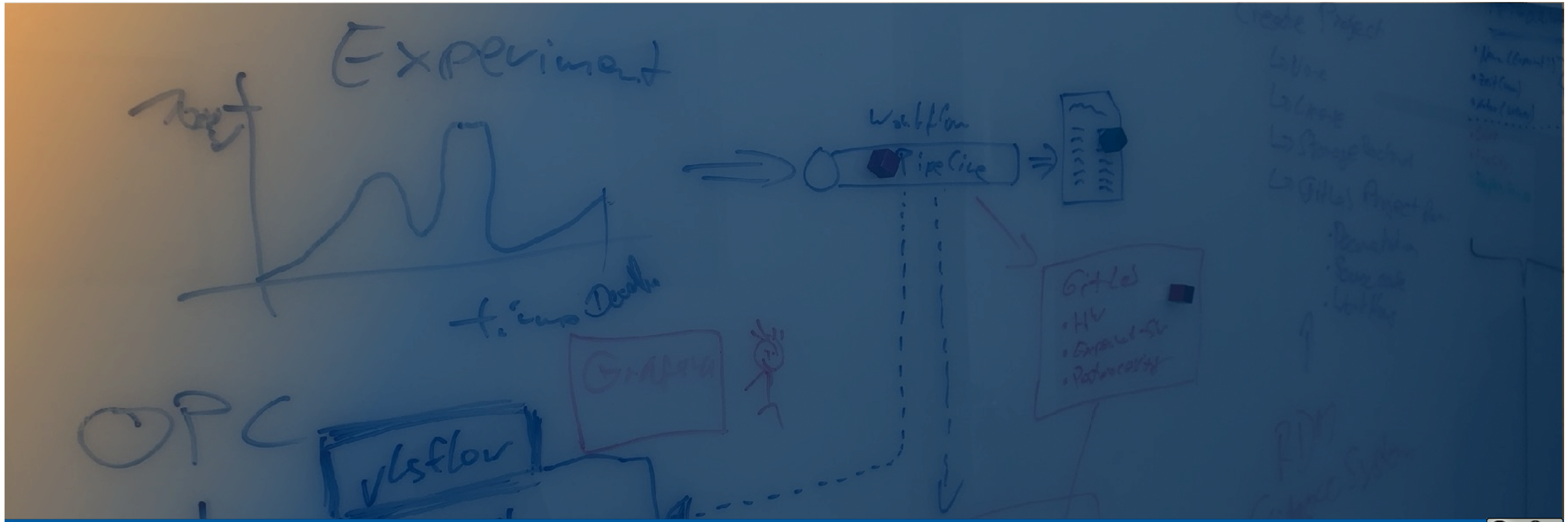
HELIPORT

HZDR
HELMHOLTZ ZENTRUM
DRESDEN ROSSENDORF



Full Integrated Research Data Lifecycle – The Project HELIPORT

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

DRESDEN concept

Our Research Facility and our Large Scale Research Infrastructures

The Helmholtz-Zentrum Dresden - Rossendorf

— Employees approx. 1,200. Thereof 600 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.

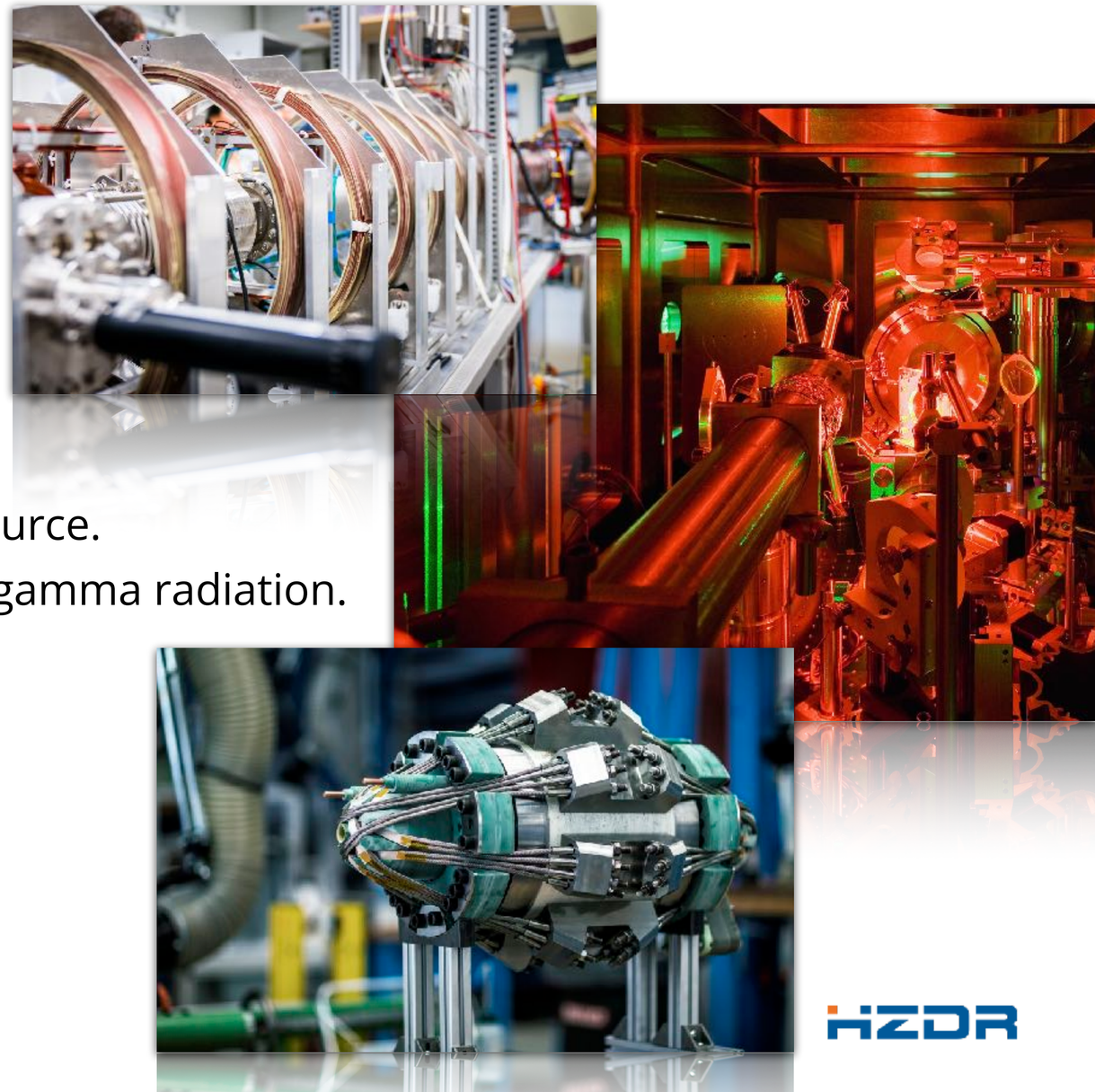
— Positrons, protons, neutrons as well as X-ray and gamma radiation.

Dresden High Magnetic Field Laboratory (HLD)

— Europe’s highest pulsed magnetic fields.

Ion Beam Center (IBC)

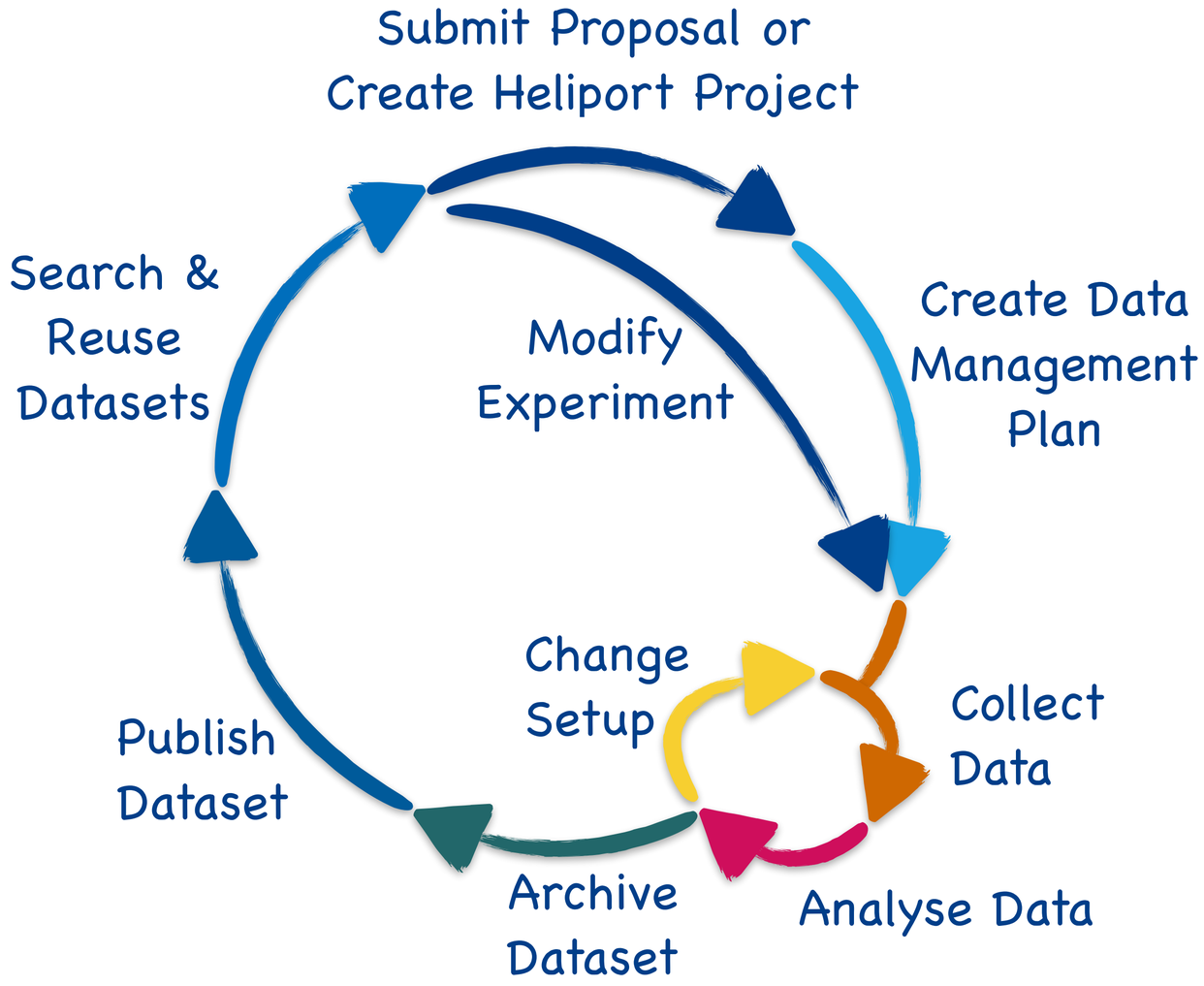
— Nanoscale surface analysis and modification.



