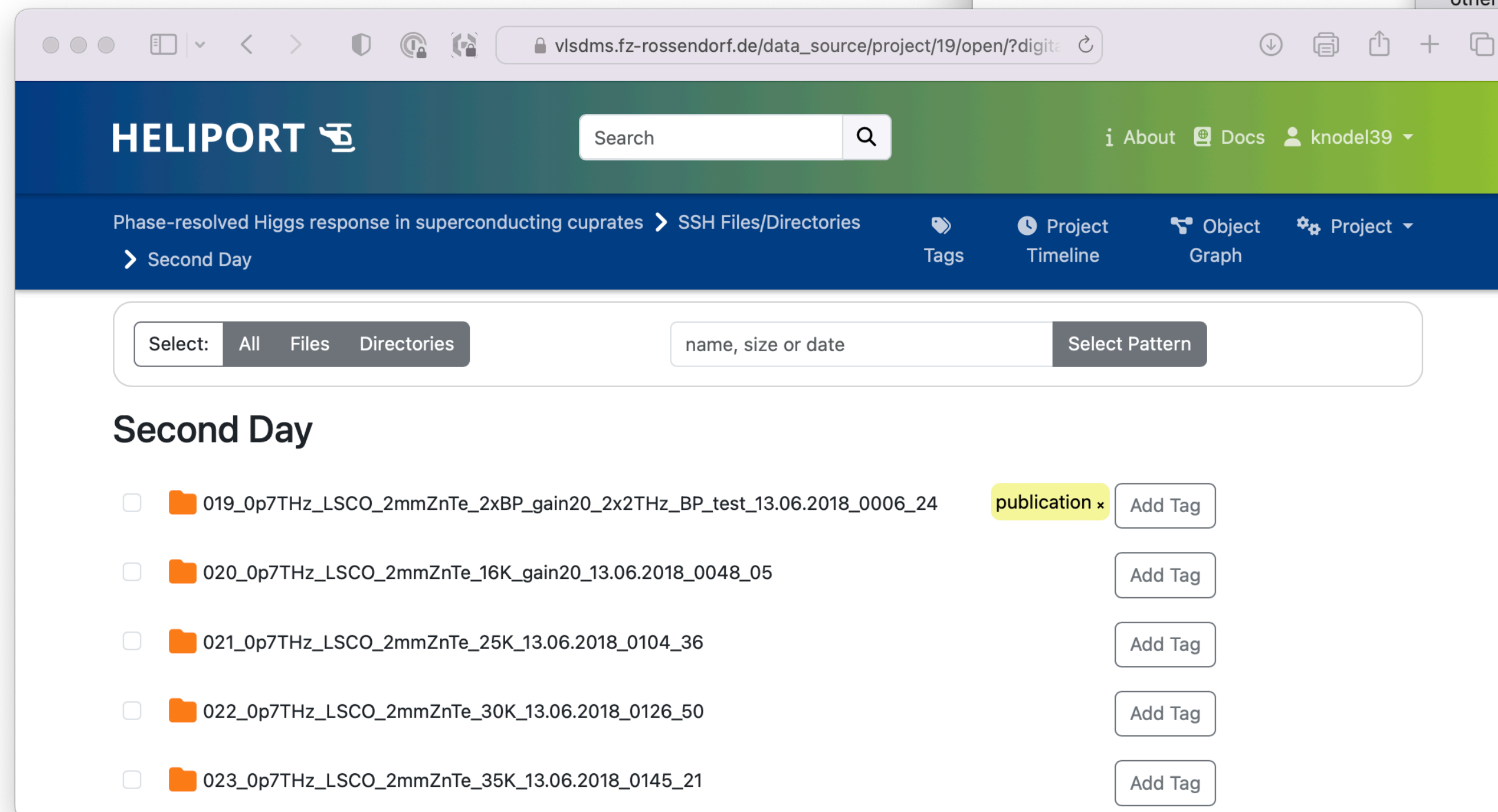
