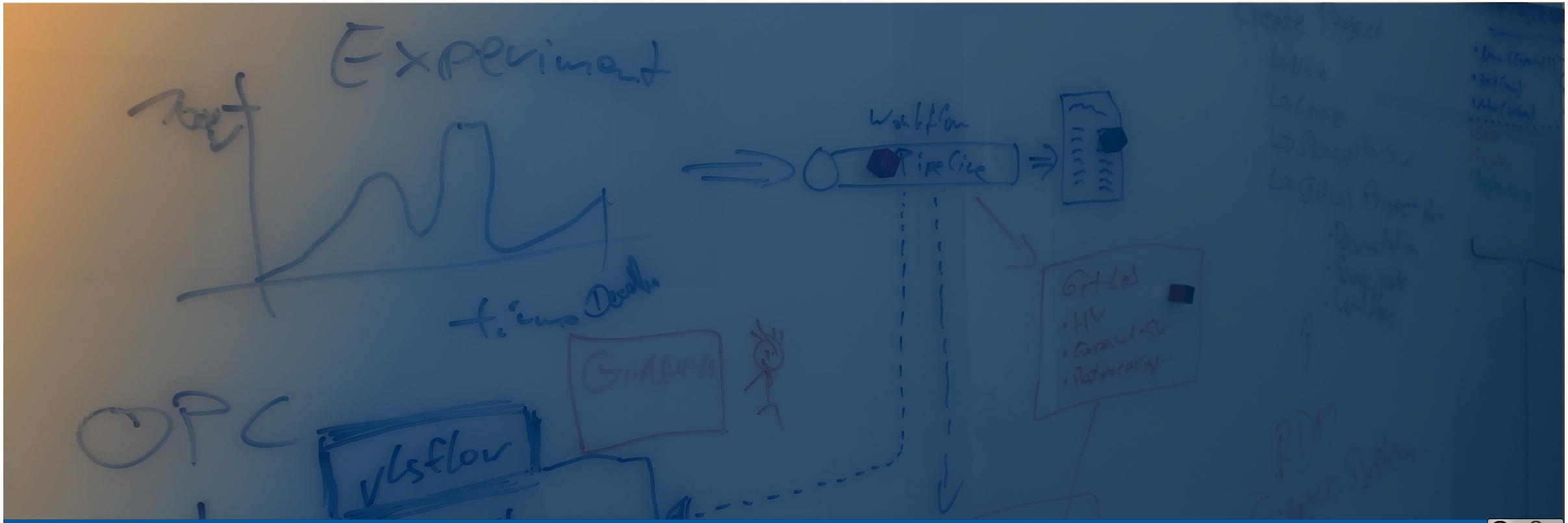
HELIPORT

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



A FAIRly Integrated Scientific Project Lifecycle

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

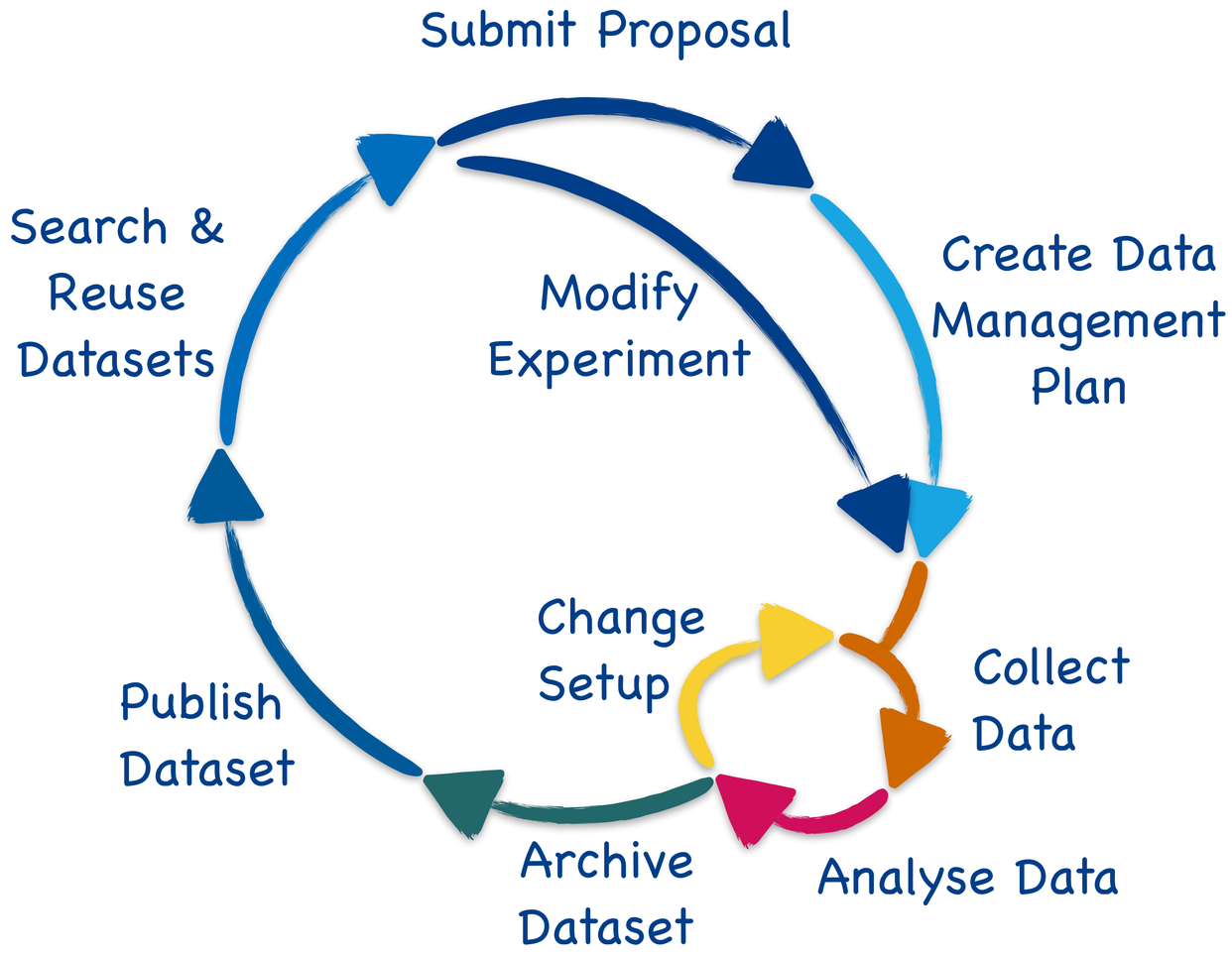
DRESDEN concept

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.

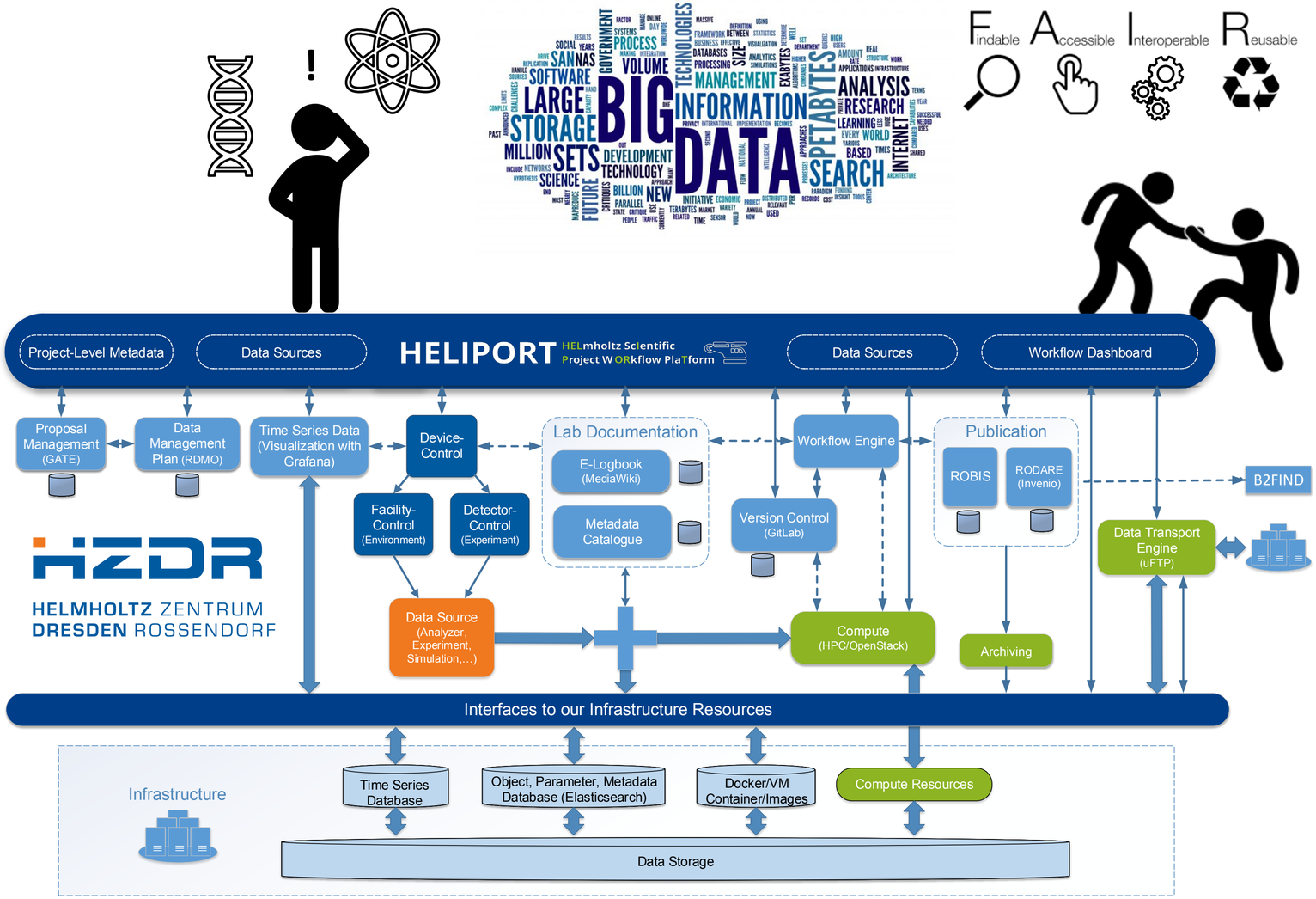
— The documentation of all these linked resources is essential to create a comprehensible and FAIR data lifecycle.



HELIPORT JÜLICH FORSCHUNGSZENTRUM HI JENA Helmholtz Institute Jena HZDR

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 (and our services) is inevitable.
- The concept of FAIR research becomes an important topic for our scientists.



HELIPORT JÜLICH FORSCHUNGSZENTRUM HI JENA Helmholtz Institute Jena HZDR

The Requirements and Conditions

- Our guidance system was originally intended to provide only the **proposal's metadata**, from internal and external scientists, to allow the assignment of resources.
- Over the time we decided to use the guidance system to answer the most important questions of our scientists:

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

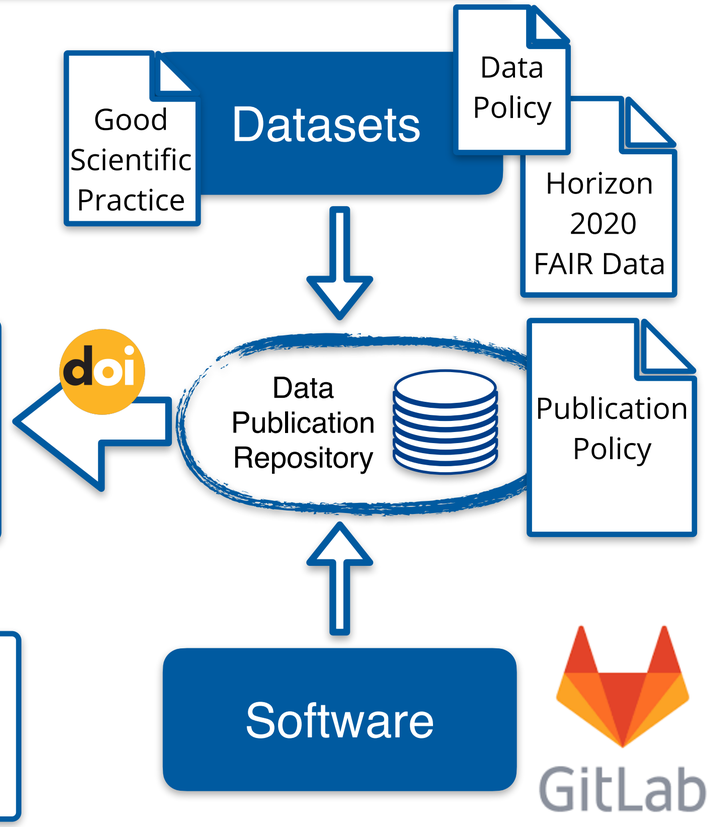
And how we can support them?!