HZDR

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

— We support many steps of a research experiment with tools:
  - electronic lab books,
  - interactive analysis,
  - publication of datasets,
  - scientific workflow management,
  - Handle generation and management.
— A uniform access to all services and systems 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

…and is Your Research Data Usable?

Make your data

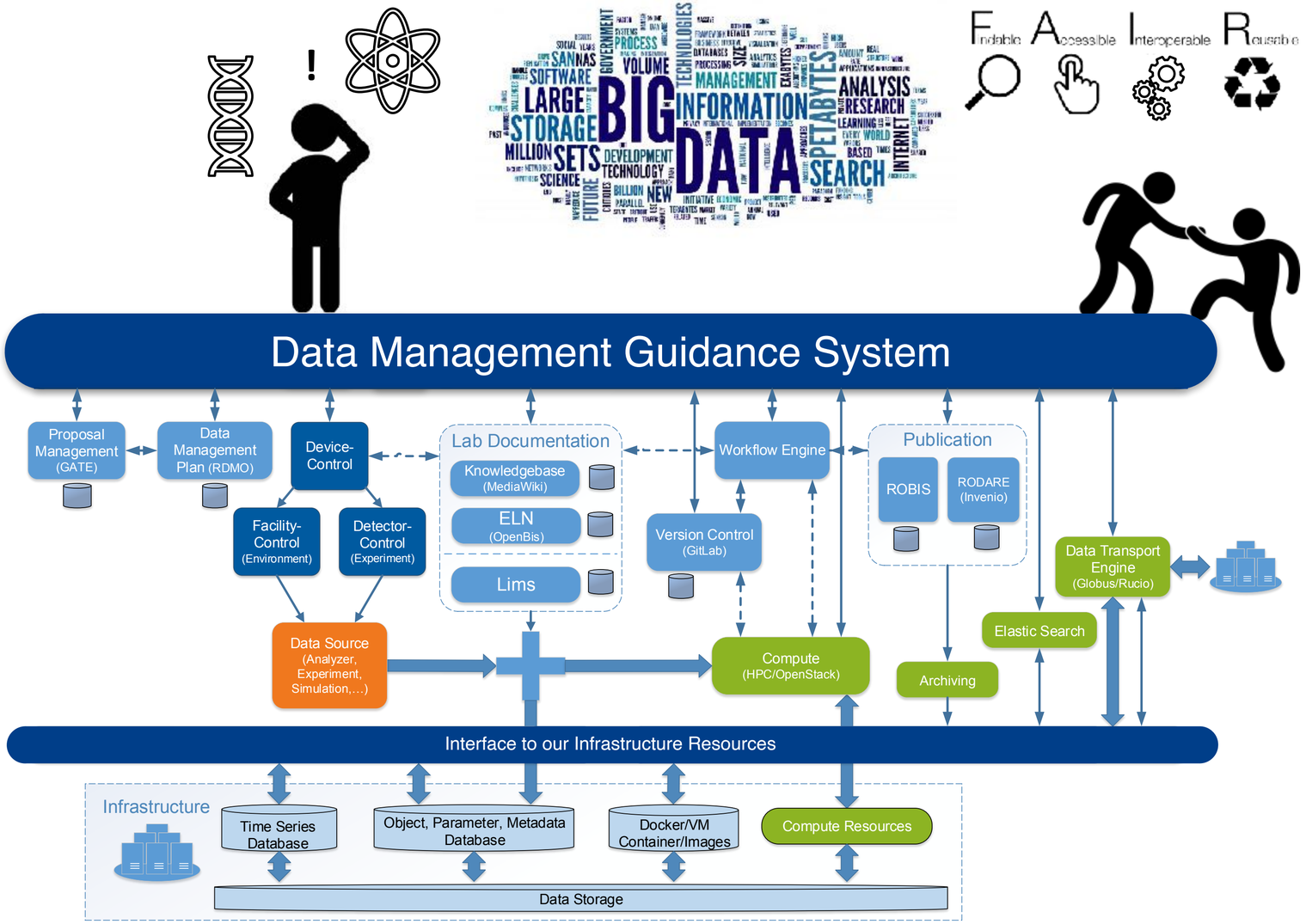
Findable, **A**ccessible, **I**nteroperable and **R**e-usable

FAIR data principles

Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016).

Our Observations and Experiences

- Our HZDR IT infrastructure can support various experiments, but it is 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?

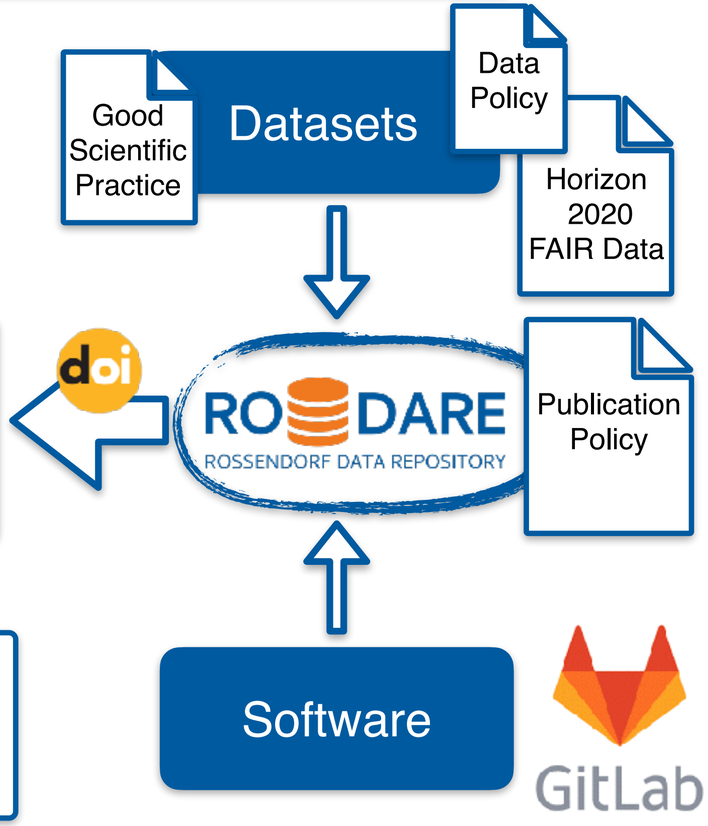
And how we can support them?!

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?

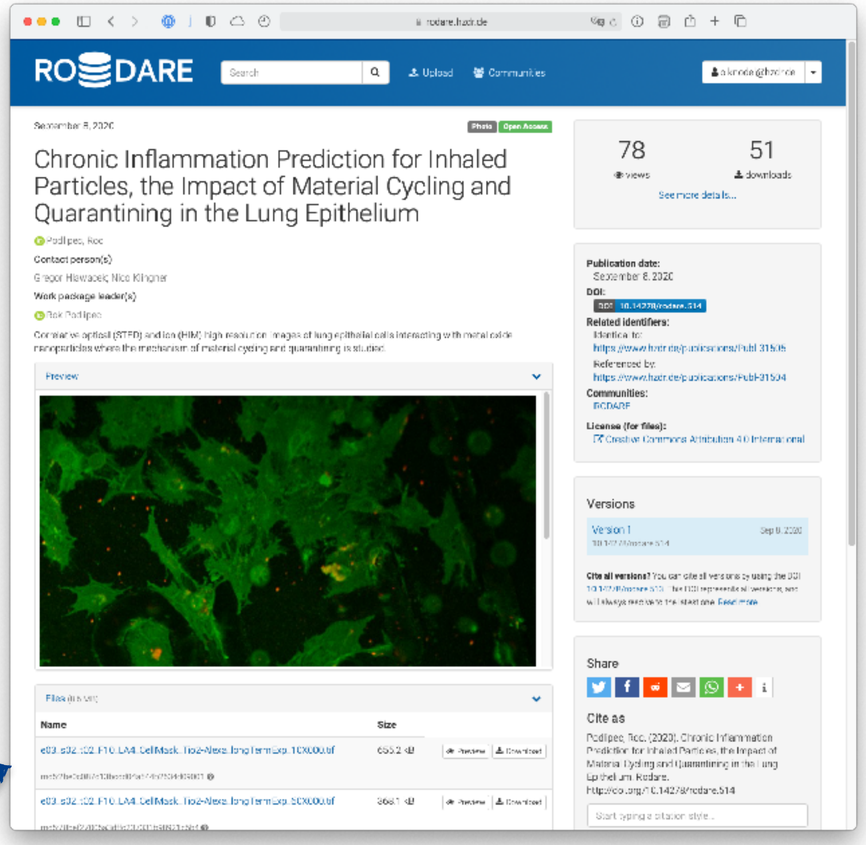
Where are data, software and how can I gain **access** to both of them?



HELIPORT JÜLICH FORSCHUNGSZENTRUM HI JENA Helmholtz Institute Jena HZDR

How we went on with our "~~Data Management Guidance System~~"

- We required a management environment supporting our project lifecycles.
- Based on our observations and experiences in the field we started developing Heliport:
  - Metadata becomes important in modern research to make every founded project comprehensible and FAIR,
  - The publication of all data products and the Data Management Plan (DMP) becomes inevitable.
  - The abstraction from all underlying services with the use of REST APIs and workflows is a key concept of Heliport.
- Heliport can fill the gap between all **data products** stored in our various systems and the **final publications** of these products in our data repository RODARE.