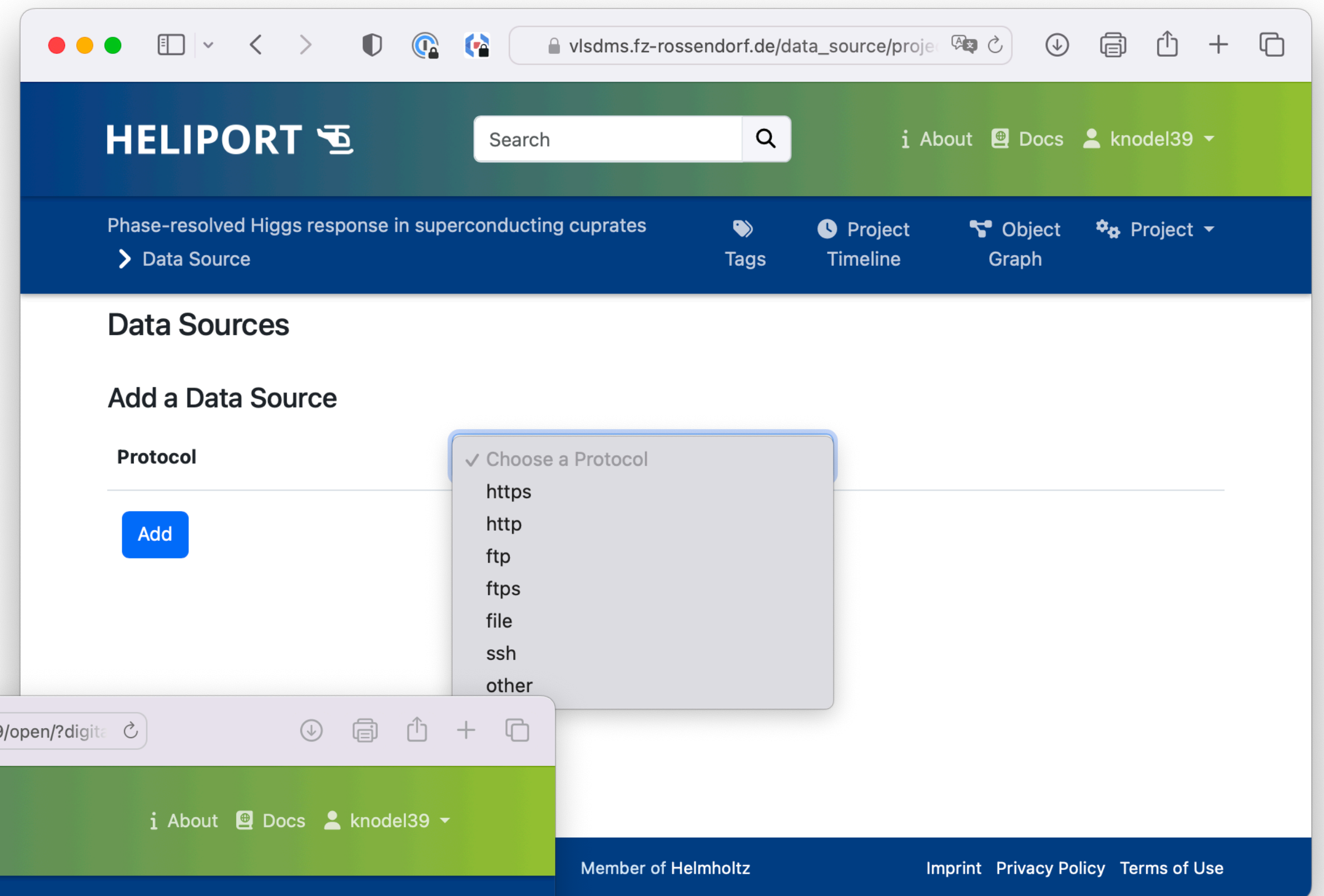
IV. Detector Control and Workflows