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 all associated tools?

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



HELIPORT JÜLICH FORSCHUNGSZENTRUM HI JENA Helmholtz Institute Jena HZDR

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:

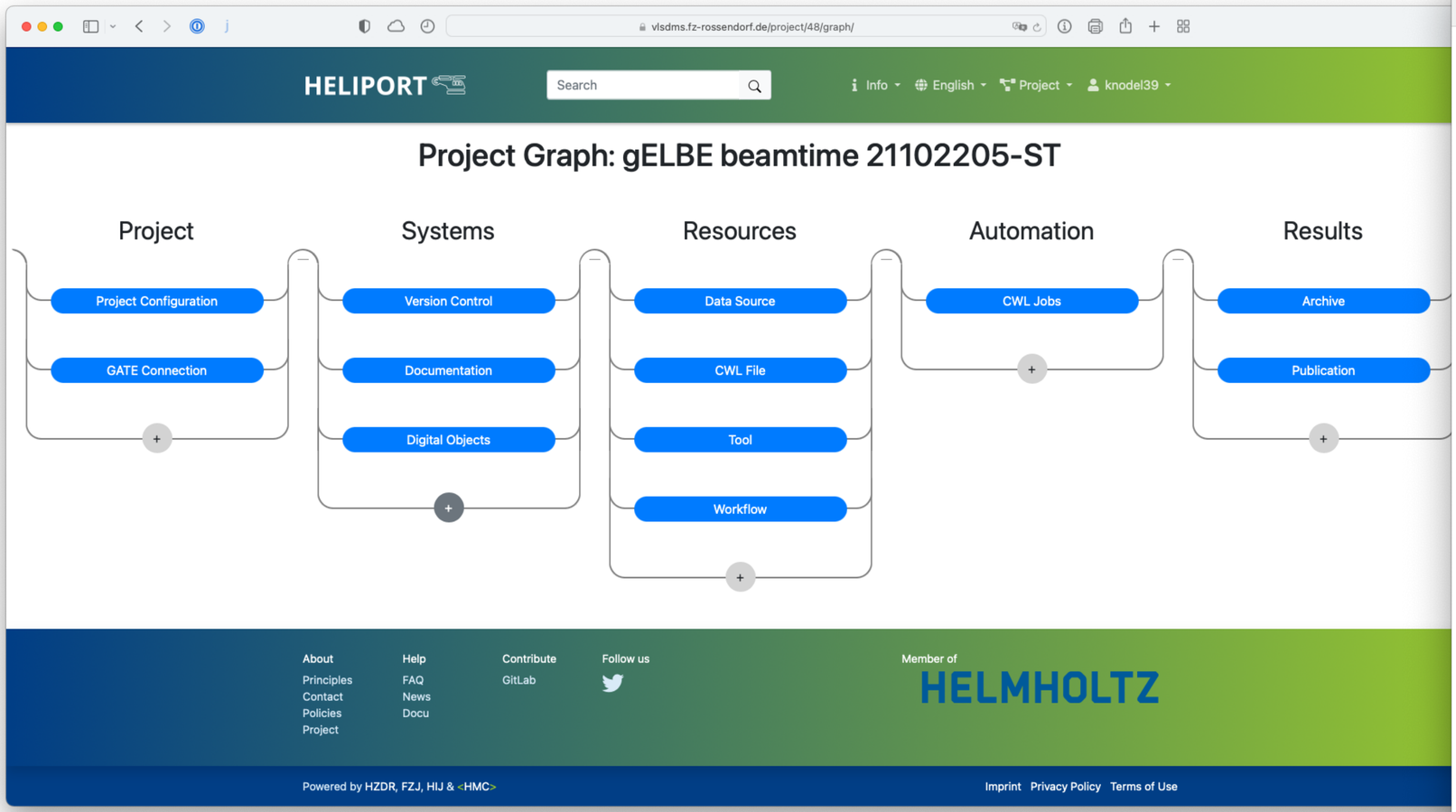
HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

HI JENA Helmholtz Institute Jena

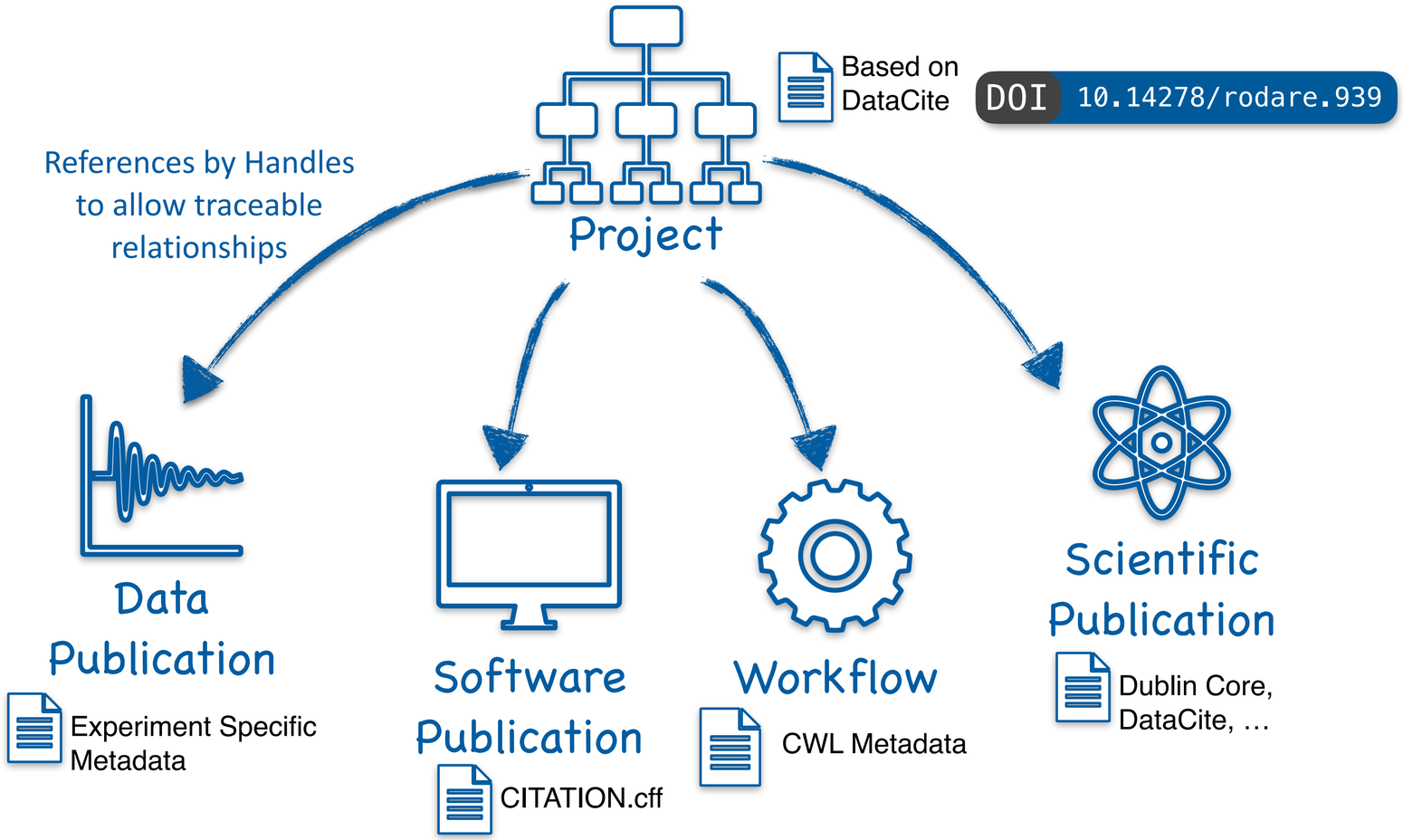
JÜLICH FORSCHUNGSZENTRUM

Founded by:

<HMC> HELMHOLTZ METADATA COLLABORATION



Heliport Metadata Ecosystem



Our Concept

- In all stages of an experiment Heliport combines information about involved services with PIDs.
- Metadata (stored *near* the PID) is used to transfer information between different systems and a documentation of the project-level workflow is possible.
- The project-level metadata is distributed over all linked third-party systems.
- The metadata from all the involved systems can be exported into an overall schema.

HELIPORT JÜLICH FORSCHUNGSZENTRUM HI JENA Helmholtz Institute Jena HZDR

HELIPORT

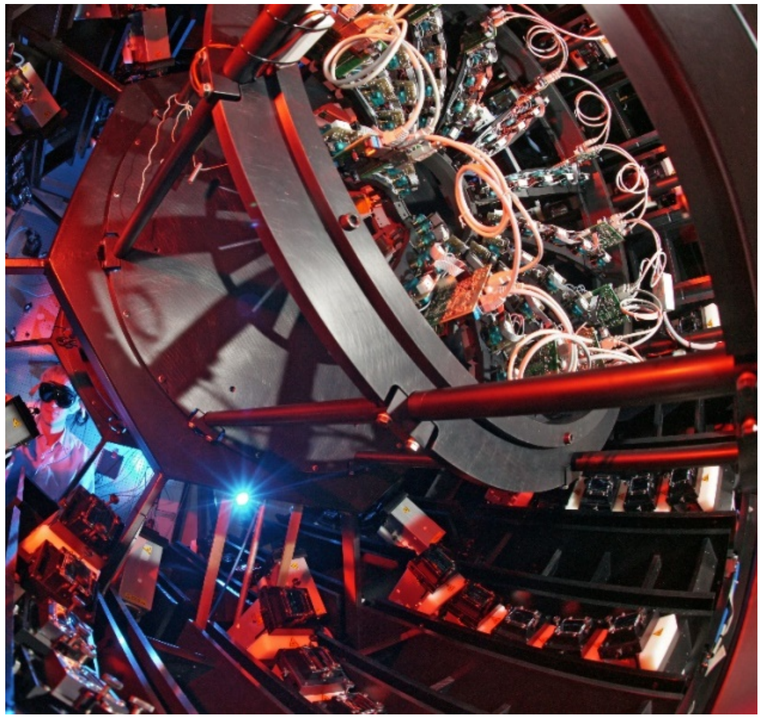
JÜLICH FORSCHUNGSZENTRUM
HI JENA Helmholtz Institute Jena
HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

Example I: POLARIS

DRESDEN concept

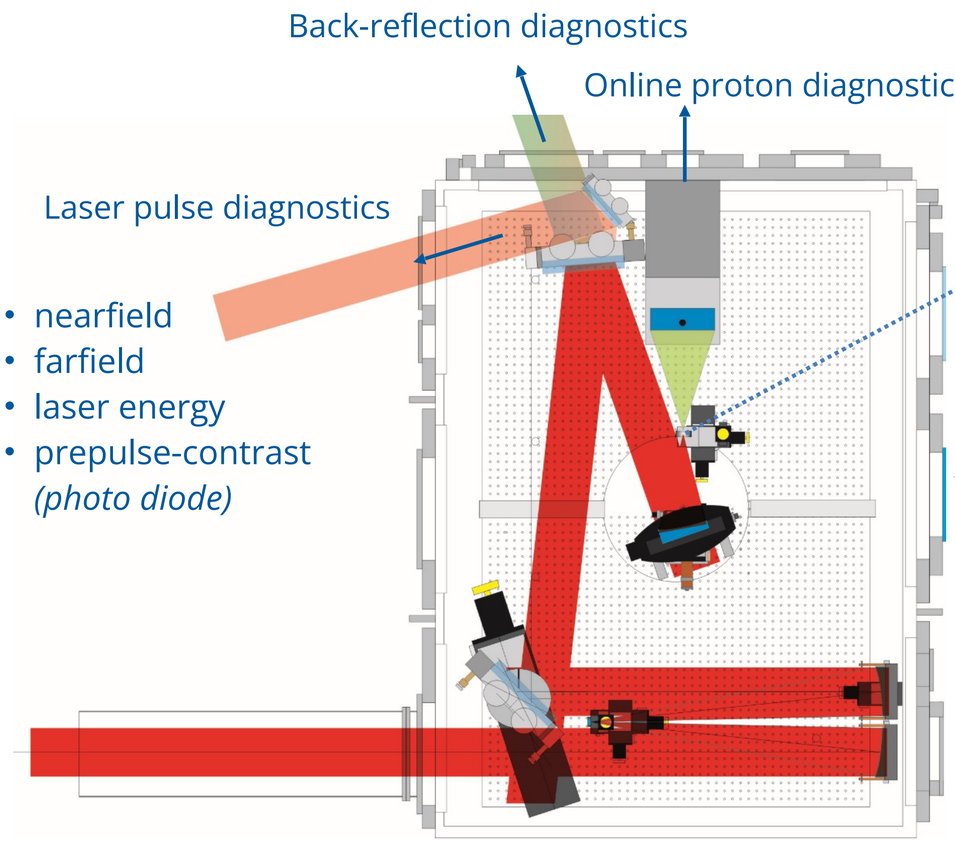
Example: The POLARIS Experiment @ HI Jena

- The **P**etawatt **O**ptical **L**aser **A**mplifier for **R**adiation **I**ntensive Experiment**S (POLARIS)** is the only fully diode-pumped, double-CPA laser system worldwide.
- Used for Laser Particle Acceleration Experiments.
- Research project of the Faculty of Physics and Astronomy at the University of Jena and the Helmholtz Institute Jena (HI-Jena).