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:

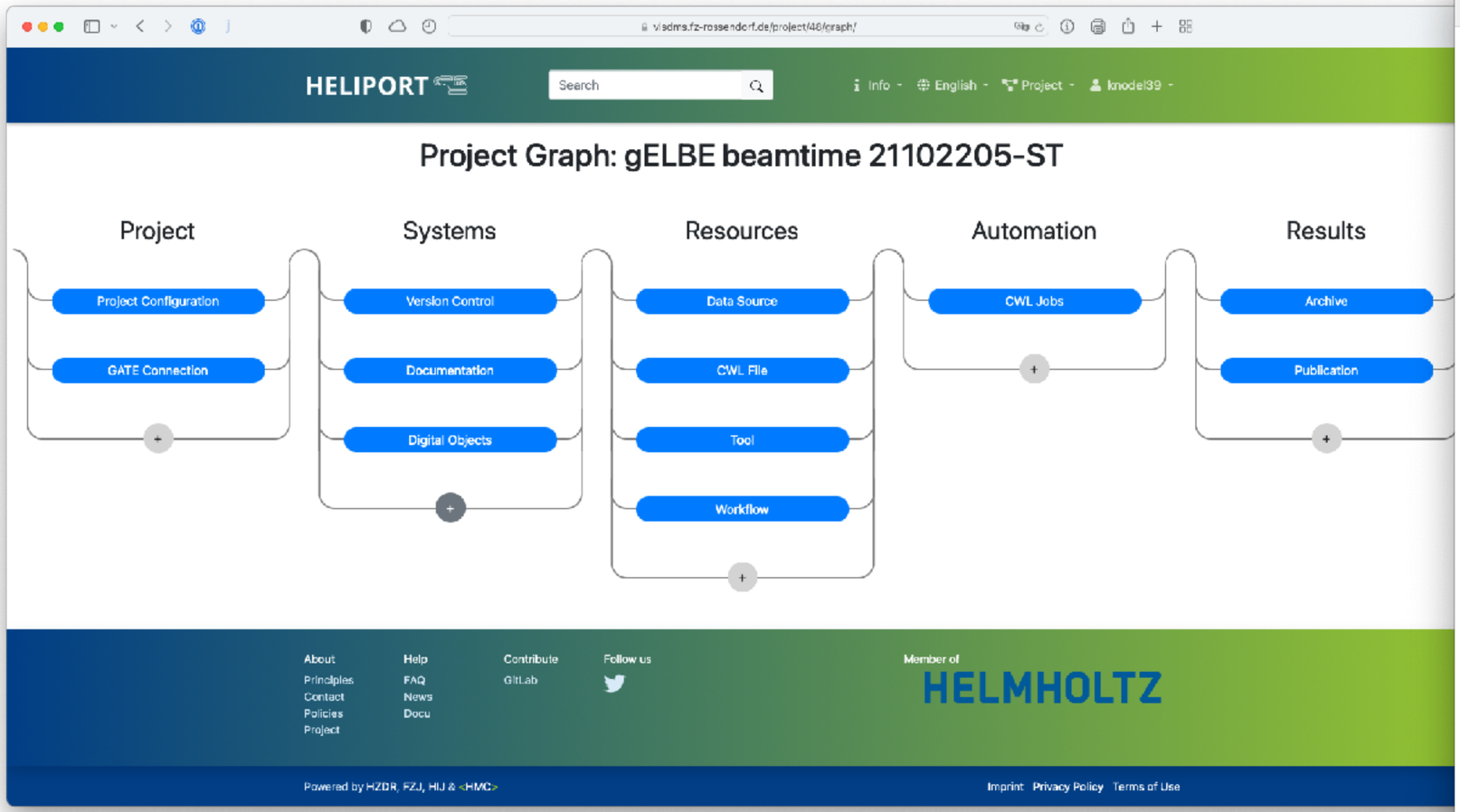
HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

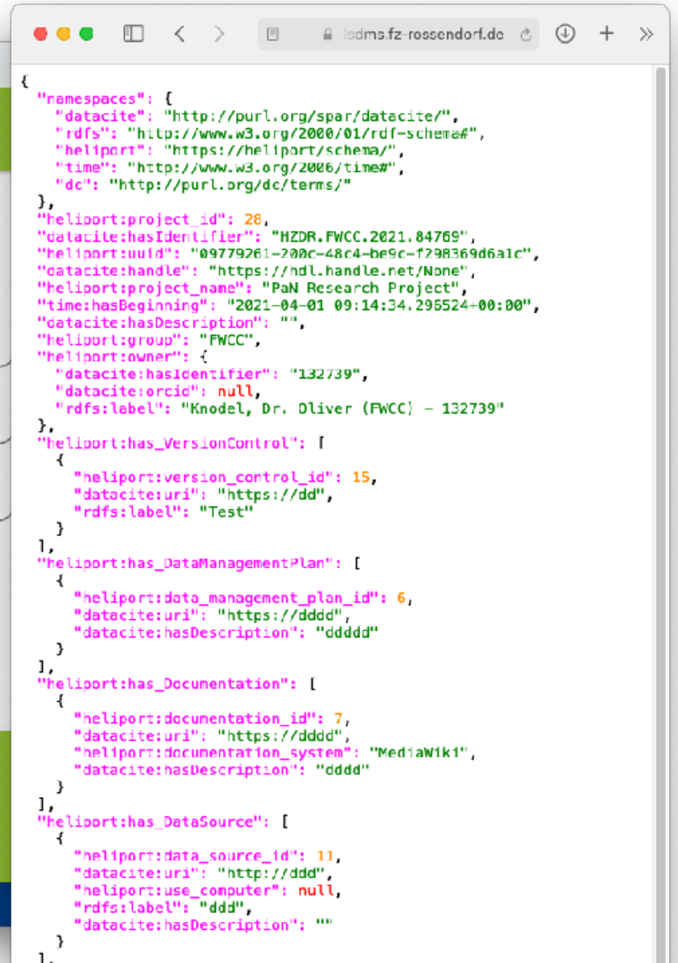
HI JENA Helmholtz Institute Jena

JÜLICH FORSCHUNGSZENTRUM

Founded by:

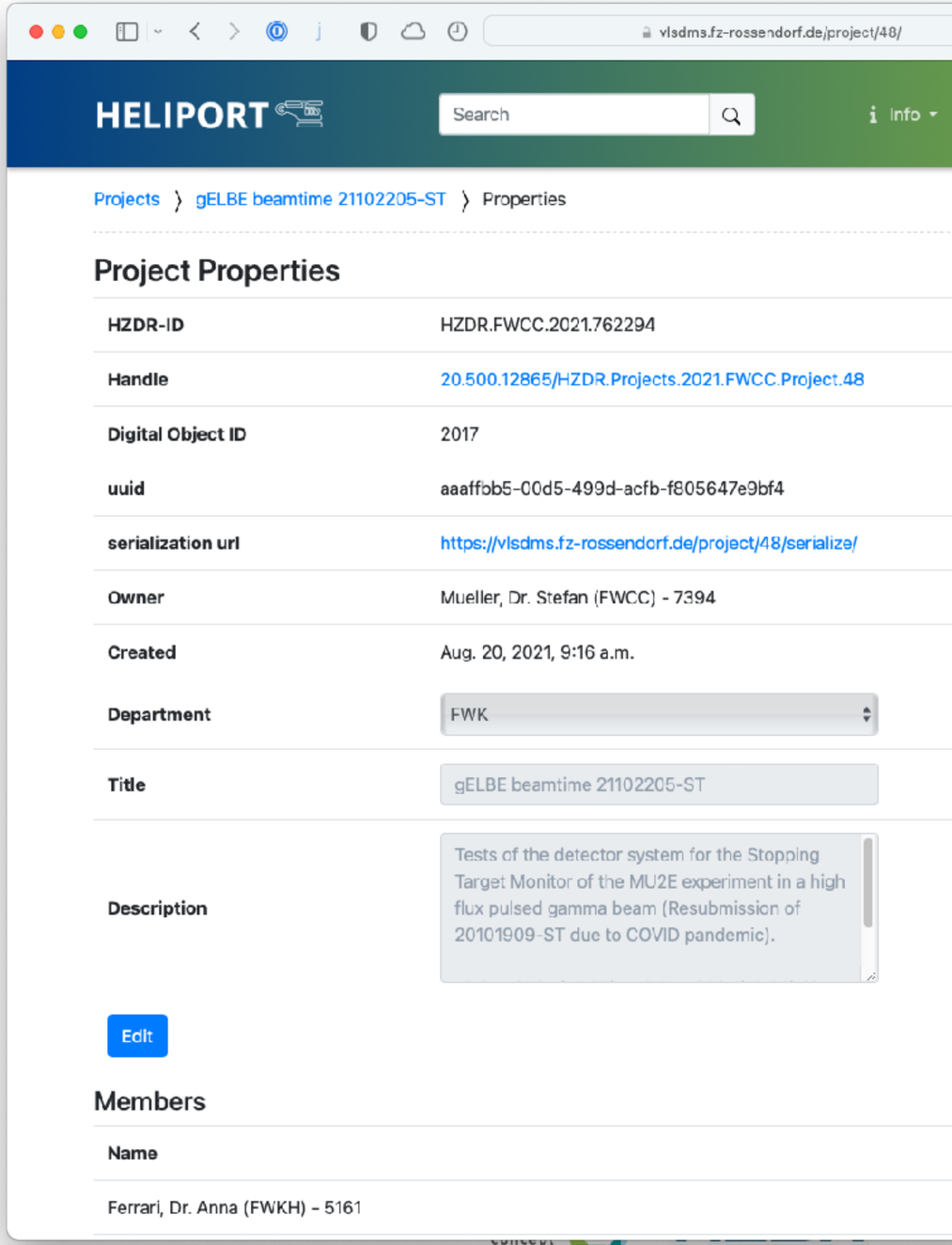
HMC HELMHOLTZ METADATA COLLABORATION





The Project Metadata Scheme


```
{
  "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": 9,
  "datacite:hasIdentifier": "HZDR.FWCC.2021.95018",
  "heliport:uuid": "8fab8a14-0f2f-414d-bbe0-747c38129bc4",
  "datacite:handle": "https://hdl.handle.net/20.500.12865/HZDR.FWCC.2021.95018",
  "heliport:label": "An Example Project",
  "time:hasBeginning": "2021-05-18 13:03:34.378458+00:00",
  "datacite:hasDescription": "This Project has the sole purpose of demonstrating the functionality of HELIPORT",
  "heliport:group": "FWCC",
  "heliport:owner": {
    "datacite:hasIdentifier": "141575",
    "datacite:orcid": "https://orcid.org/0000-0001-5556-838X",
    "rdfs:label": "Voigt, Martin (FWCC) - 141575"
  },
  "heliport:co_owners": [
    {
      "datacite:hasIdentifier": "132739",
      "datacite:orcid": "https://orcid.org/0000-0001-8174-7795",
      "rdfs:label": "Knodel, Dr. Oliver (FWCC) - 132739"
    },
    . . .
  ],
  "heliport:has_GATEProject": [
    {
      "heliport:gate_id": 283747364,
      "dc:title": "An Example GATE Project",
      "heliport:status": "Continuesly Improving",
      "dc:abstract": "There is an urgent need to … (ref.: 10.1038/sdata.2016.18)",
      "datacite:hasIdentifier": "sdata.2016.18",
      "dc:accessRights": true,
      "dc:creator": {
        "datacite:hasIdentifier": "141575",
        "datacite:orcid": "https://orcid.org/0000-0001-5556-838X",
        "rdfs:label": "Voigt, Martin (FWCC) - 141575"
      },
      "heliport:responsible_experimentalist": {
        . . .
      },
. . .
```