- 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 HELIPOINT as data source.
- Each member of a HELIPOINT 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

Project Properties

HZDR-ID	HZDR.FWCC.2021.114636
Digital Object ID	83
uuid	12215397-437a-468a-a95d-1a1d3f1d92ea
Landing Page	https://vlsdms.fz-rossendorf.de/object/83/?format=landing_page
Created	May 18, 2021, 5:03 p.m.
Department	FWCC
Title	Phase-resolved Higgs response in superconducting ci

Systems

- Version Control
- Data Management Plan
- Documentation
- Digital Objects

Resources

- Data Source
- SSH Files/Directories
- UNICORE Storages

Phase-resolved Higgs response in superconducting cuprates

December 16, 2021

2,980 views | 10,619 downloads

Publication date: December 16, 2021

DOI: DOI: 10.14278/rodare.1289

Keyword(s): Superconductors, terahertz, Higgs, Nonlinear dynamics, ultrafast

Related identifiers: Identical to: <https://www.hzdr.de/publications/Publ-30902>, Referenced by: <https://www.hzdr.de/publications/Publ-29647>, 10.1038/s41467-020-15613-1

Communities: Research field: Matter, RODARE

License (for files): Creative Commons Attribution 4.0 International

Versions

Version 2	Dec 16, 2021
10.14278/rodare.1289	

Second Day

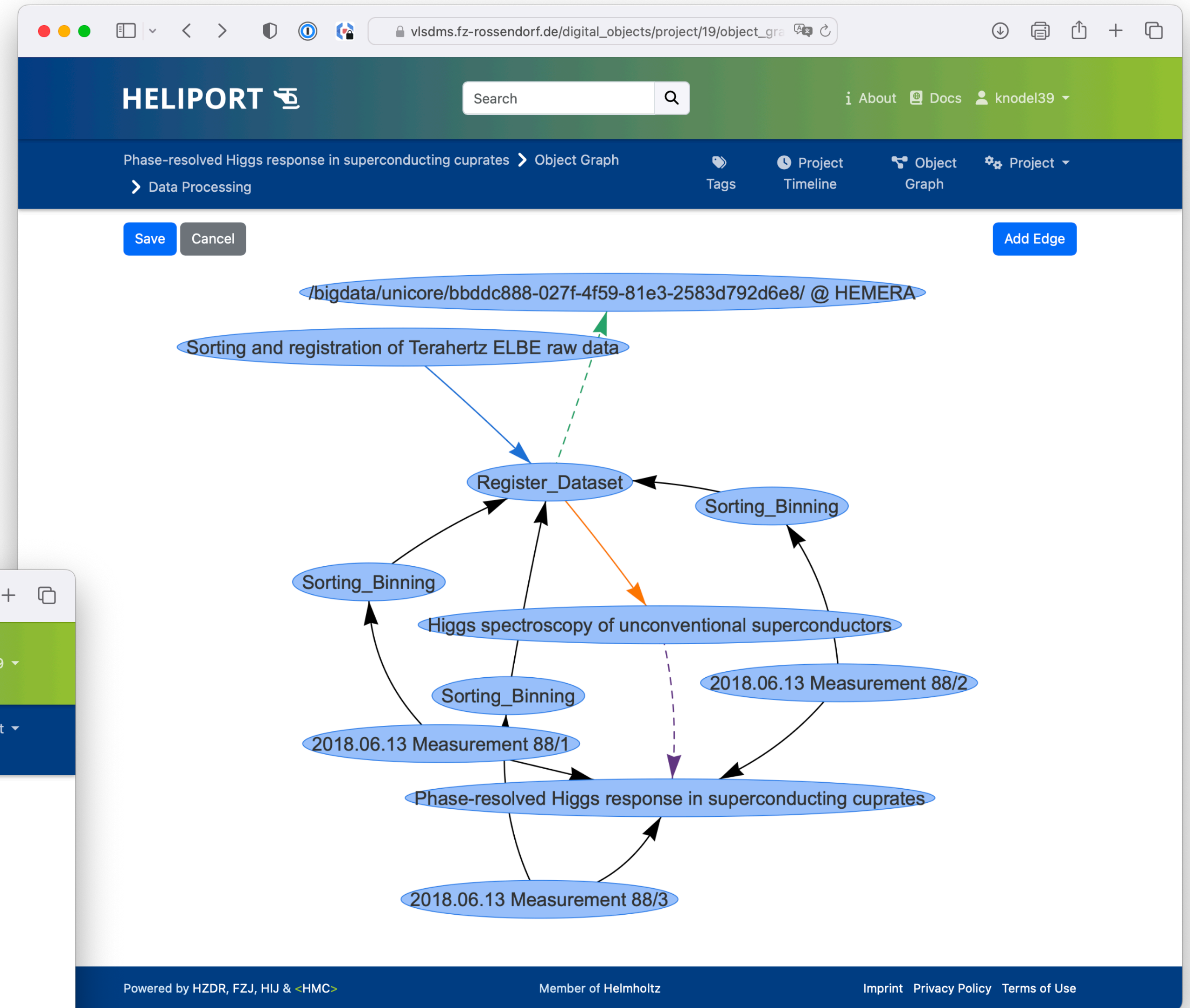
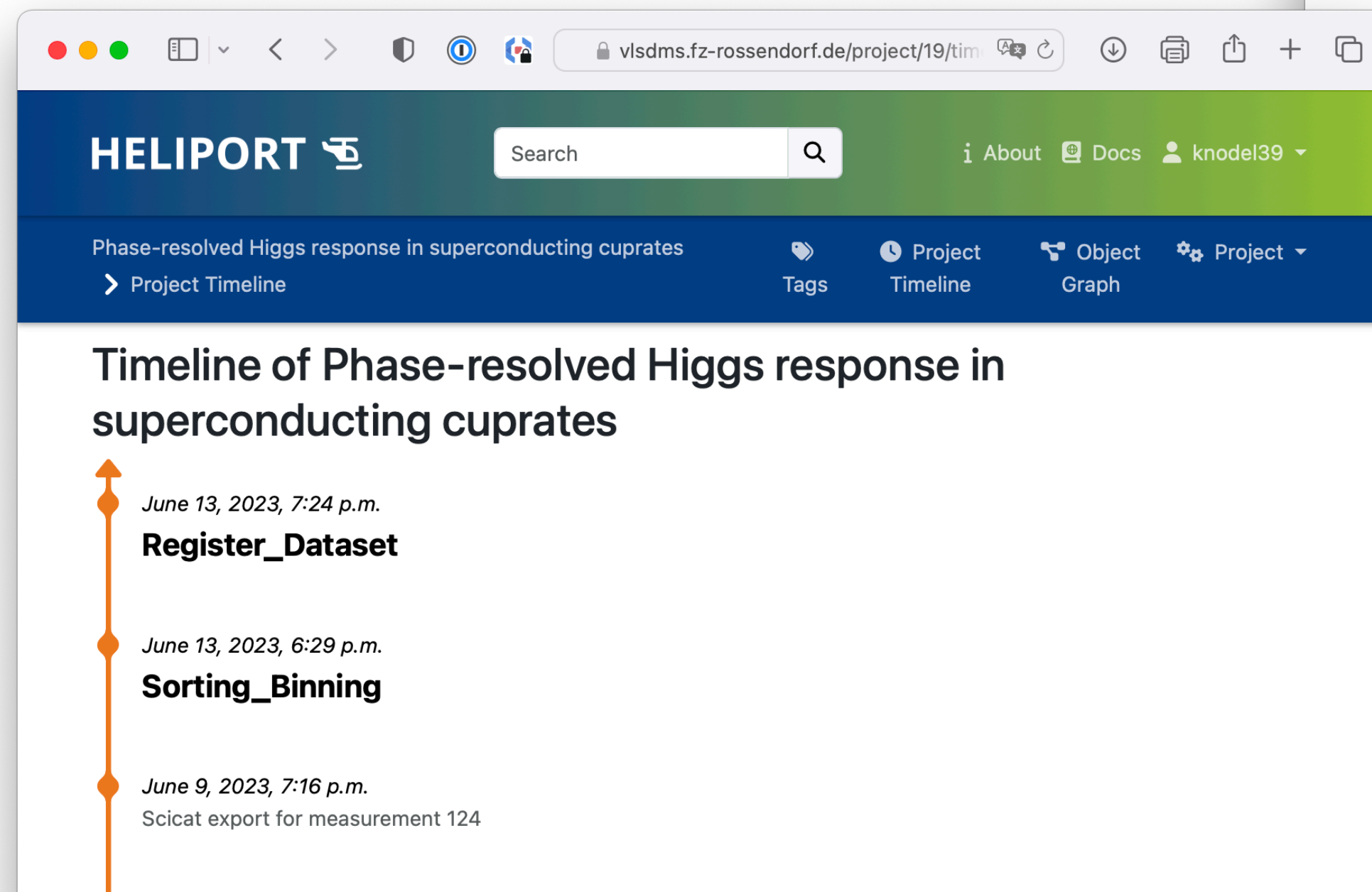
- 019_0p7THz_LSCO_2mmZnTe_2xBP_gain20_2x2THz_BP_test_13.06.2018_0006_24 **publication** Add Tag
- 020_0p7THz_LSCO_2mmZnTe_16K_gain20_13.06.2018_0048_05 Add Tag
- 021_0p7THz_LSCO_2mmZnTe_25K_13.06.2018_0104_36 Add Tag
- 022_0p7THz_LSCO_2mmZnTe_30K_13.06.2018_0126_50 Add Tag
- 023_0p7THz_LSCO_2mmZnTe_35K_13.06.2018_0145_21 Add Tag