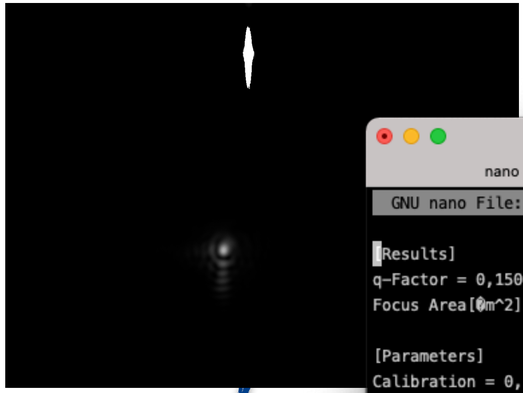


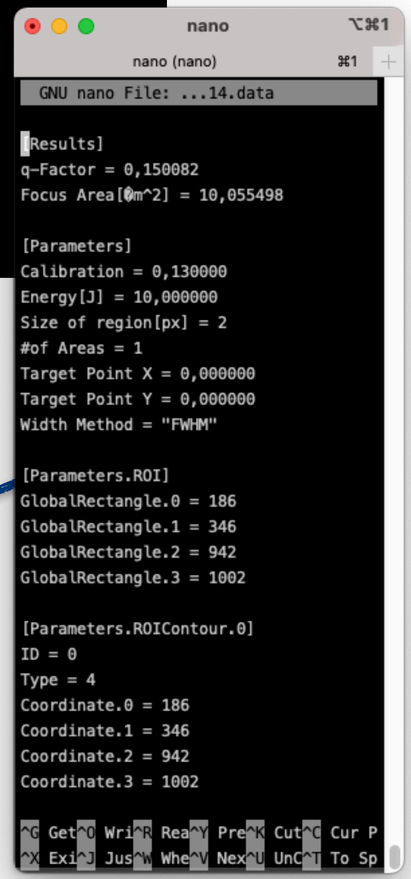
Malte C. Kaluza, Contrast Dependence of Laser-Driven Proton Acceleration, 18th Advanced Accelerators Concepts Workshop, Breckenridge, US, (2018)

POLARIS Experimental Setup

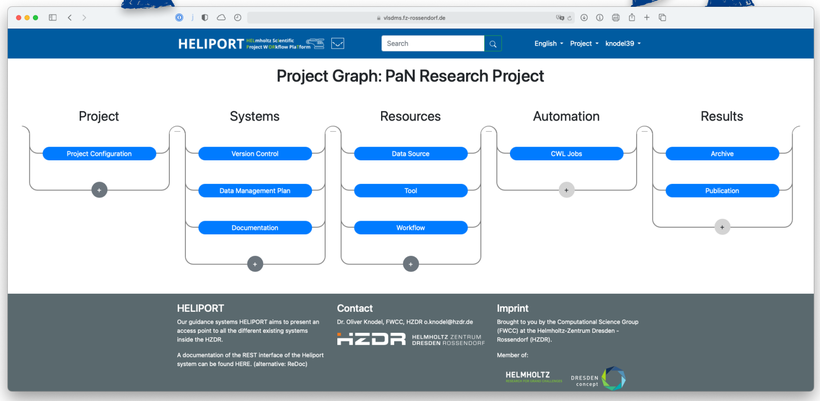


- nearfield
- farfield
- laser energy
- prepulse-contrast *(photo diode)*





In the HELIPORT project, our goal is to bring all together: images, settings, target metadata and everything else.

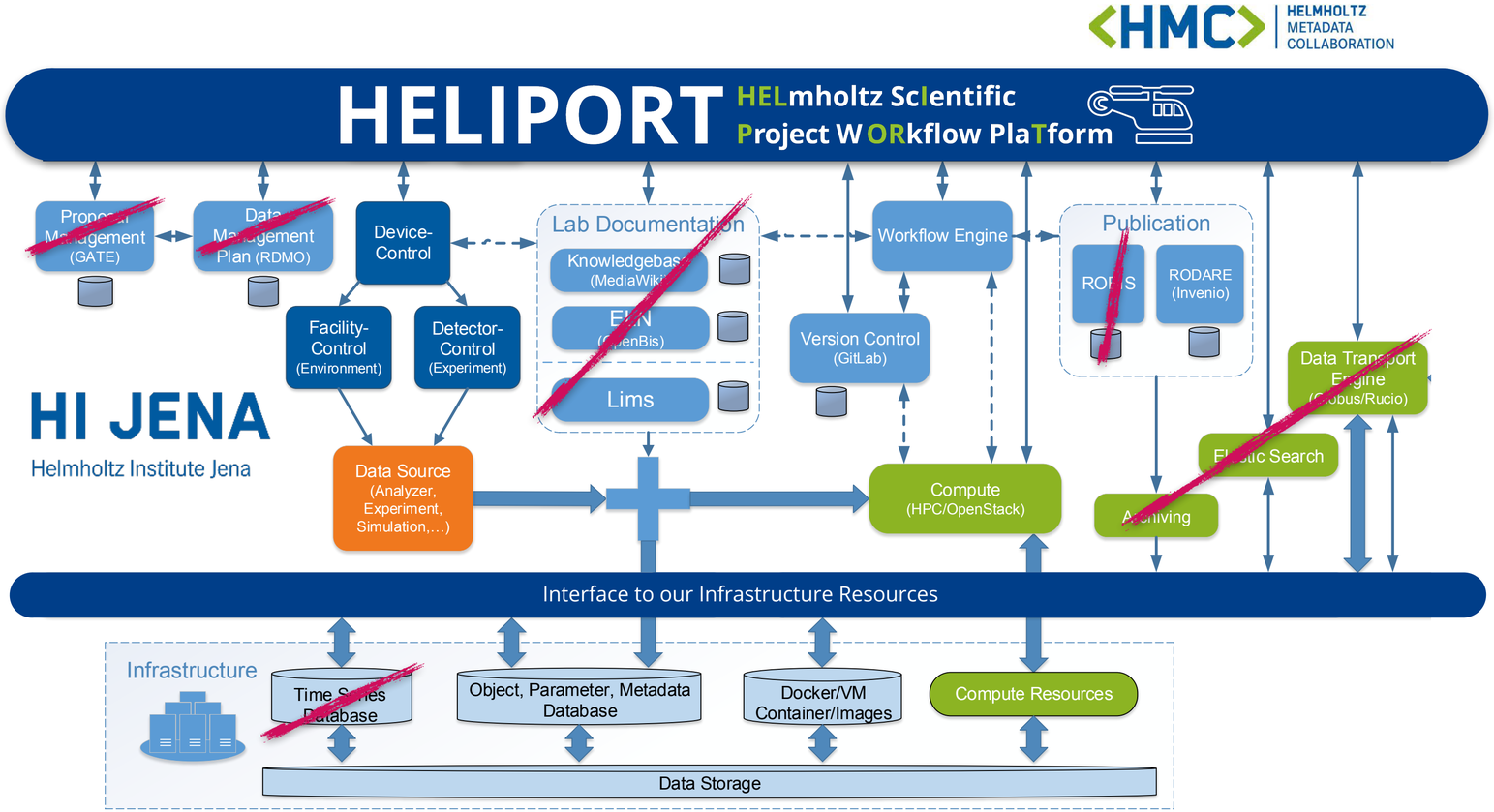


Malte C. Kaluza, Contrast Dependence of Laser-Driven Proton Acceleration, 18th Advanced Accelerators Concepts Workshop, Breckenridge, US, (2018)

HELIPORT JÜLICH FORSCHUNGSZENTRUM HI JENA Helmholtz Institute Jena HZDR

HELIPORT and the Infrastructure at Helmholtz Institute Jena

- HELIPORT was initially developed to provide access to the computing Infrastructure at the HZDR.
- The infrastructure at Jena is different, but the important systems are available:
  - HPC Infrastructure and Data Management*,
  - Data Acquisition, and
  - Git Repositories
- And others are in development:
  - Lab Documentation,
  - Workflow engine, and
  - Data Repository.



*Sponsored by the Thüringer Aufbaubank, Projekt **2019 FGI 0013 HDA**

HELIPORT

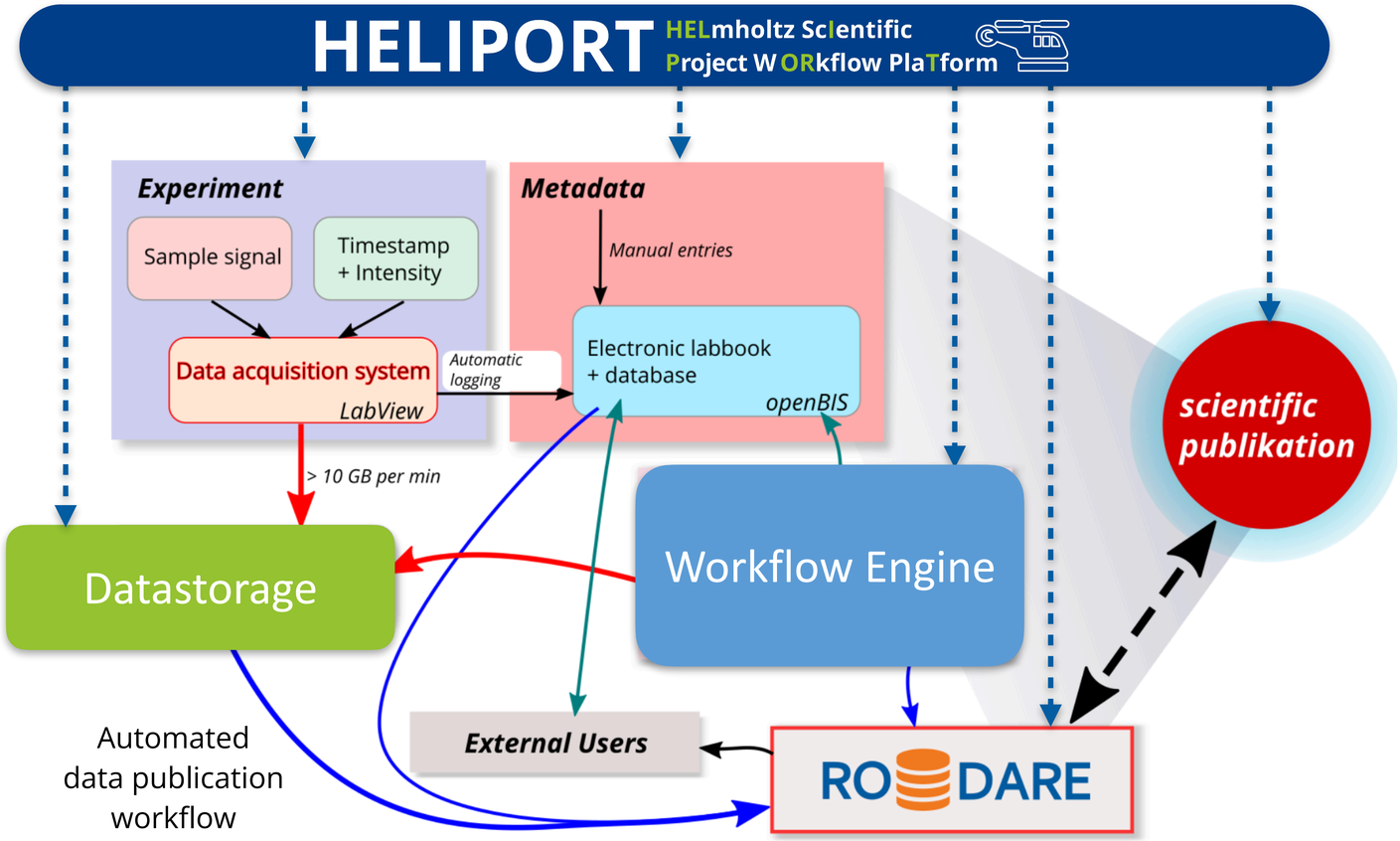
JÜLICH FORSCHUNGSZENTRUM
HI JENA Helmholtz Institute Jena
HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

Example II: TELBE

DRESDEN concept

TELBE Data Flow

- Terahertz facility at the ELBE center for High-Power Radiation Sources.
- In the future HELIPORT should guide external scientists through the complete experiment.