Available Proposal Metadata

Proposal Metadata

— Title, abstract, status

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

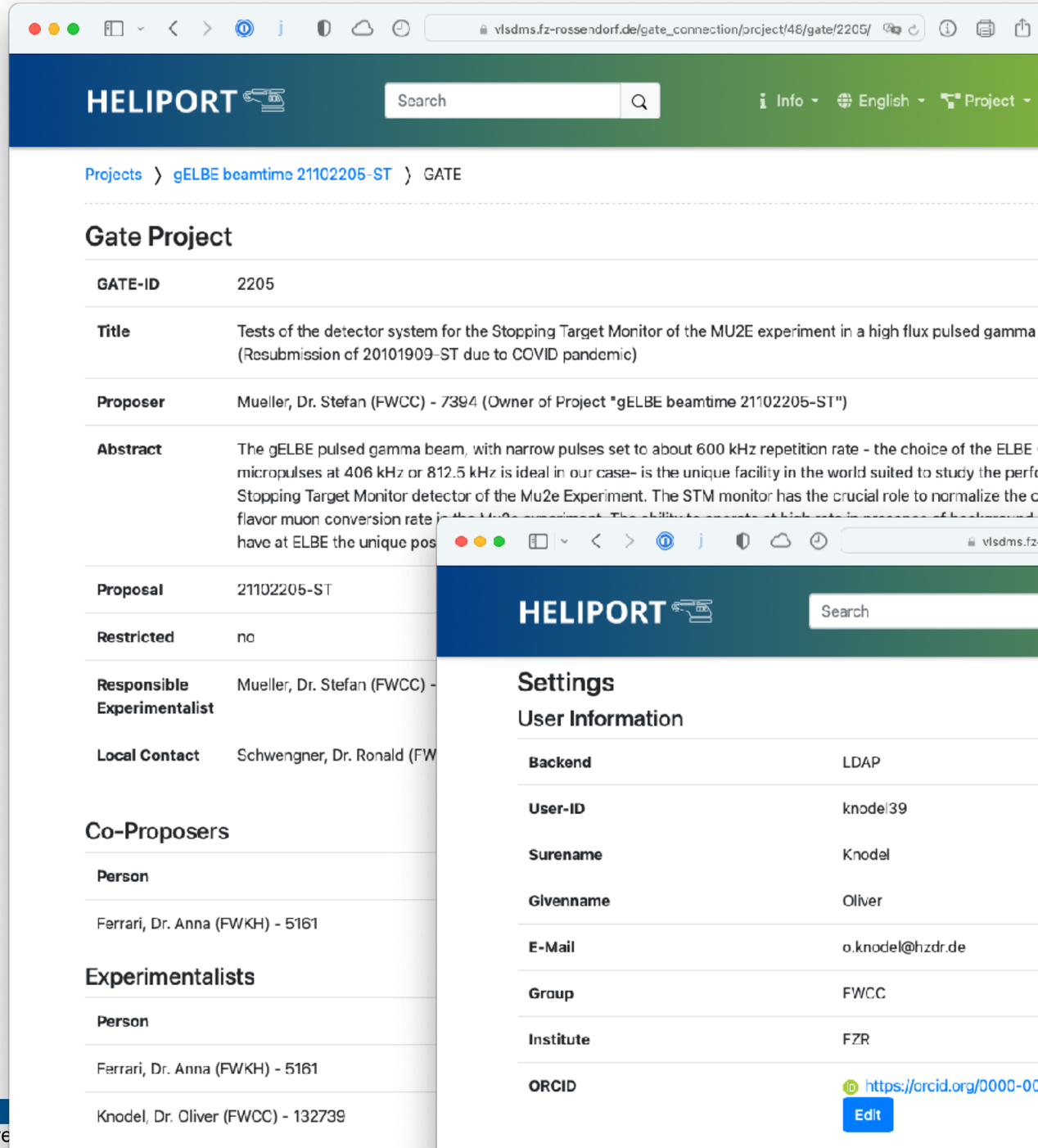
Future Metadata Fields

— Instrument or beam-line,

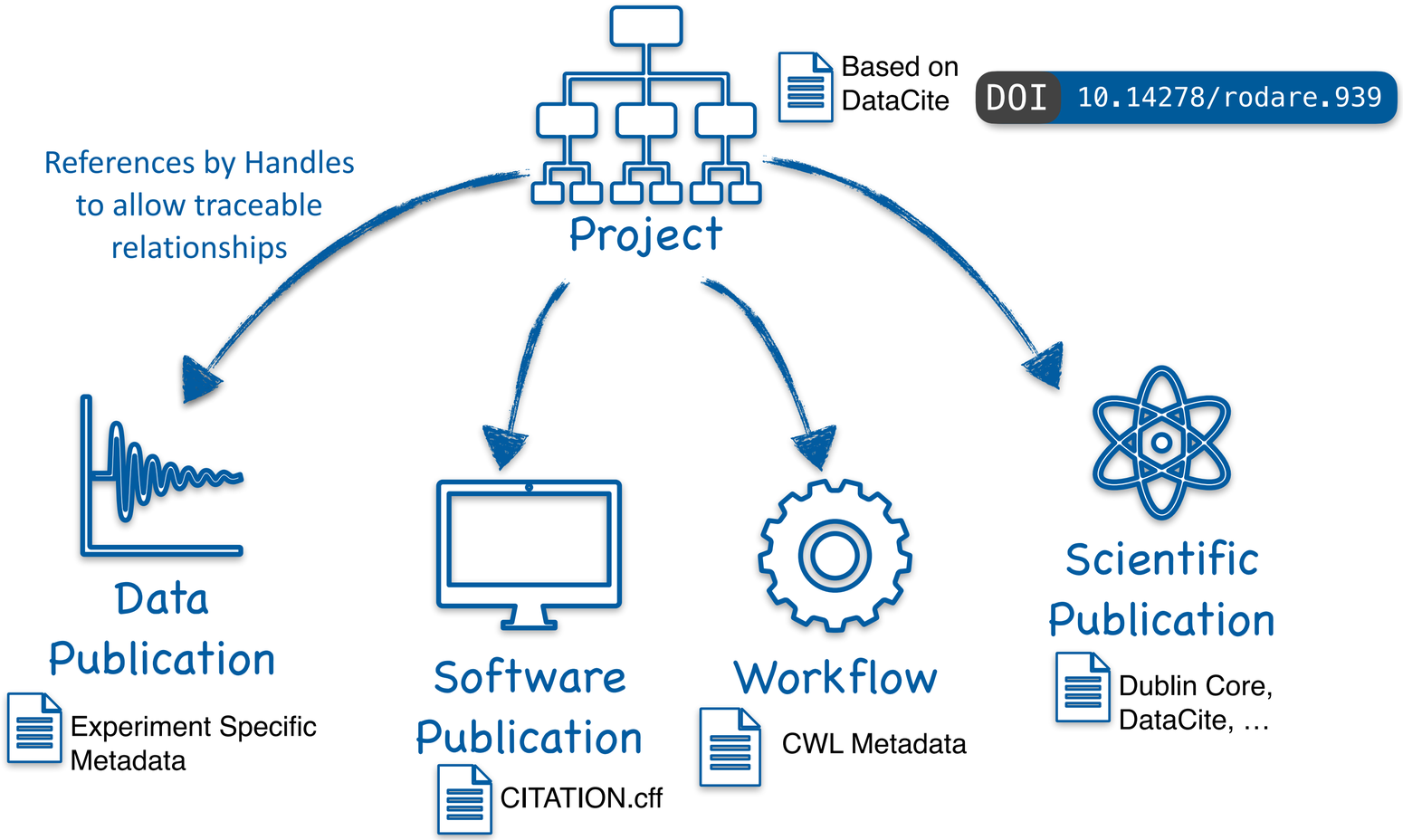
— Funding,

— Scientific method,

— ….