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.



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, ...)

The screenshot shows the HELIPOINT web interface. The browser address bar displays 'visdms.fz-rossendorf.de/project/74/timeline/'. The page header includes the HELIPOINT logo, a search bar, and navigation links for 'Info', 'Project', and the user 'knodel39'. The main content area is titled 'Timeline of gELBE beamtime 21202619-ST' and includes a checkbox for 'Show timeline in project graph'. The timeline consists of several entries:

- May 3, 2022, 2:28 p.m.**
TRCprocess
- April 29, 2022, 12:47 p.m.**
Alex Keshavarzi's github repo (use branch McrDev)
- April 29, 2022, 12:43 p.m.**
/bigdata/GATE21202619ST/Data
Filesystem on /bigdata containing the beamtime data.
- April 28, 2022, 1:19 p.m.**
nothing
- April 20, 2022, 11:17 a.m.**
DSPEC_LaBr
This folder contains the DSPEC-runs taken with the LaBr detector, exported to the HZDR cloud. Password is "ELBE2022"
- April 14, 2022, 12:04 p.m.**
Run logbook
- April 6, 2022, 5:07 p.m.**
Cloud folder (Password: ELBE2022)

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.

The screenshot displays the Heliport REST API documentation interface. The browser address bar shows the URL: `vlsdms.fz-rossendorf.de/redoc/#operation/createDigitalObject`. The interface is divided into several sections:

- Left Sidebar:** A navigation menu with a search bar and a list of API endpoints. The 'digital-objects' section is expanded, showing endpoints like `listDigitalObjects`, `createDigitalObject` (highlighted), `retrieveDigitalObject`, `updateDigitalObject`, `partialUpdateDigitalObject`, and `destroyDigitalObject`.
- Main Content Area:**
 - Endpoint:** `createDigitalObject` (POST) under the 'Digital Objects' group.
 - Request Body Schema:** A table defining the JSON payload structure:

Field	Type	Constraints
<code>project</code>	integer	required
<code>handle</code>	string	≤ 100 characters, Nullable
<code>relation</code>	string	required
<code>category</code>	string	required
<code>description</code>	string	required
 - Responses:** A section showing a response code of `201`.
- Right Panel:** A dark-themed sidebar with two sections:
 - Request samples:** Shows a JSON payload: `{ "project": 0, "handle": "string", "relation": "string", "category": "string", "description": "string" }`.
 - Response samples:** Shows a JSON response for status `201`: `{ "digital_object_id": 0, "project": 0, "handle": "string", "relation": "string", "category": "string", "description": "string" }`.