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

HELIPORT

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

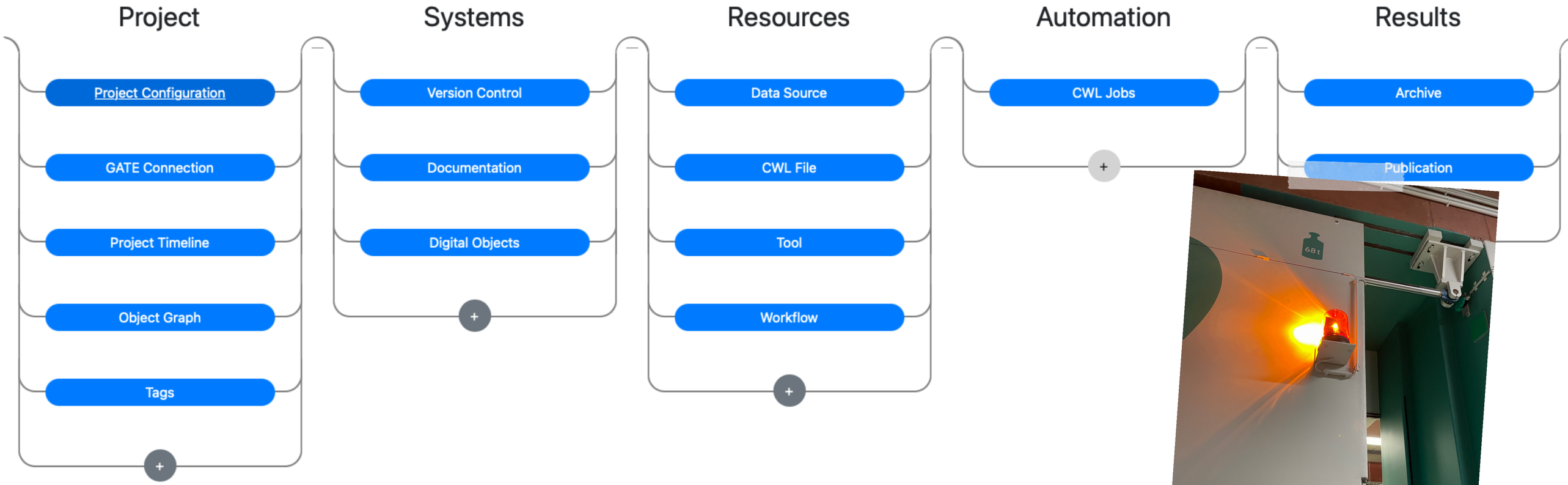
Appearance and Components

…explained with Example III

DRESDEN concept

Use-Case Project III: gELBE Beamtime 21102205-ST

- A simple detector test for the Mu2e experiment at Fermilab at our ELBE facility.
- All related resources are summarized in HELIPORT:



HELIPORT JÜLICH FORSCHUNGSZENTRUM HI JENA Helmholtz Institute Jena HZDR

The Initial Proposal Metadata

Proposal Metadata

— Title, abstract, status, ID,

— Proposer, experimentalist, local contact, experimental team, and

— Experiment schedule.

Additional Metadata within HELIPORT

— Extended user information (name, mail, OrcID),

— HZDR ID (Heliport Handle),

— Additional members with contributions,

— And: documentation, datasets, workflows, systems (infrastructure), software repositories, publications, …

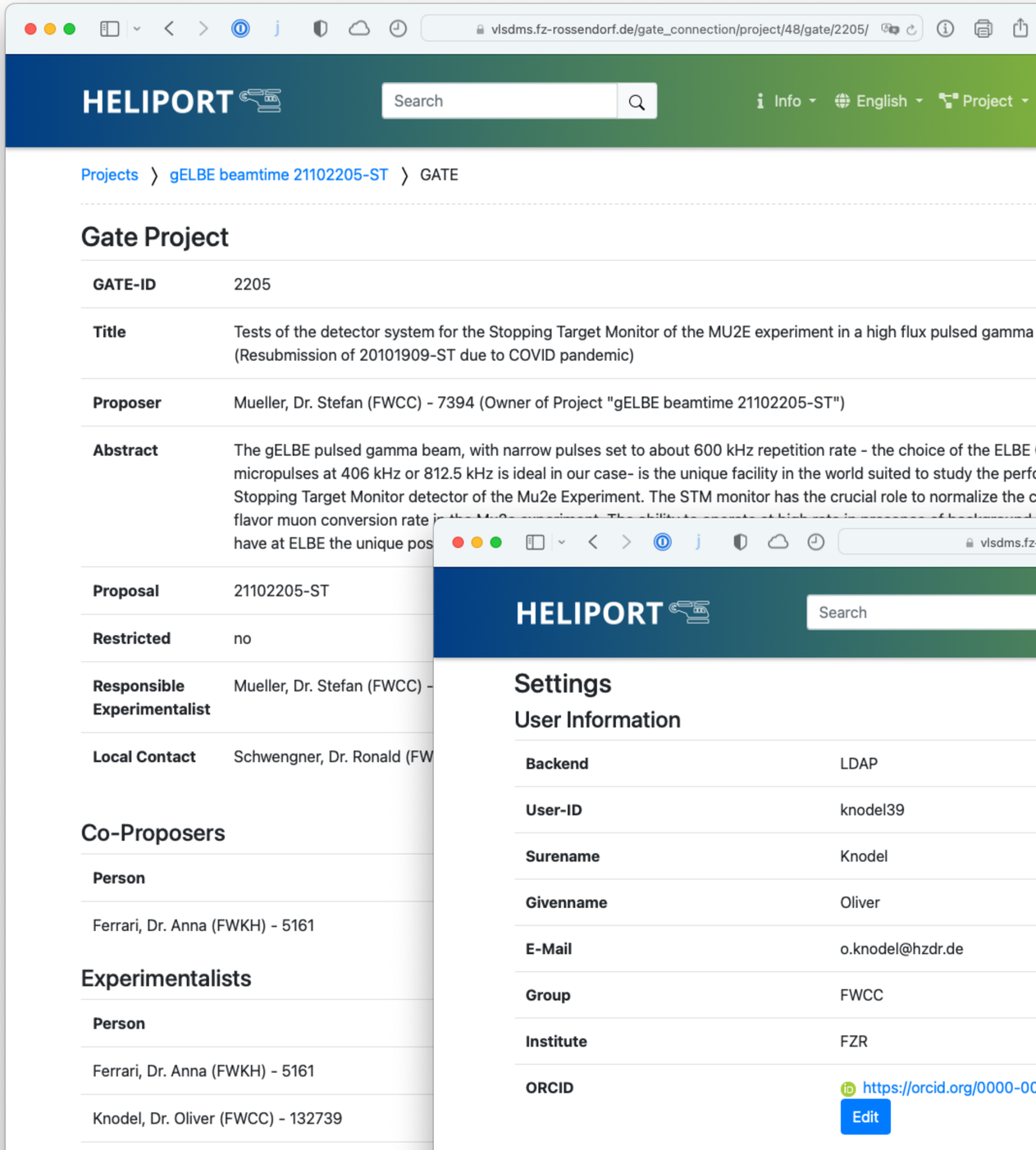
Next Metadata Fields

— Instrument or beam-line,

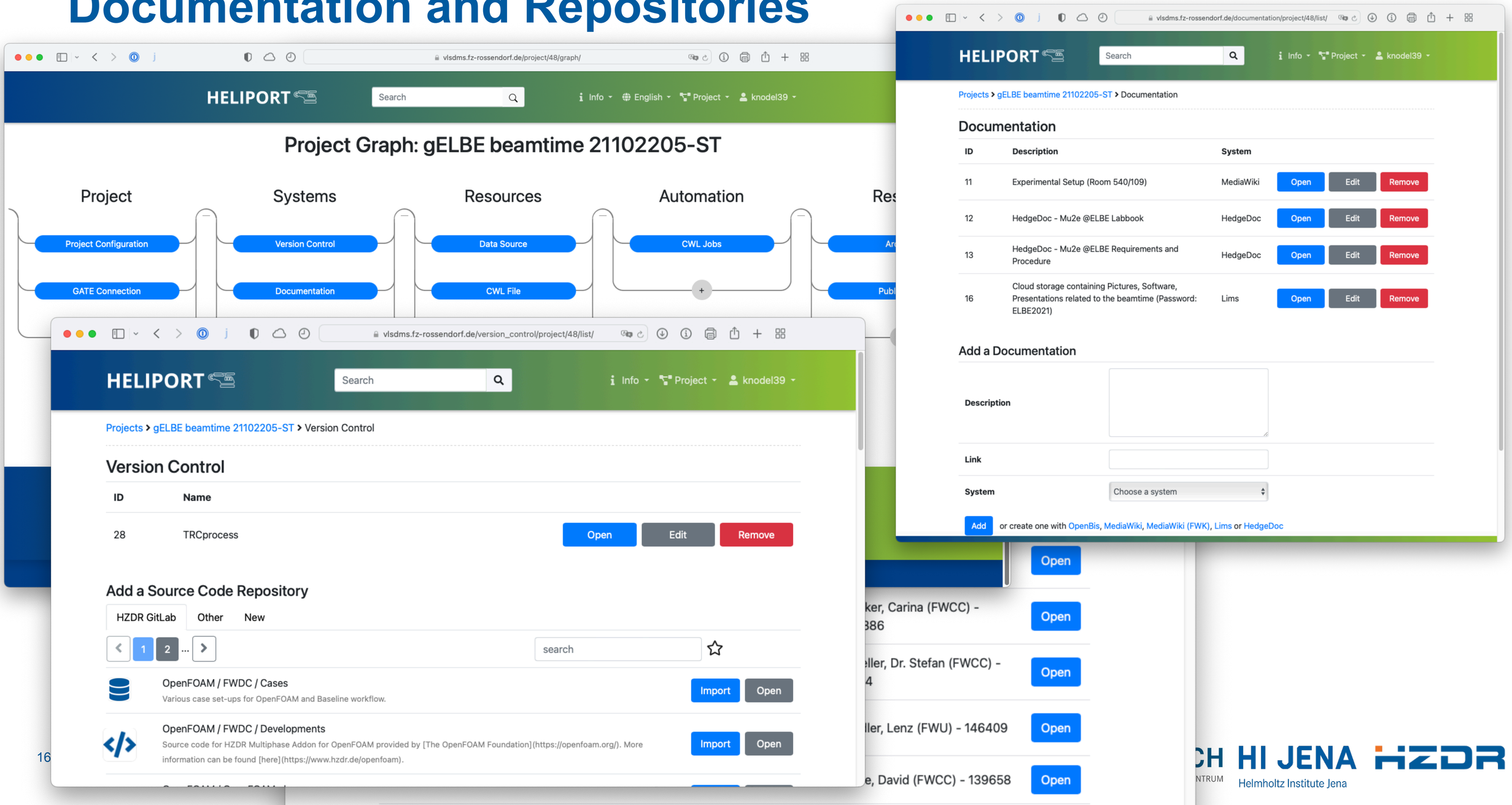
— Funding,

— Scientific method,

— ….