Heliport Metadata Ecosystem

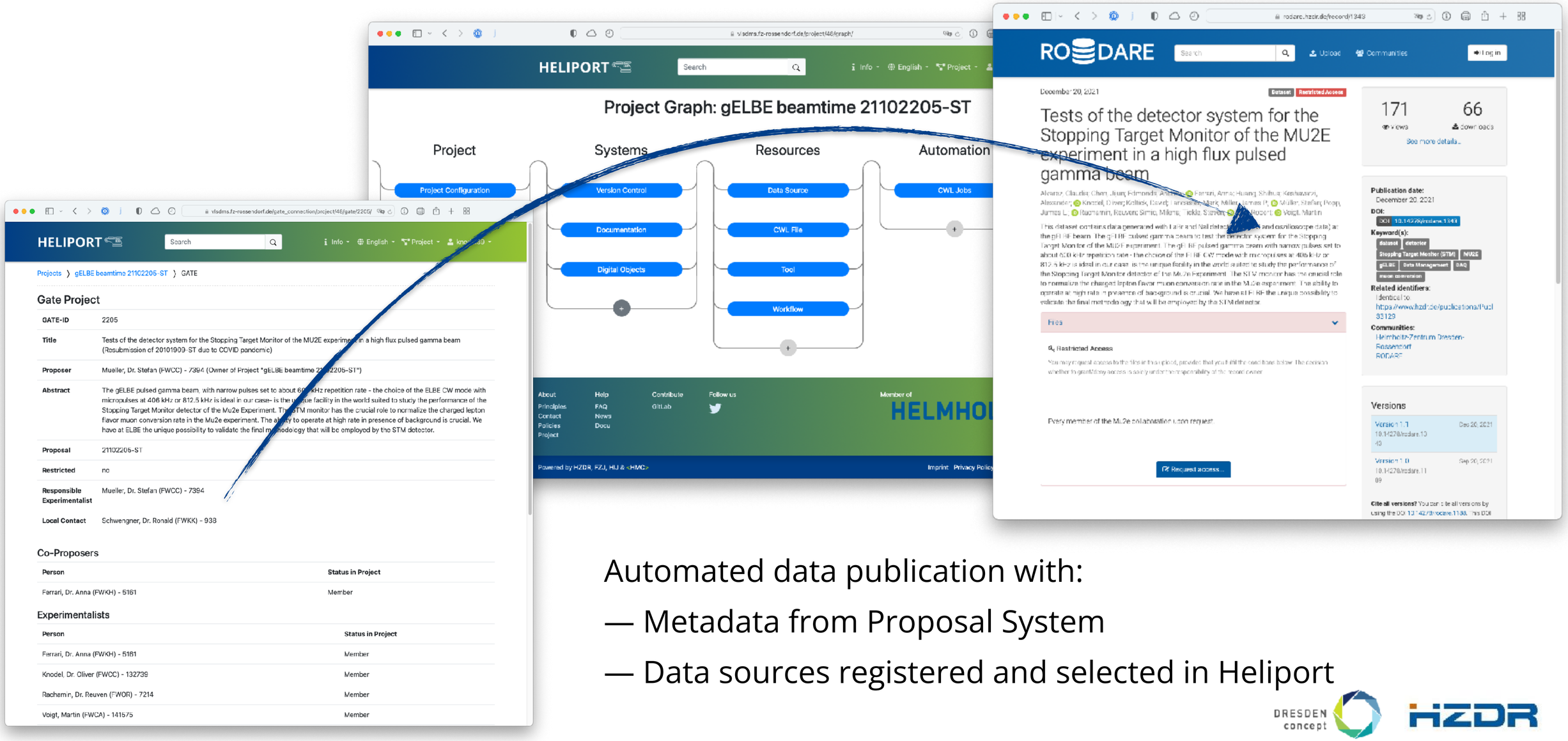


Our Concept

- 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.

HELIPORT JÜLICH FORSCHUNGSZENTRUM HI JENA Helmholtz Institute Jena HZDR

Integration in an Overall Data Management



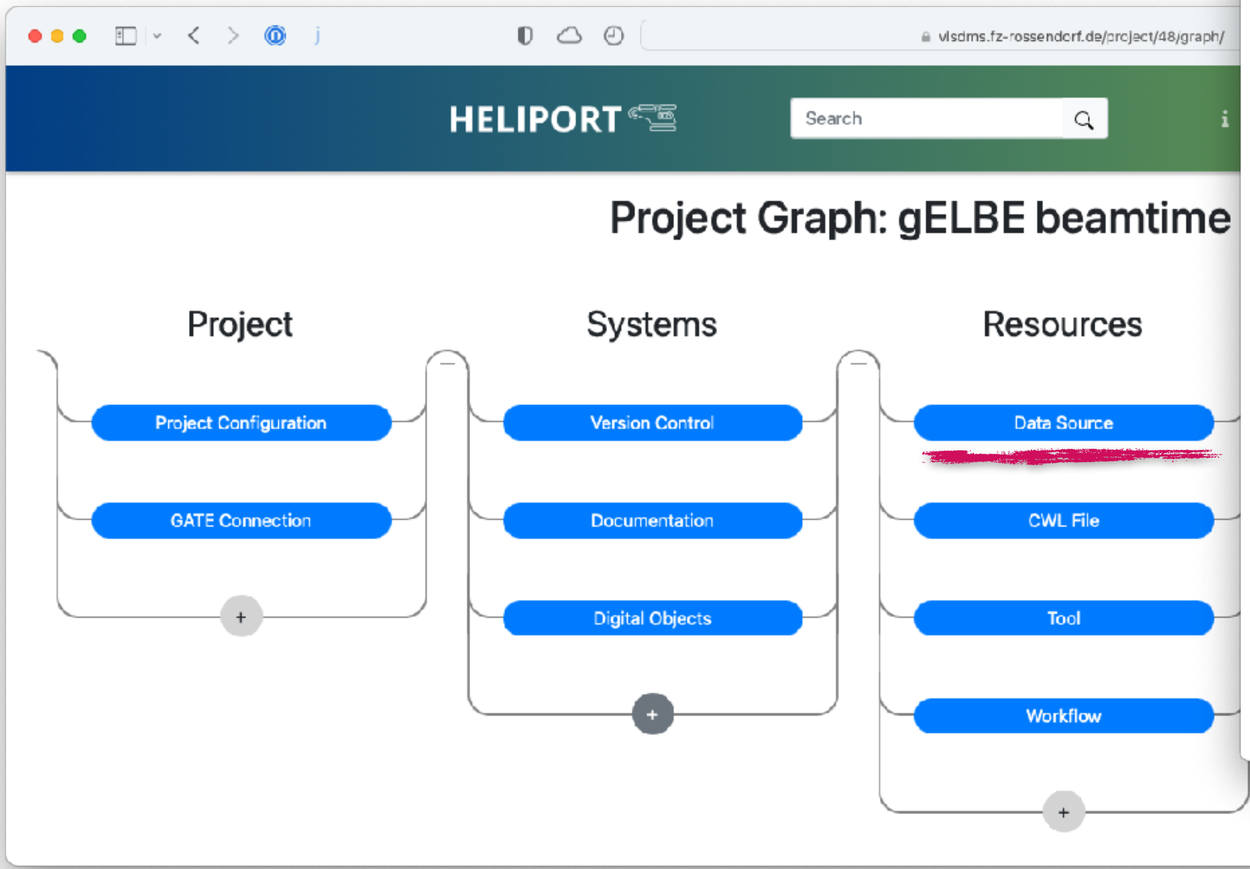
Automated data publication with:

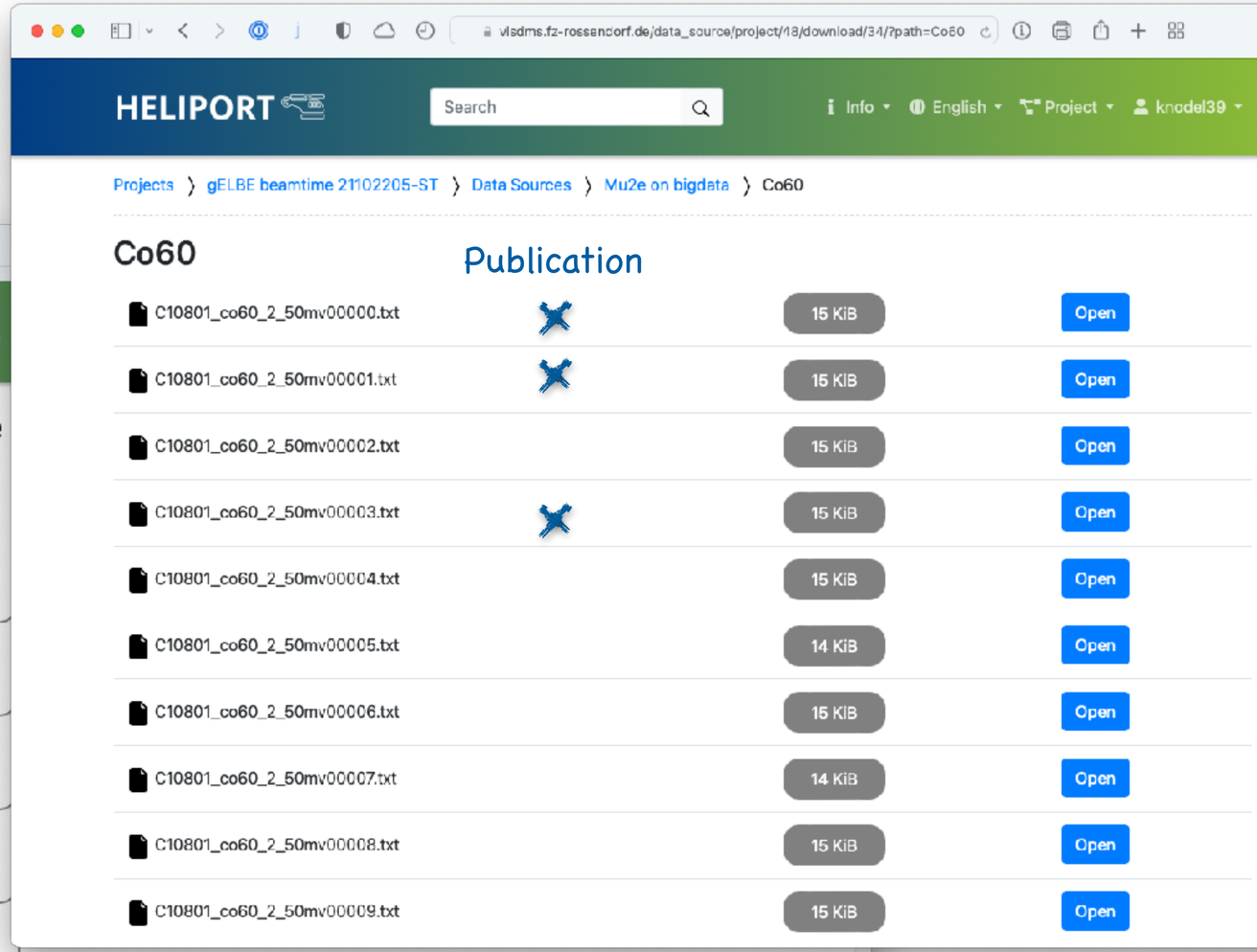
— Metadata from Proposal System

— Data sources registered and selected in Heliport

Data Sources

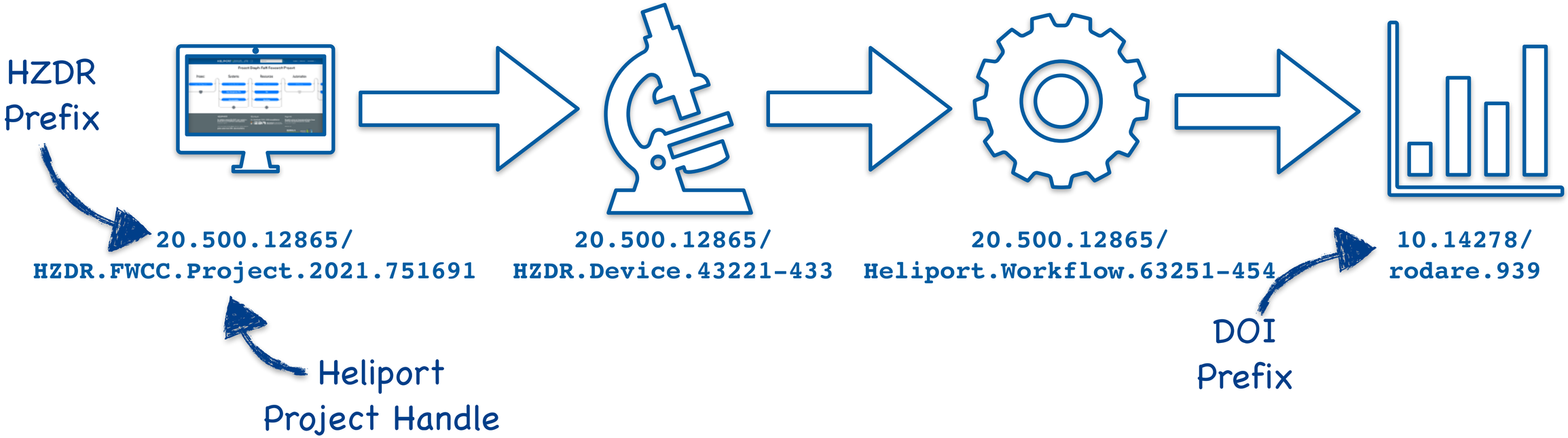
..and files (selected for publication) can be transferred directly to the data publication (development in progress).





Persistant Identifiers (Handles) in Heliport

Heliport is linked with our local Handle-Server (handle.hzdr.de) hdl enabled and generates uniform PIDs (resolvable using hdl.handle.net) from and for various systems and services. Associated information can be changed as needed without changing the identifier.



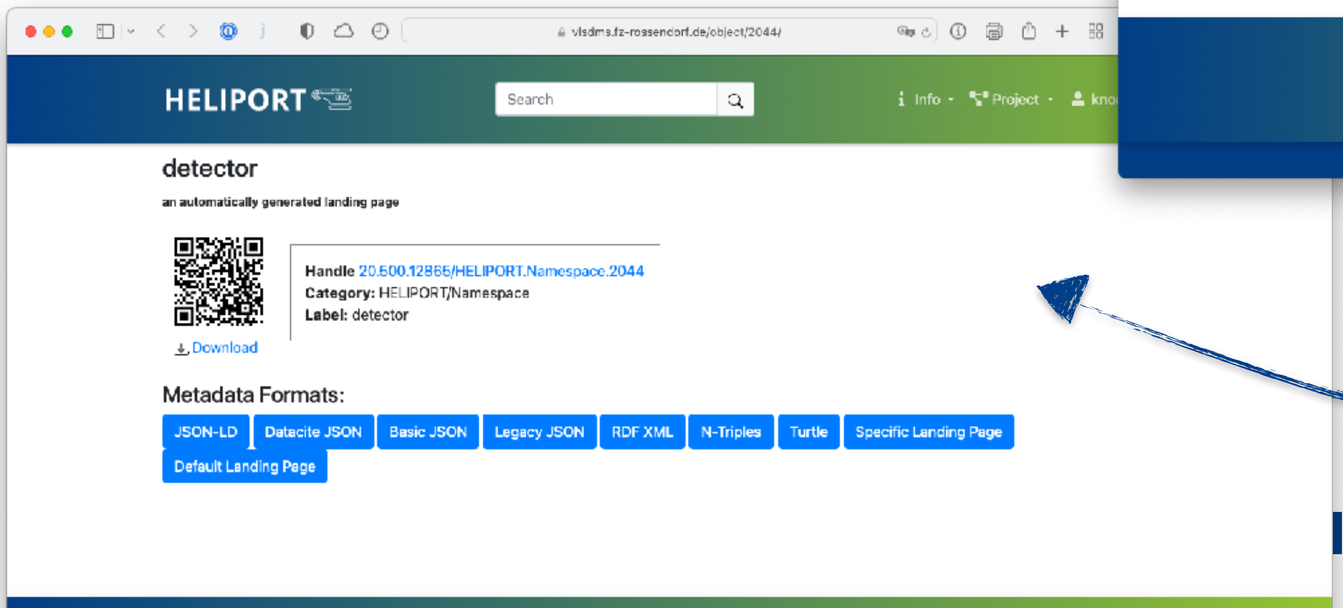
HELIPORT JÜLICH FORSCHUNGSZENTRUM HI JENA Helmholtz Institute Jena HZDR

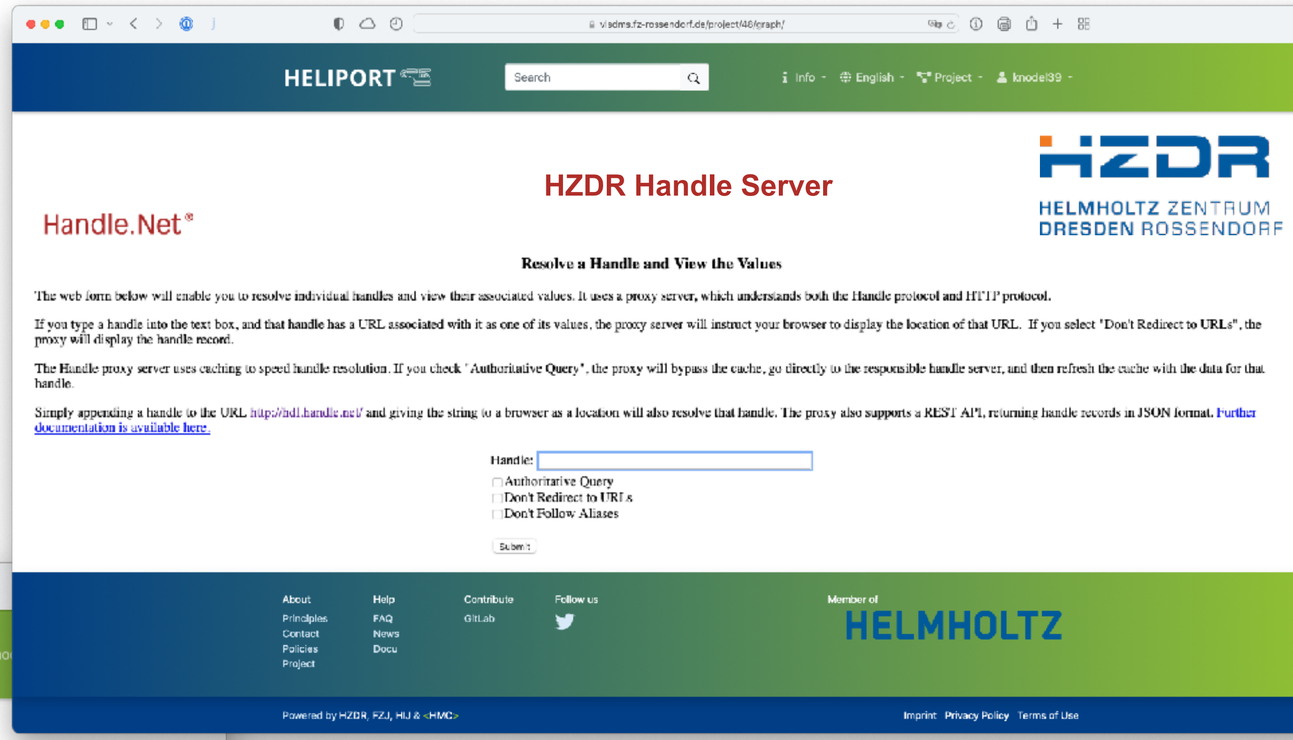
(Future) Handle Management Infrastructure in Heliport

— Every *digital object* in HELIPORT is registered at our public available Handle server: handle.hzdr.de

— For ever Handle we provide a landing page with the associated public metadata.

— A link to the object in the HELIPORT project is available when the user has access to the project and the internal HZDR network.