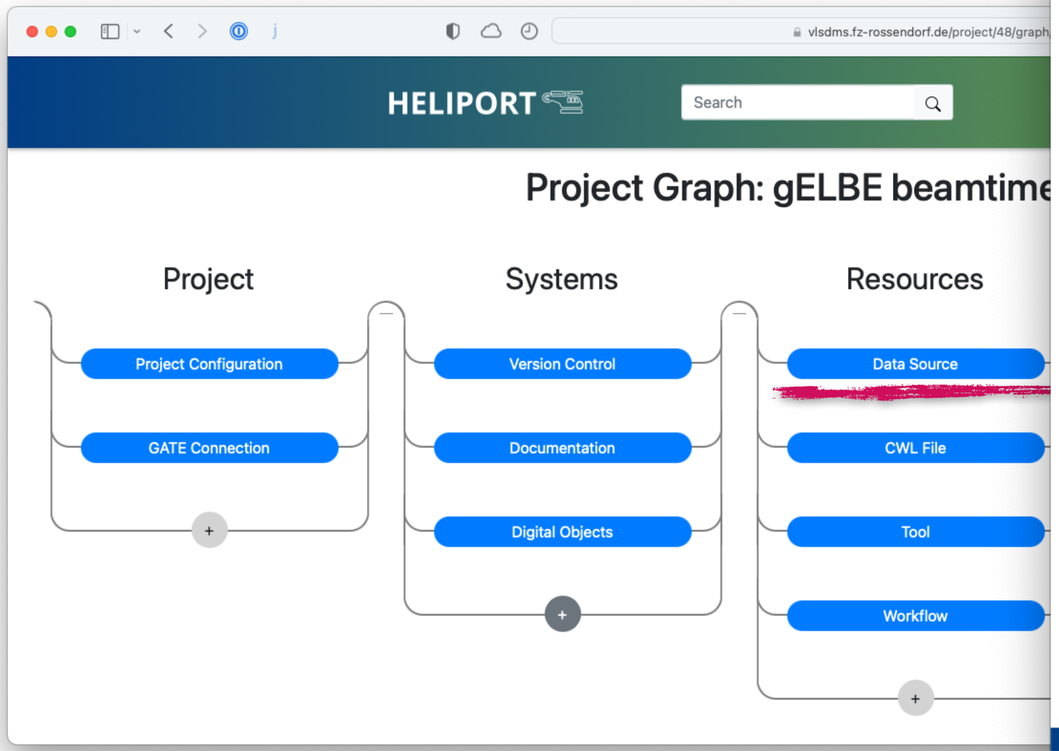
Documentation and Repositories

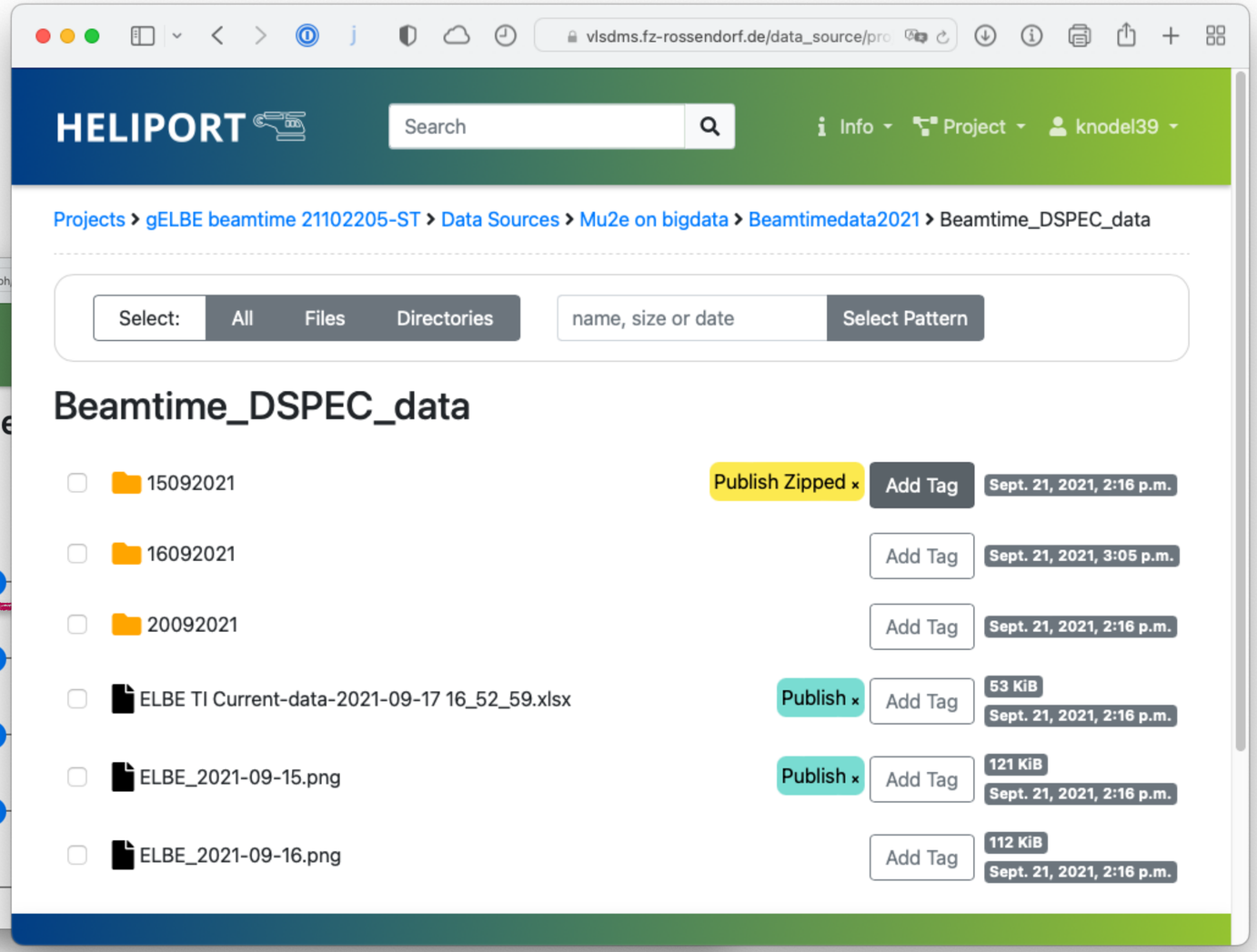


Helmholtz Institute Jena

Data Sources can be Registered…

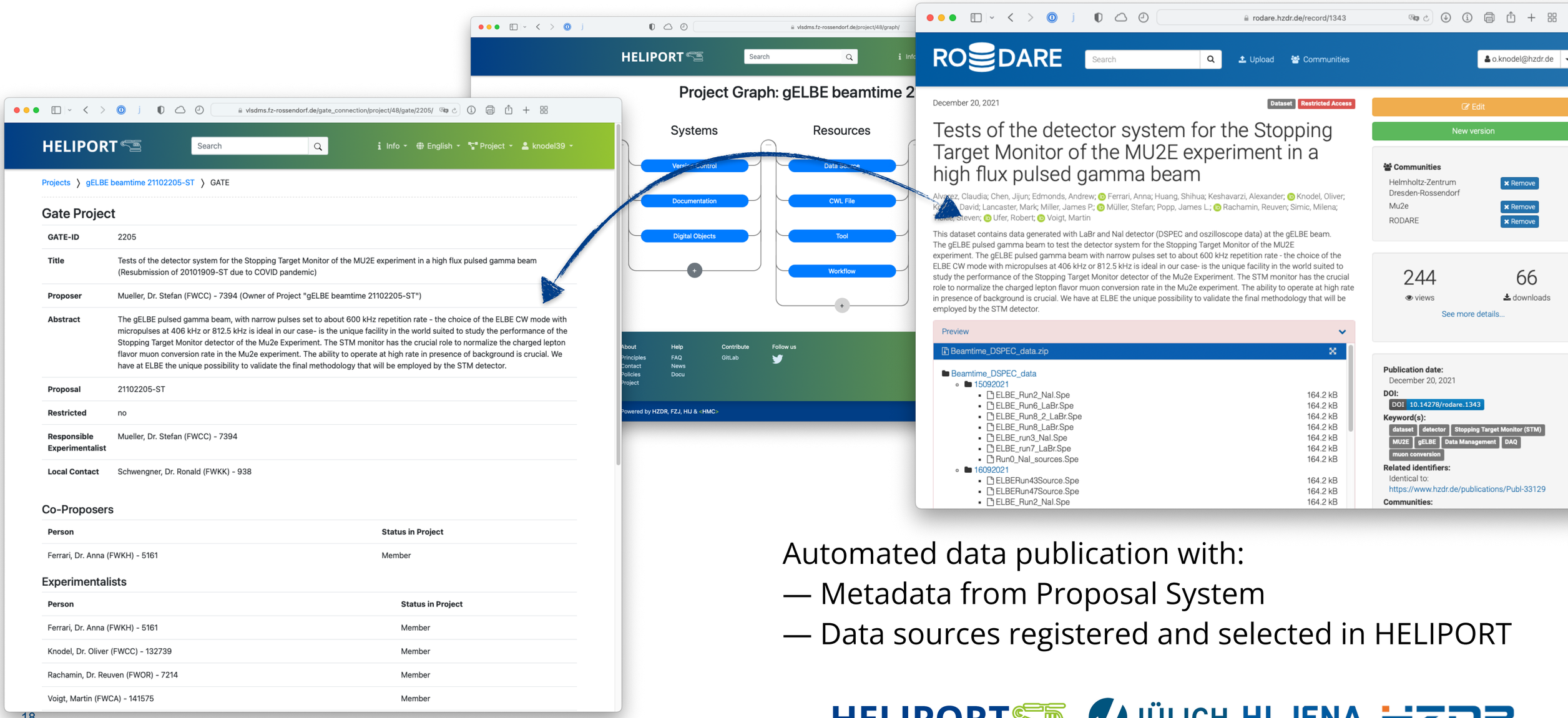
…and files (selected for publication) can be transferred directly to the data publication.





HELIPORT JÜLICH FORSCHUNGSZENTRUM HI JENA Helmholtz Institute Jena HZDR

Integration in an Overall Data Publication

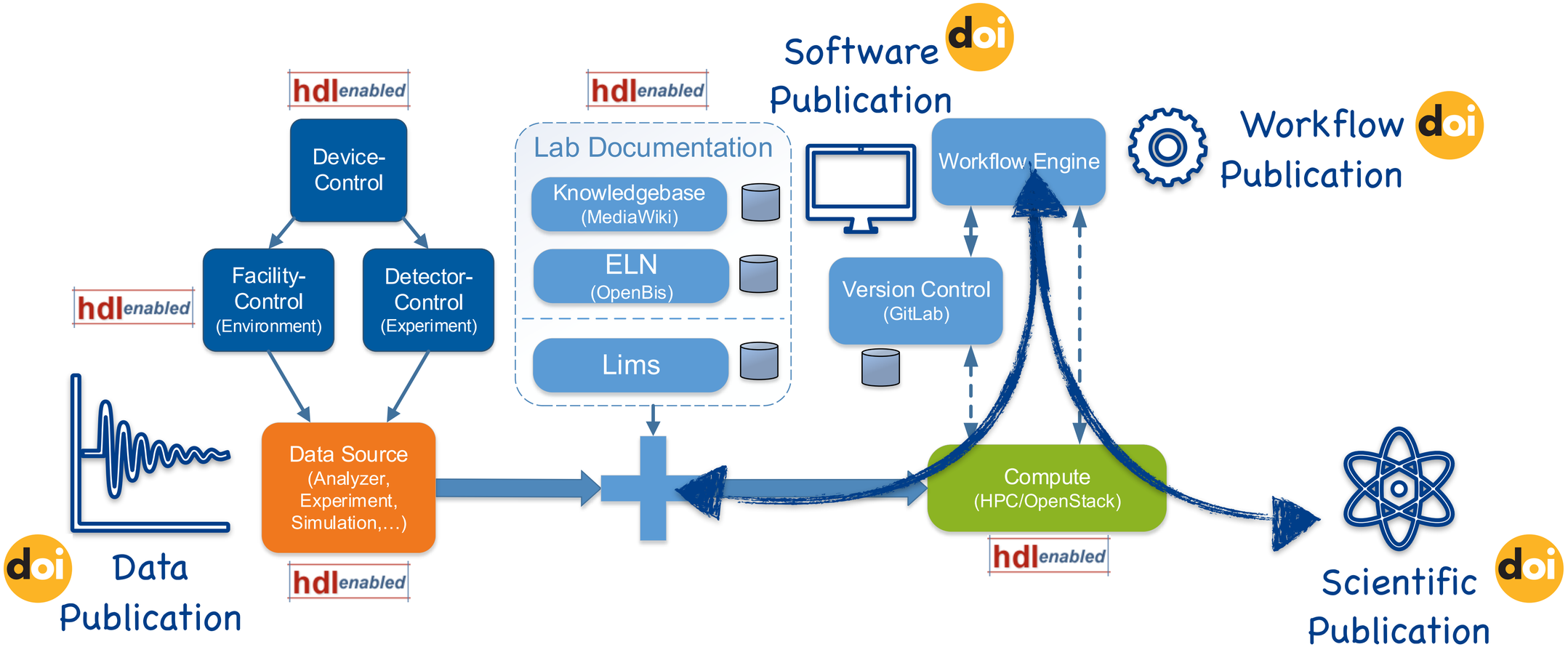


Automated data publication with:

— Metadata from Proposal System

— Data sources registered and selected in HELIPORT

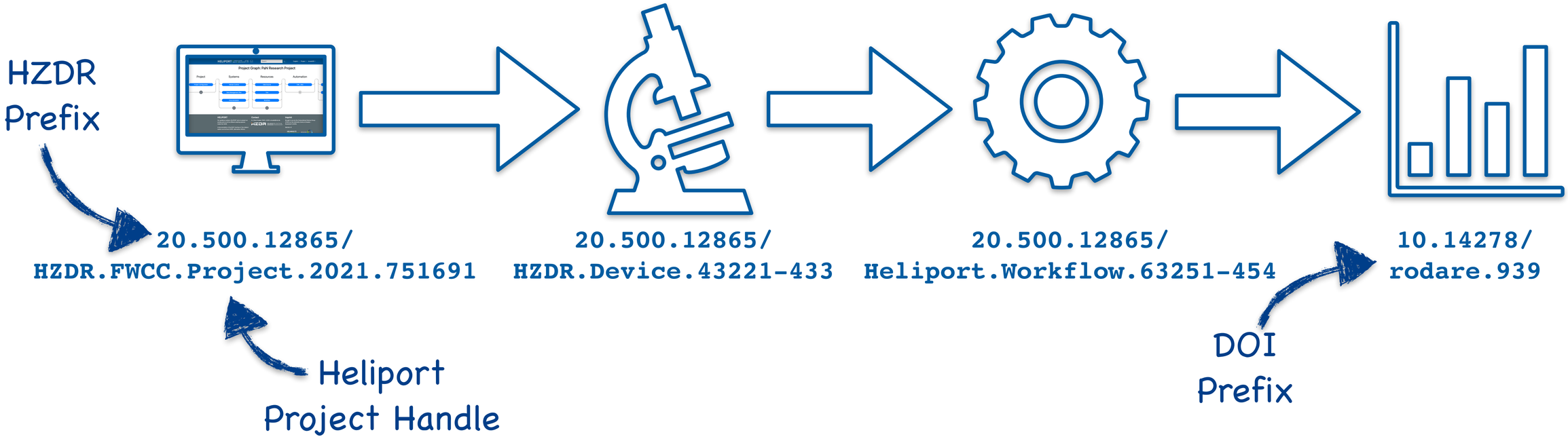
Different Types of (Data) Publications and Data Provenance



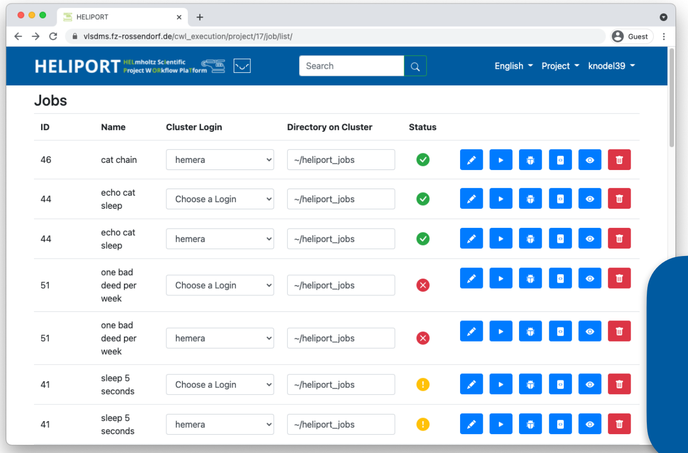
HELIPORT JÜLICH FORSCHUNGSZENTRUM HI JENA Helmholtz Institute Jena HZDR

Persistant Identifiers (Handles) and Digital Objects in Heliport

Heliport can be linked with local Handle servers (e.g. handle.hzdr.de) to generate uniform global PIDs for our digital objects and various systems and services.



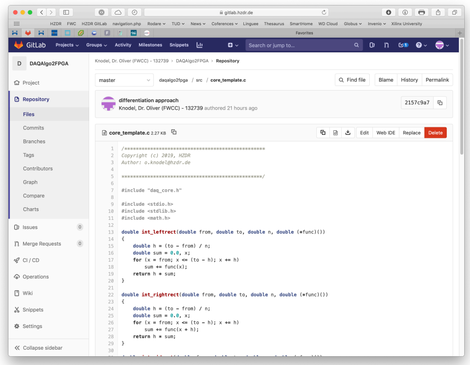
Scientific Software Development and Reproducible Workflows



Workflow Engine

Version Control

Compute (HPC, OpenStack)



GitLab

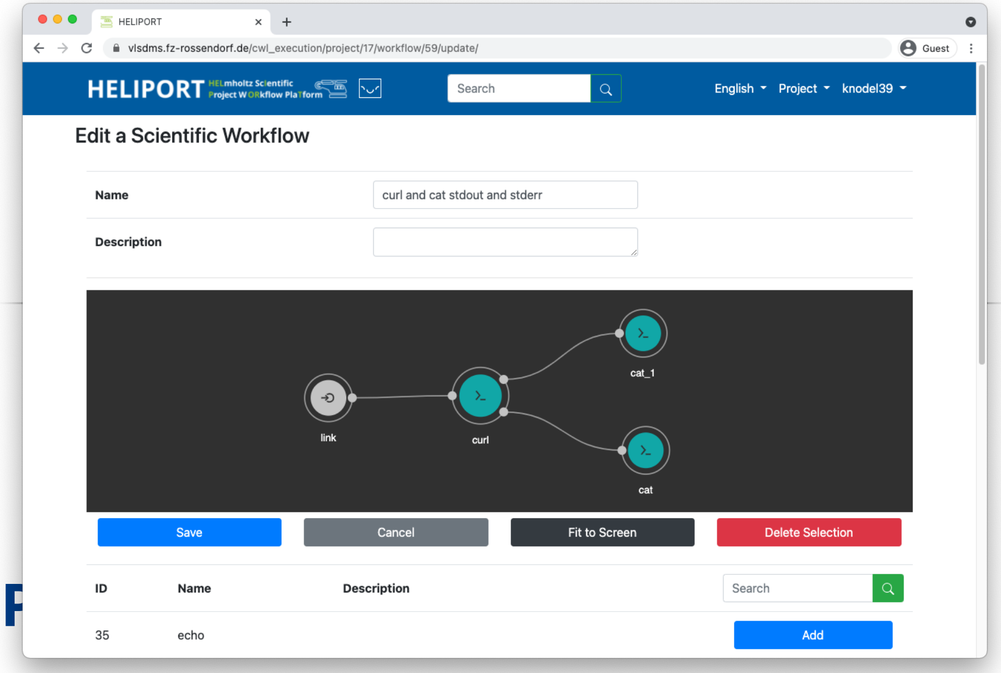
UNICORE

slurm workload manager

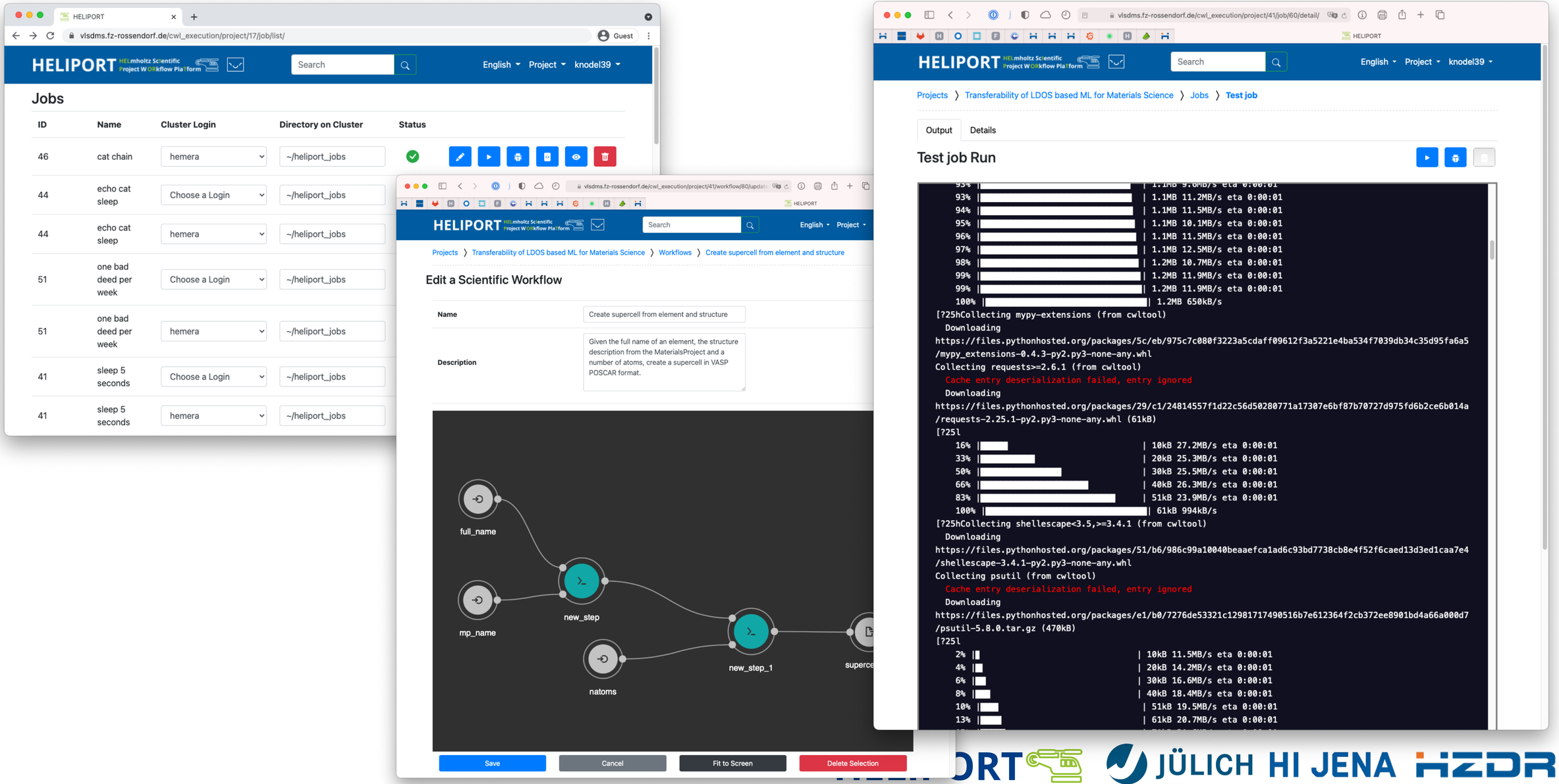
JÜLICH FORSCHUNGSZENTRUM

- Capsuling every step in a workflow adapts the FAIR principles.
- Analysis and Pre-/Postprocessing steps needs to be:
  - Documented and
  - Reproducible

COMMON WORKFLOW LANGUAGE



Scientific Software Development and Reproducible Workflows II

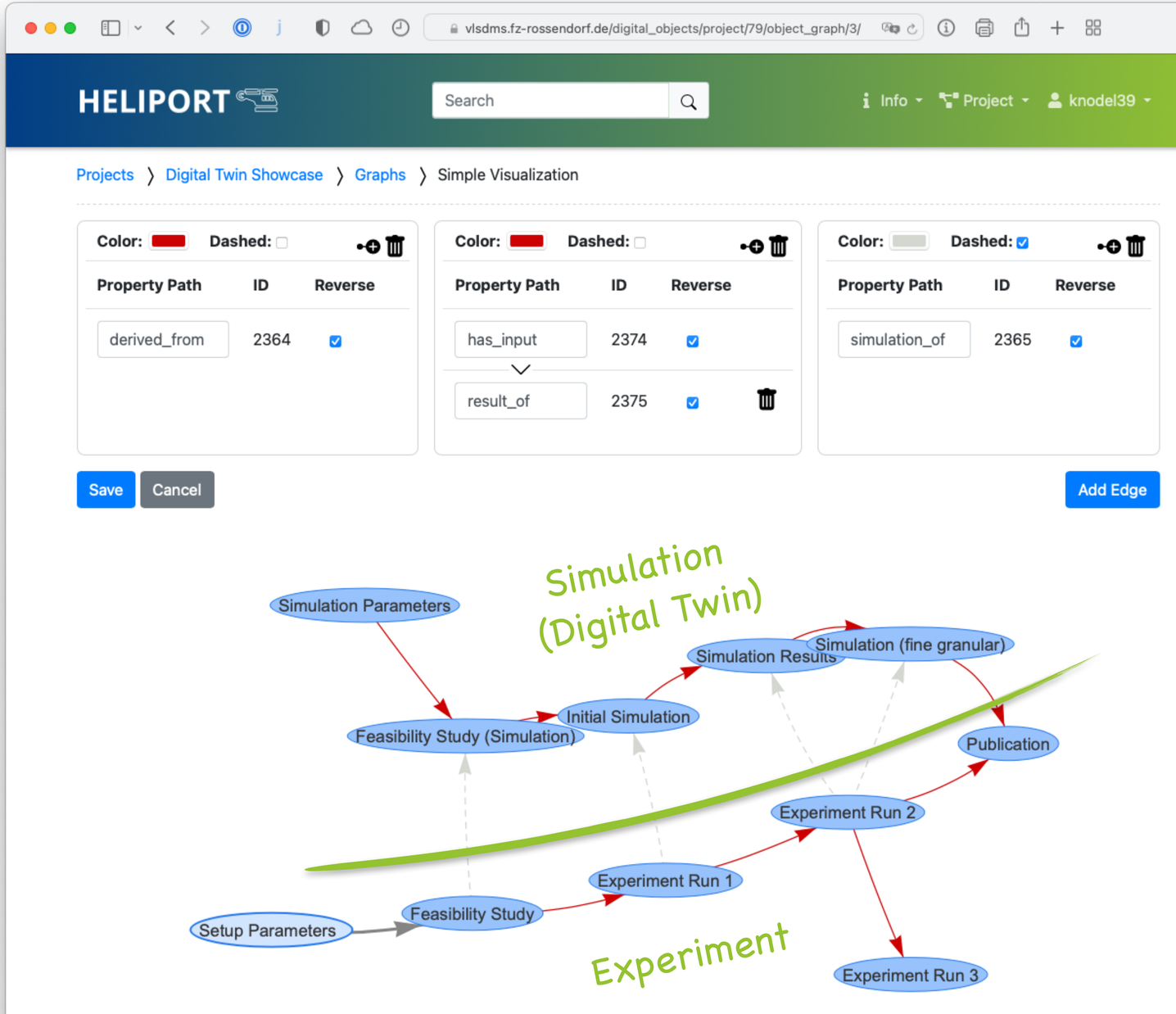


Relations Between Digital Objects

— Relations between digital objects are visualized to provide a top-level view on the project with dependencies.