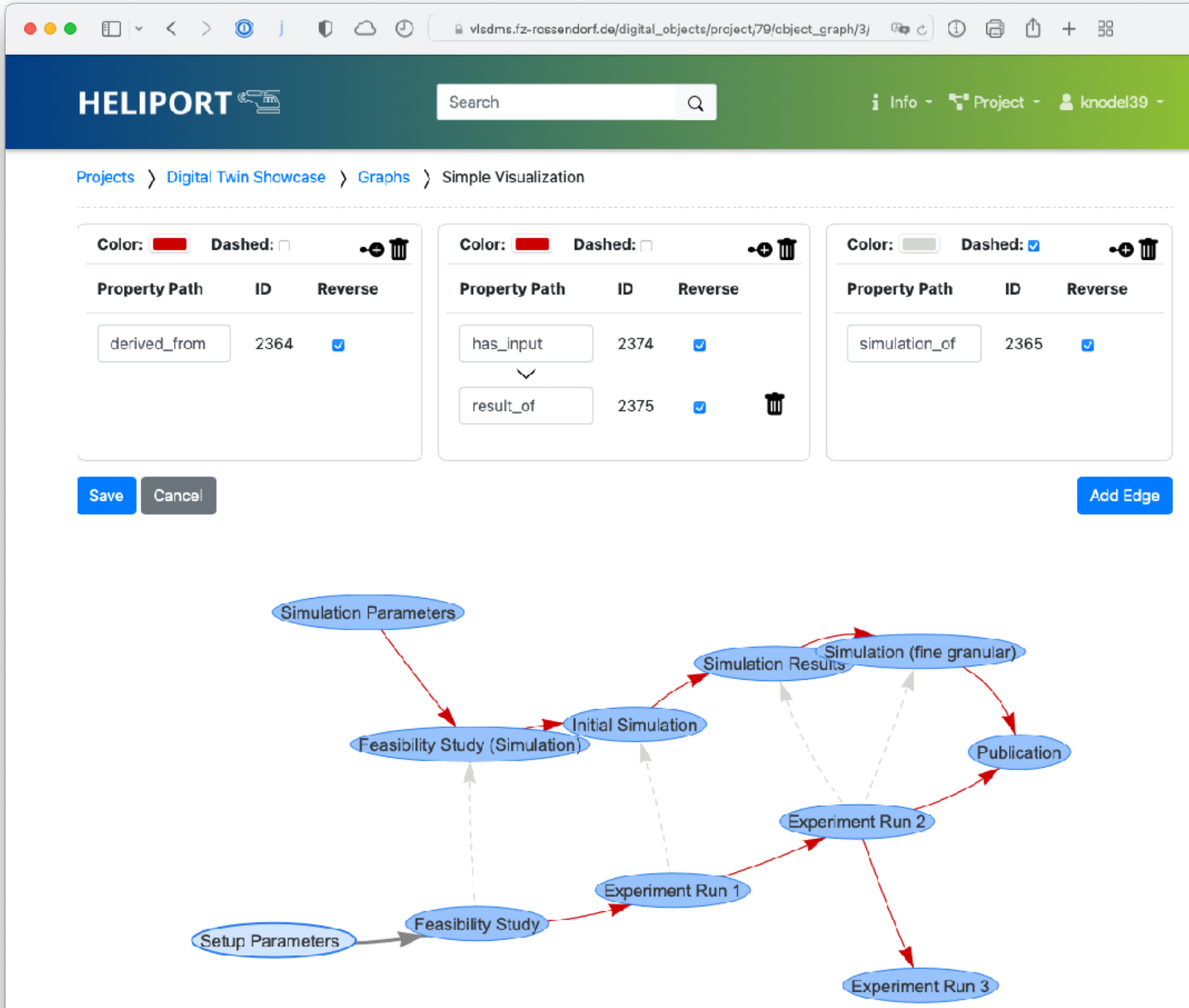




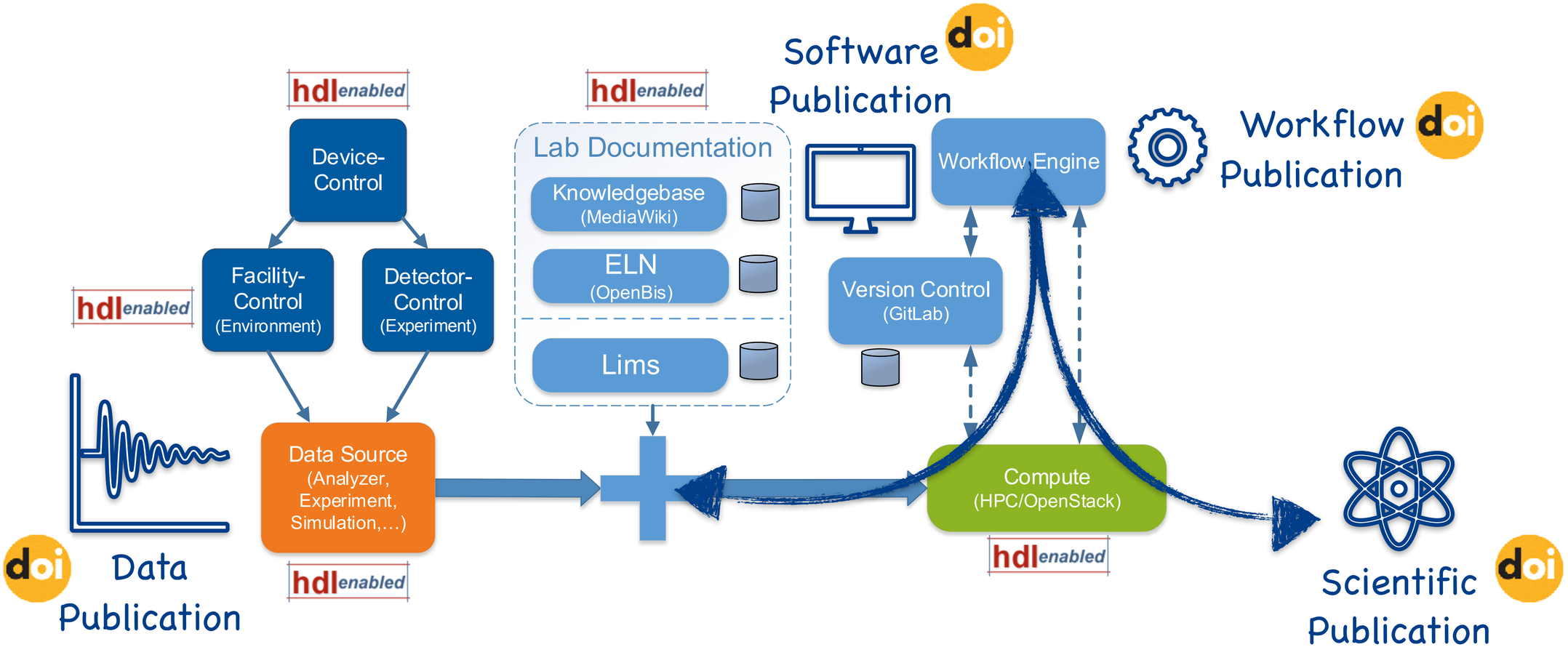
Relations Between Digital Objects

— The relations between digital objects is visualized to provide an additional view on the scientific project with dependencies.

— With the graphical representation, the objects and also the processes on our computing infrastructure can be understood more clearly.

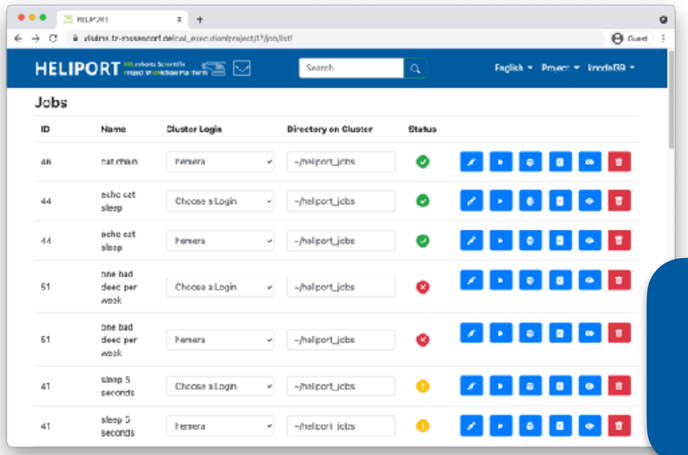


Different Types of (Data) Publications and Data Provenance



HELIPORT JÜLICH FORSCHUNGSZENTRUM HI JENA Helmholtz Institute Jena HZDR

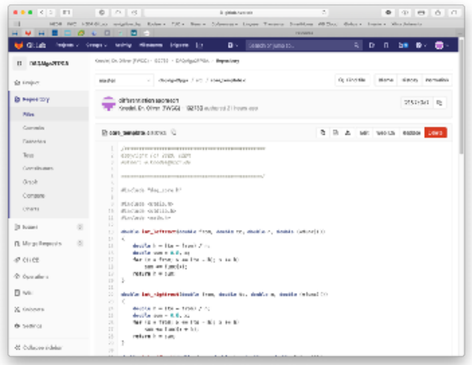
Scientific Software Development and Reproducible Workflows



Workflow Engine

Version Control

Compute (HPC, OpenStack)



GitLab

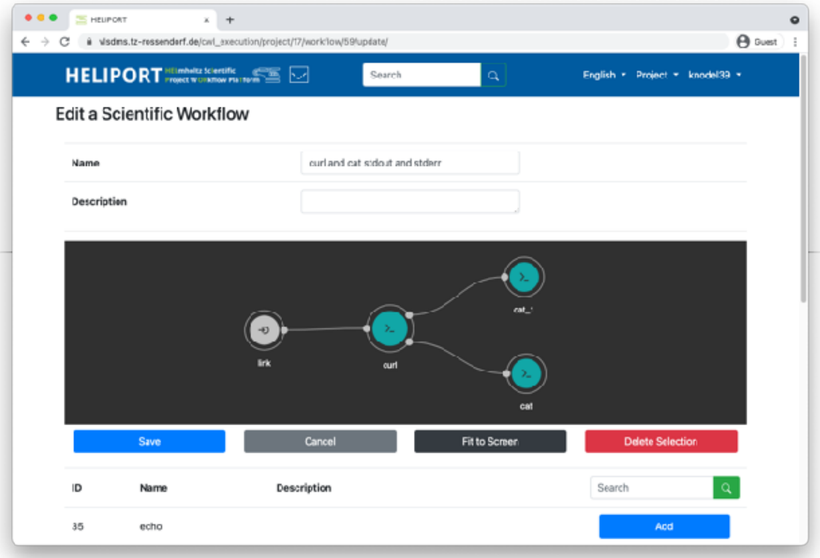
slurm workload manager

UNICORE

— Analysis and Pre-/Postprocessing steps needs to be:
  - Documented and
  - Reproducible

COMMON WORKFLOW LANGUAGE

— Capsuling every step in a workflow adapts the **FAIR** principles.