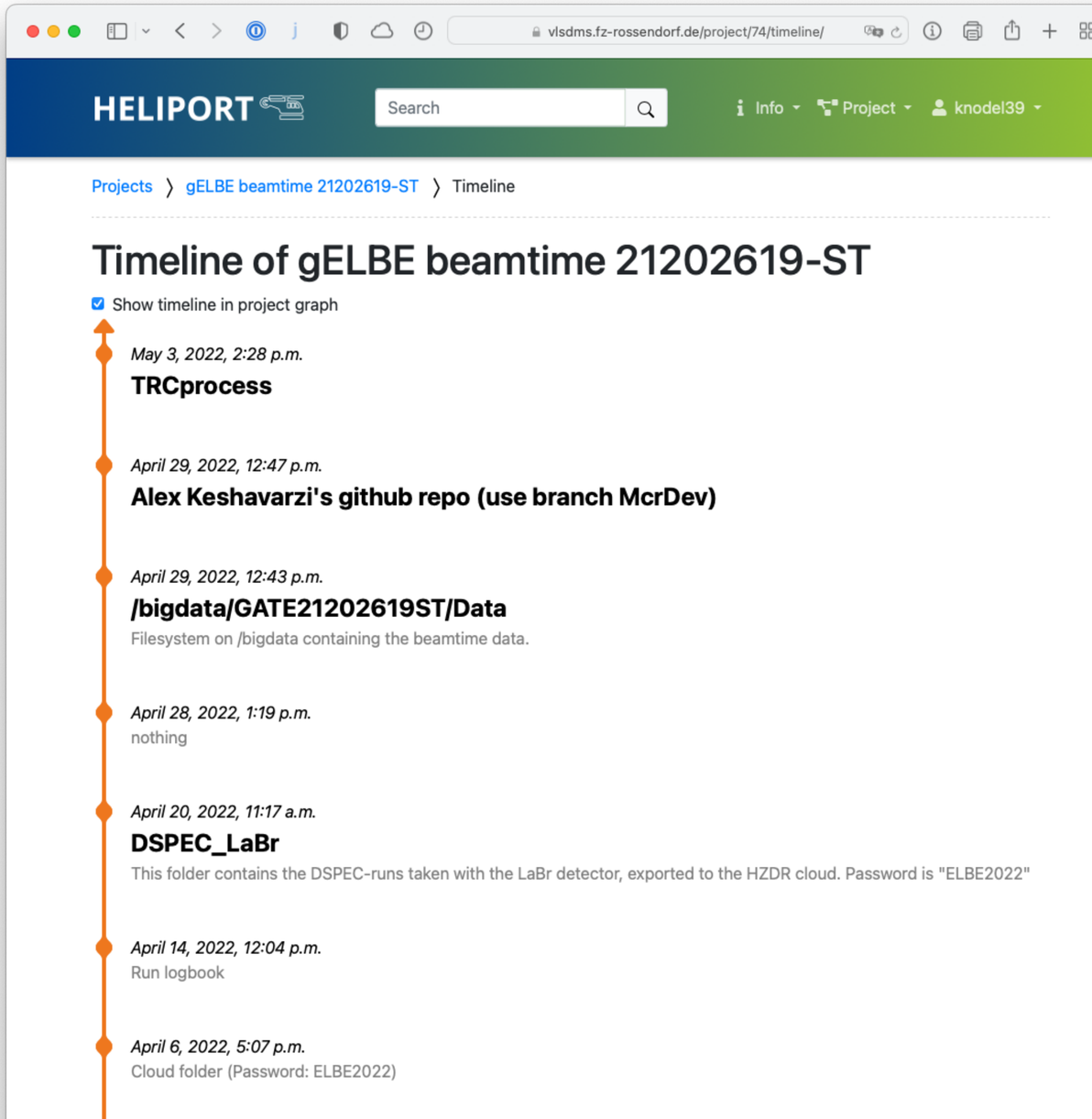
— With a graphical representation, objects and the processes on our infrastructures are comprehensible.

— Also the components and relations between simulation and experiment can be visualized.



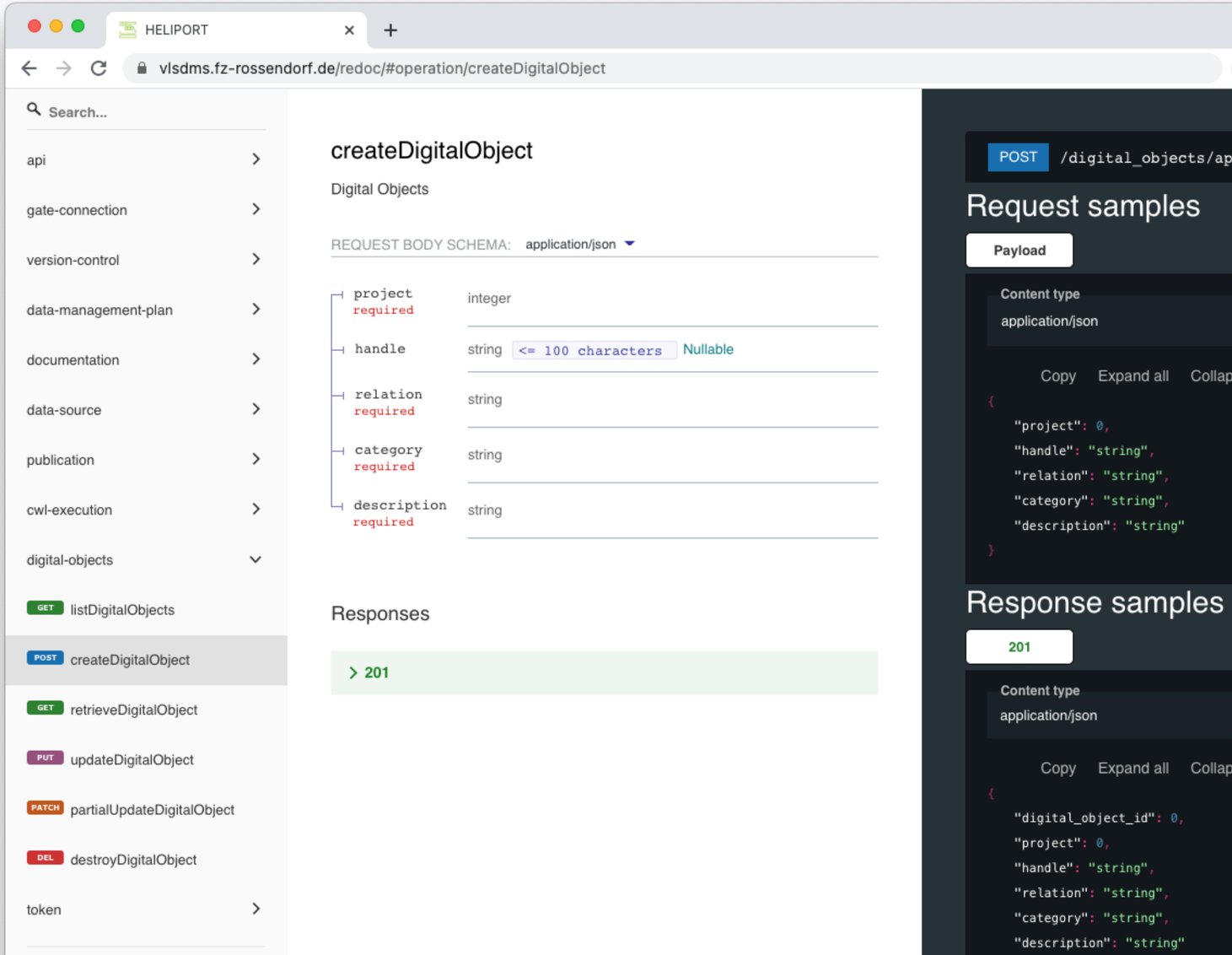
Data provenance and Comprehensibility

- 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, …)



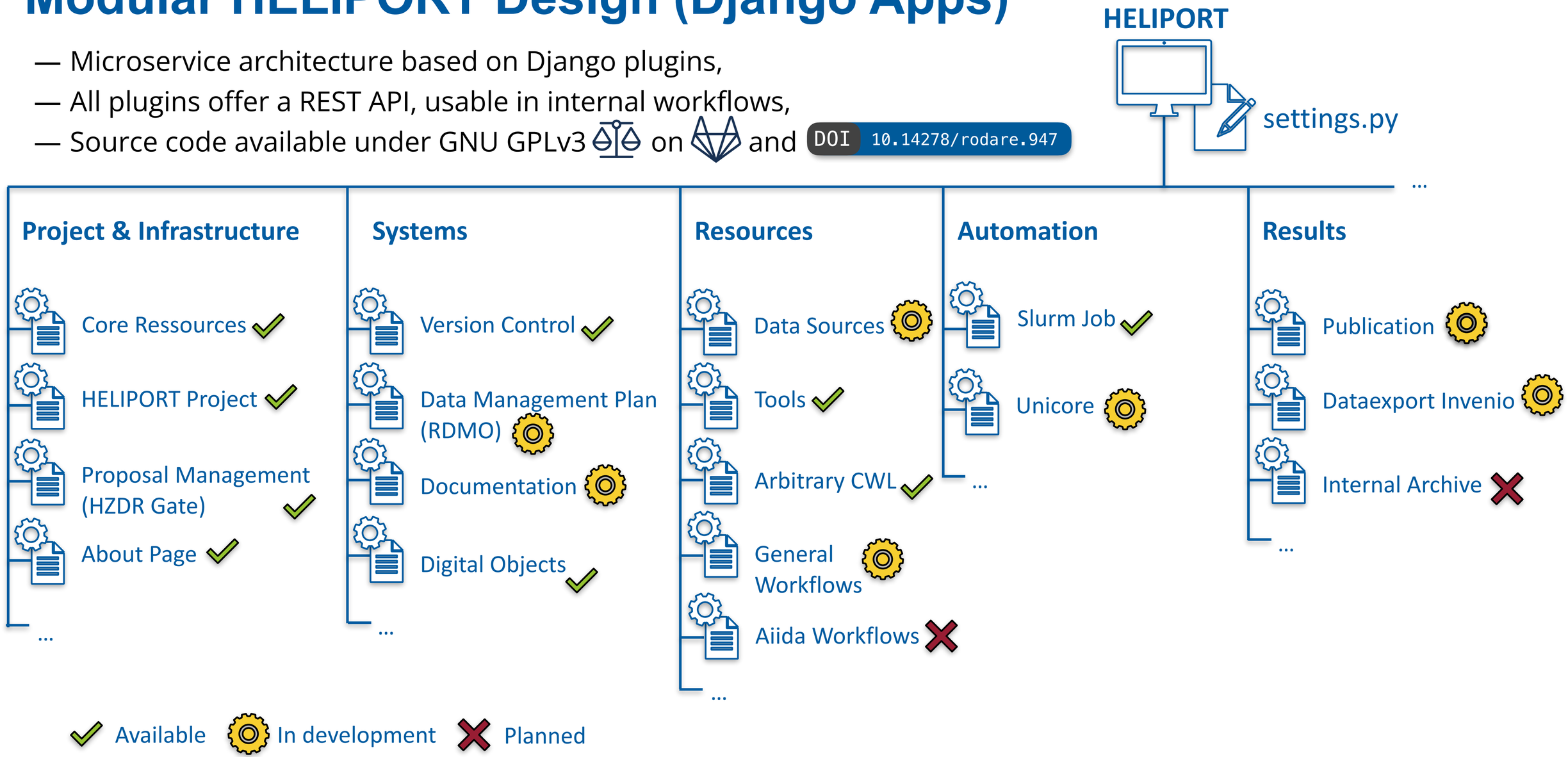
Heliport REST API

- 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.



Modular HELIPORT Design (Django Apps)

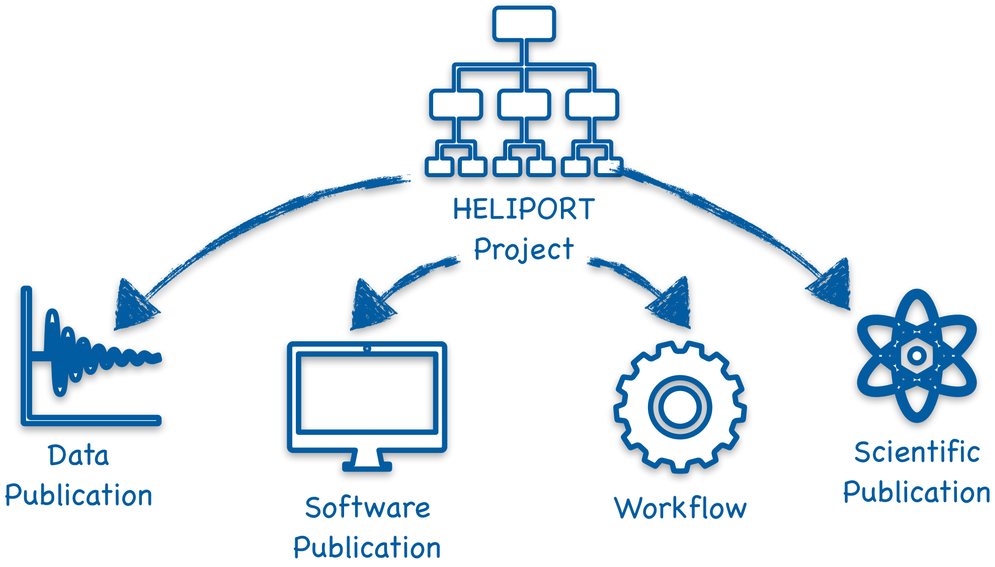
— Microservice architecture based on Django plugins,
— All plugins offer a REST API, usable in internal workflows,
— Source code available under GNU GPLv3 on and DOI 10.14278/rodare.947

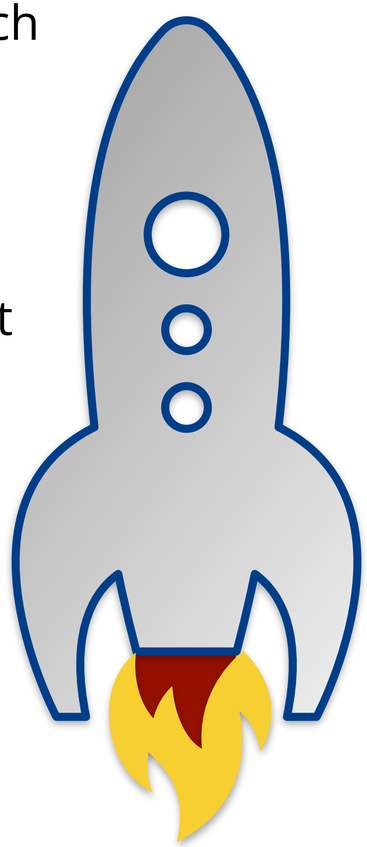


HELIPORT JÜLICH FORSCHUNGSZENTRUM HI JENA Helmholtz Institute Jena HZDR

Conclusions

- A guidance system that describes and collects all metadata from the systems involved 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.
- With all data products registered in one system we can promote the different data publications to make the research more visible and comprehensible.

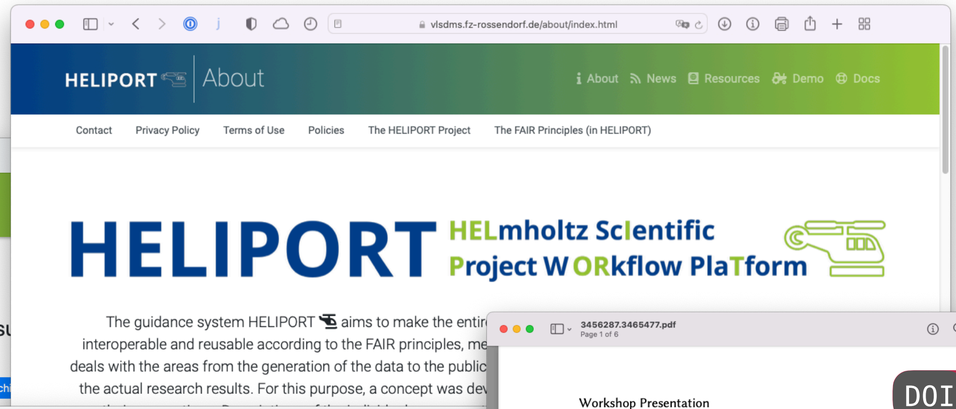


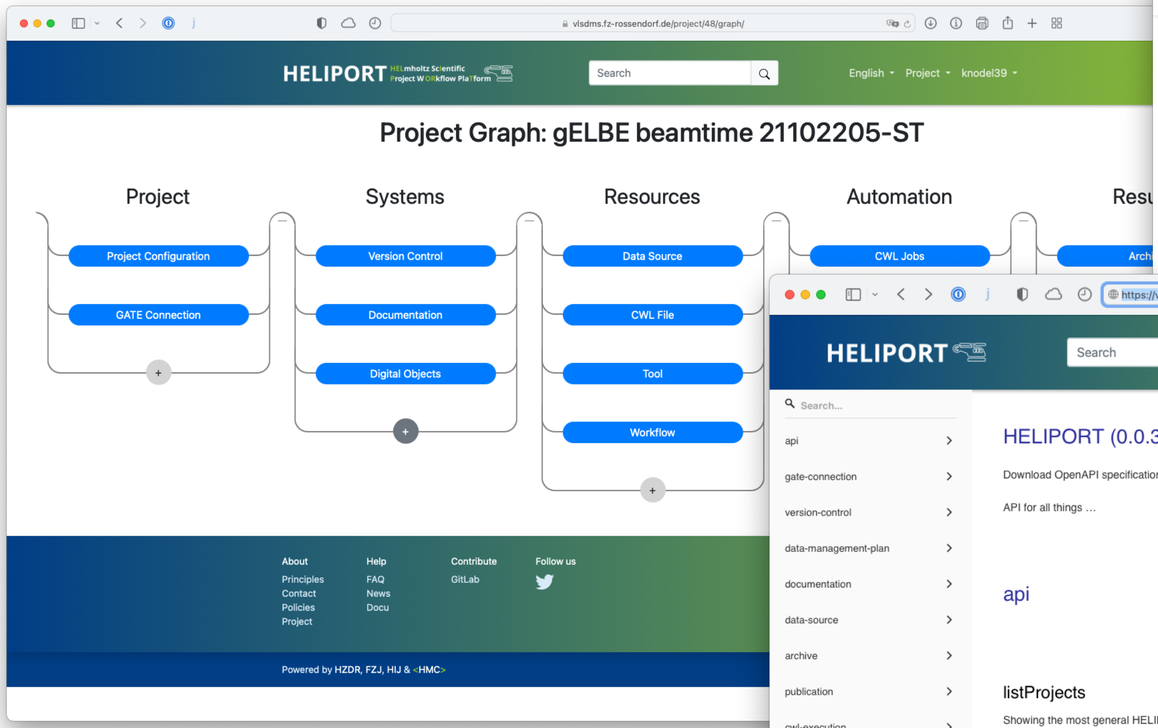


HELIPORT JÜLICH FORSCHUNGSZENTRUM HI JENA Helmholtz Institute Jena HZDR

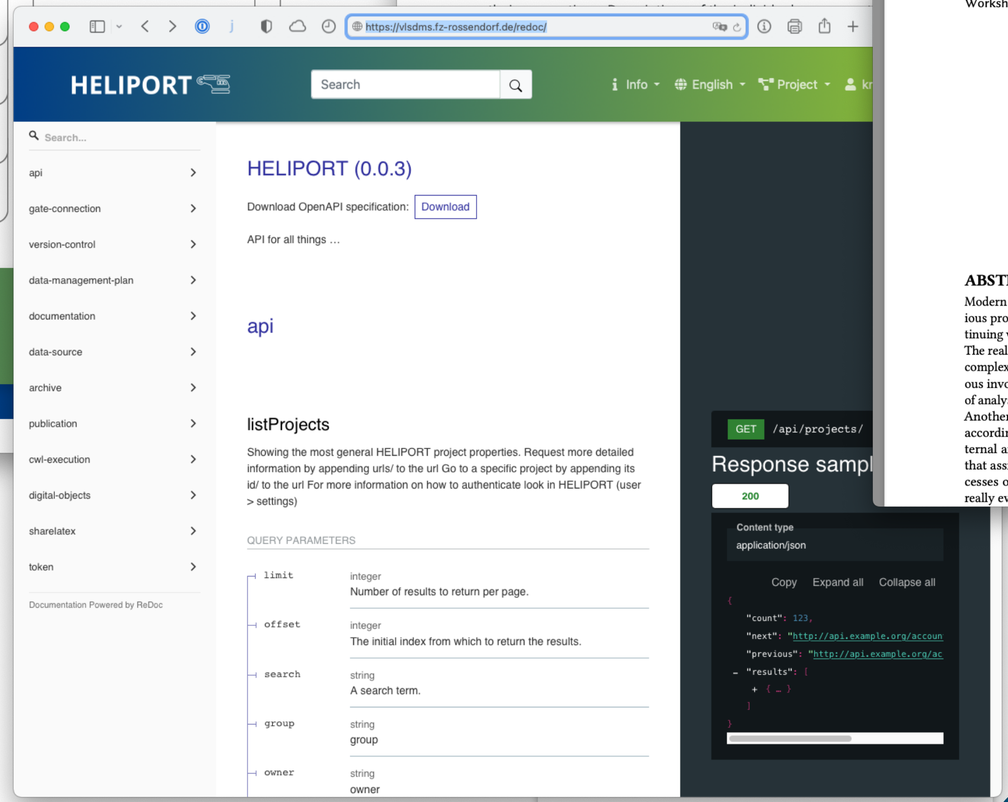
Resources

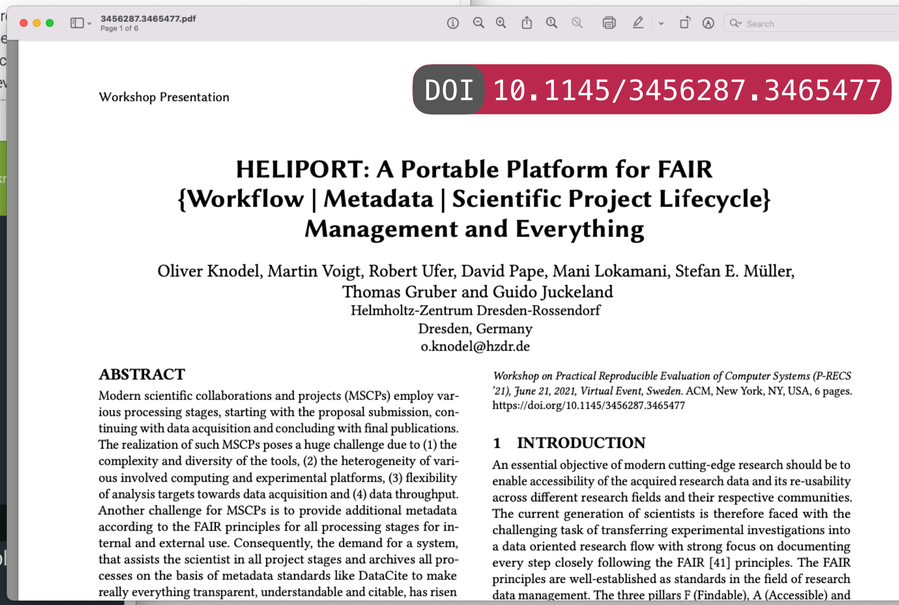
Website: heliport.hzdr.de





Demo: heliport.hzdr.de/app





API Doc: heliport.hzdr.de/redoc/

JÜLICH FORSCHUNGSZENTRUM HI JENA Helmholtz Institute Jena HZDR