HELIPORT

JÜLICH FORSCHUNGSZENTRUM

HI JENA Helmholtz Institute Jena

HZDR

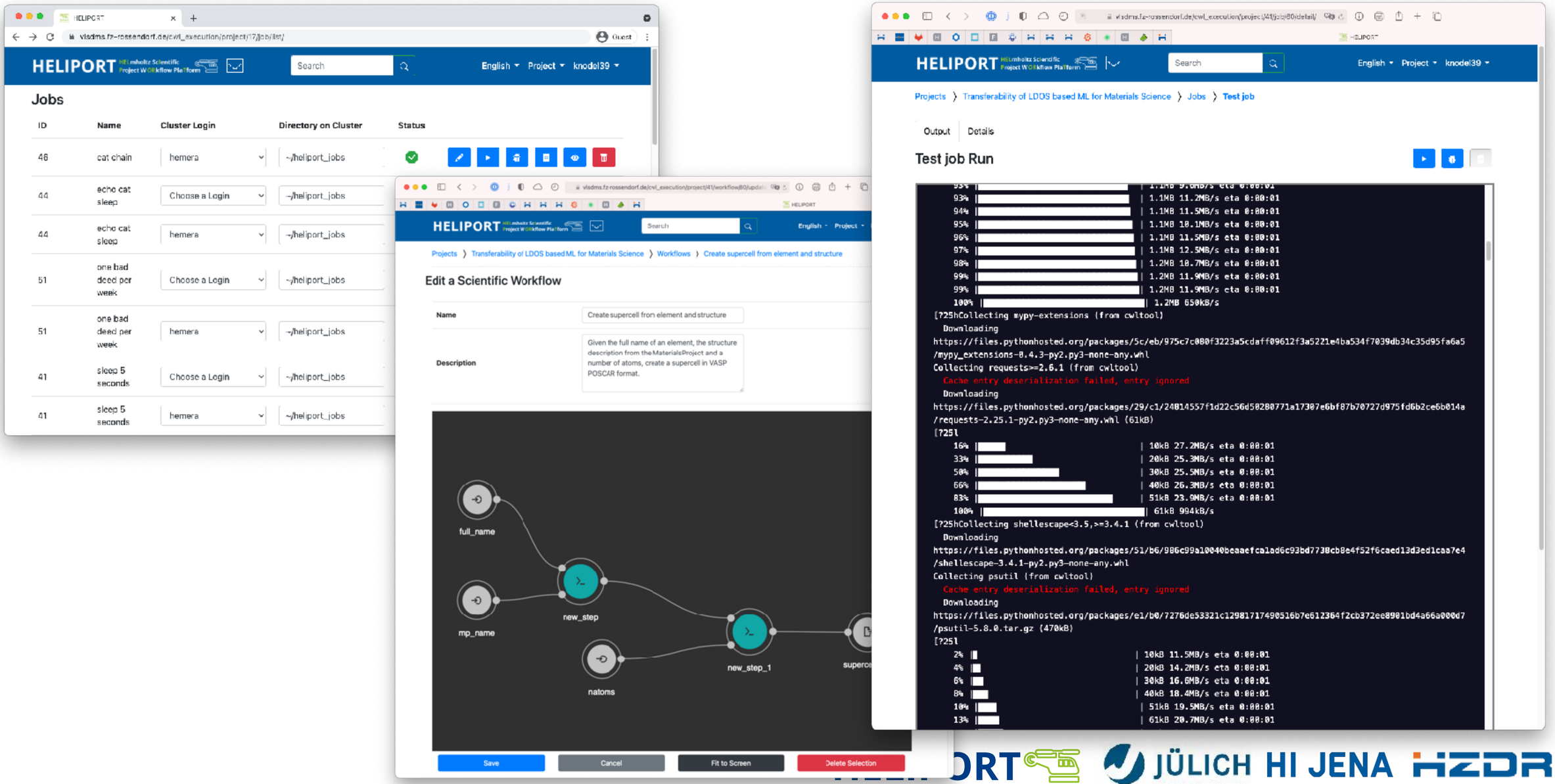
Scientific Workflows are Managed in the back-end by UNICORE

— We need access to our computing infrastructure at HZDR, the Jülich Supercomputing Center (JSC) and also at ZIH for our users from the Center for Advanced Systems Understanding (CASUS).

— UNICORE makes distributed computing and data resources available in a seamless and secure way in intranets and the internet.

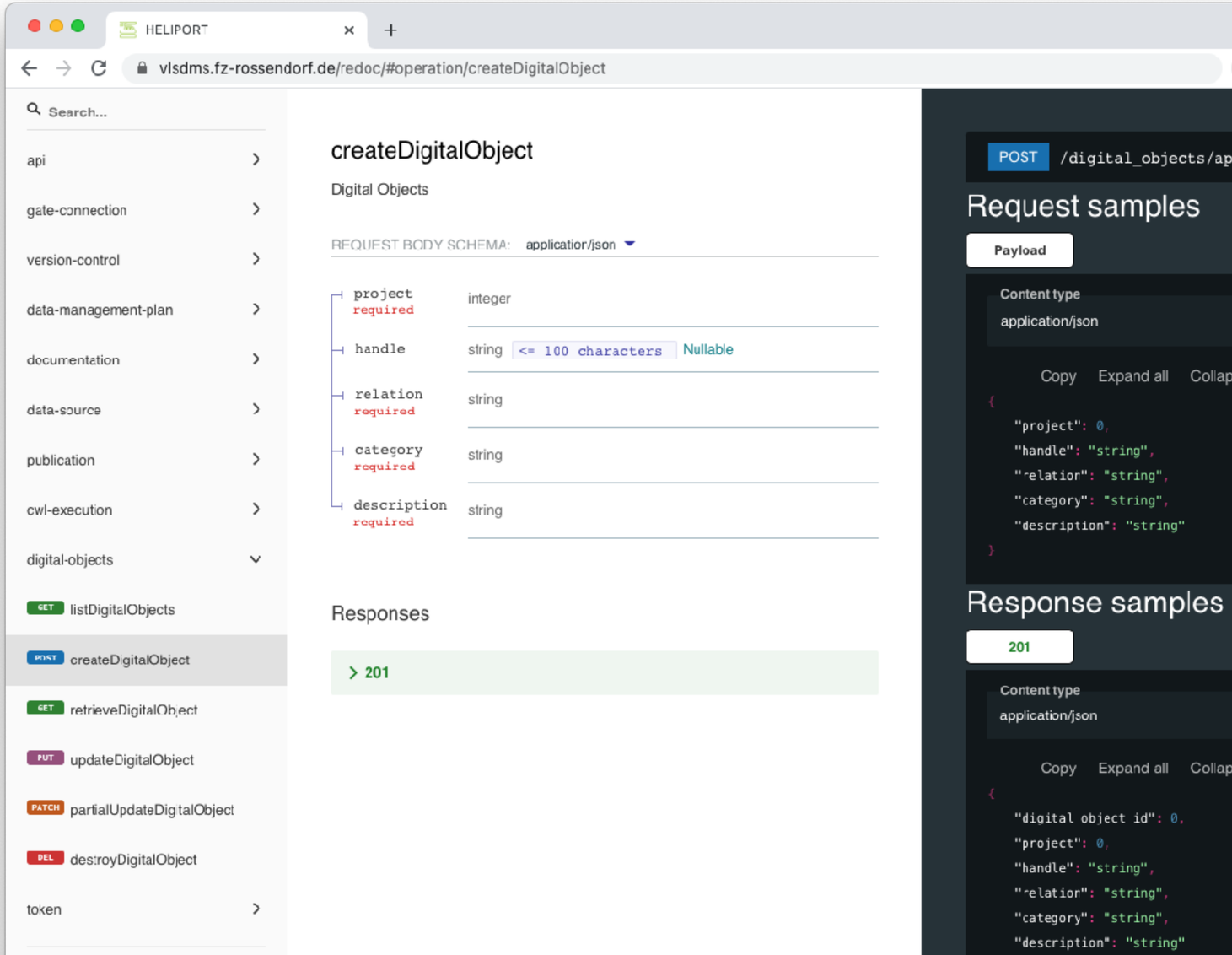
— We integrated UNICORE in our backend to decouple HELIPORT from the different HPC infrastructures.

Heliport Manages Workflows, their Execution and the Metadata



Heliport REST API

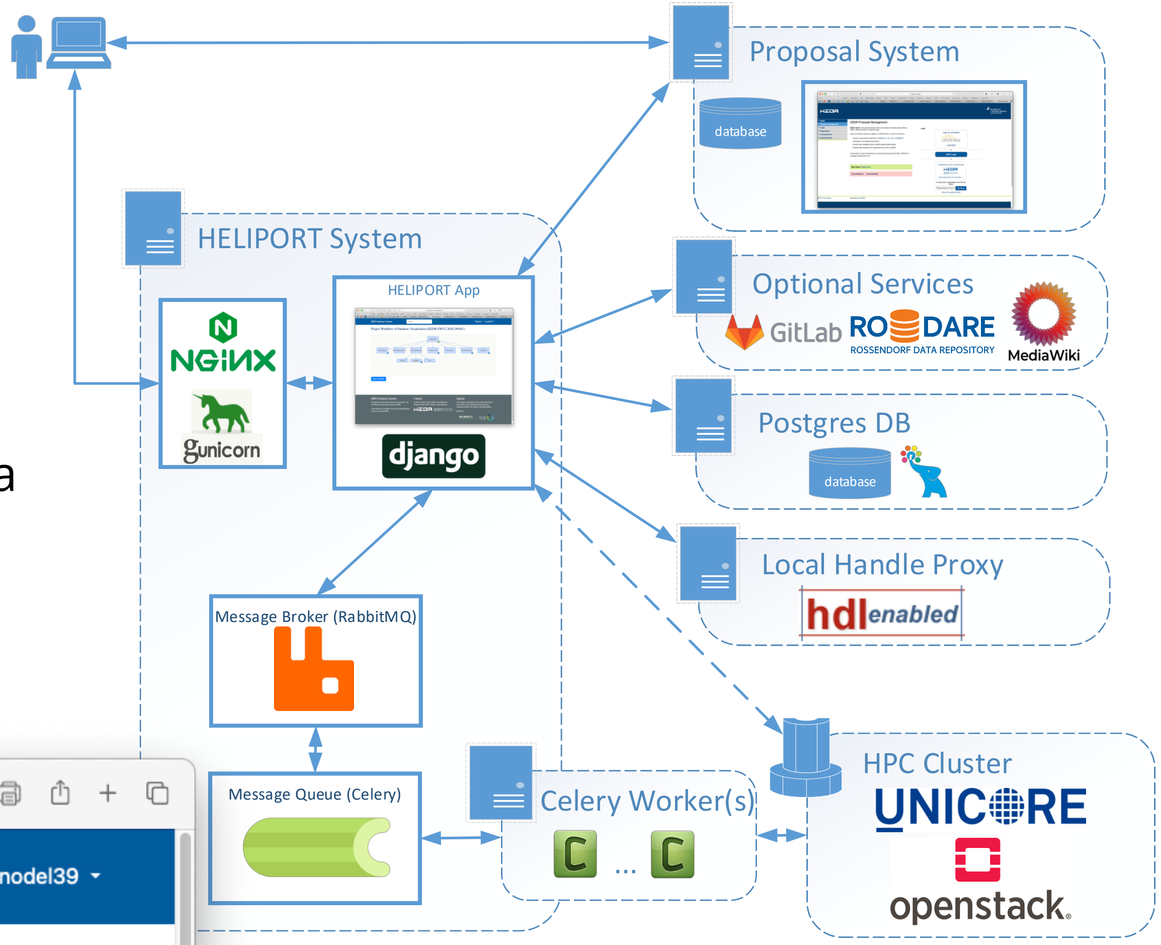
— The API provides access to our full Heliport infrastructure:
  - Proposal access (GATE),
  - Handle management,
  - CWL execution and monitoring,
  - Project metadata export,
  - Digital Object and
  - Lifecycle management.
— API documentation (ReDOC) available.
— Essential to integrate the Heliport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

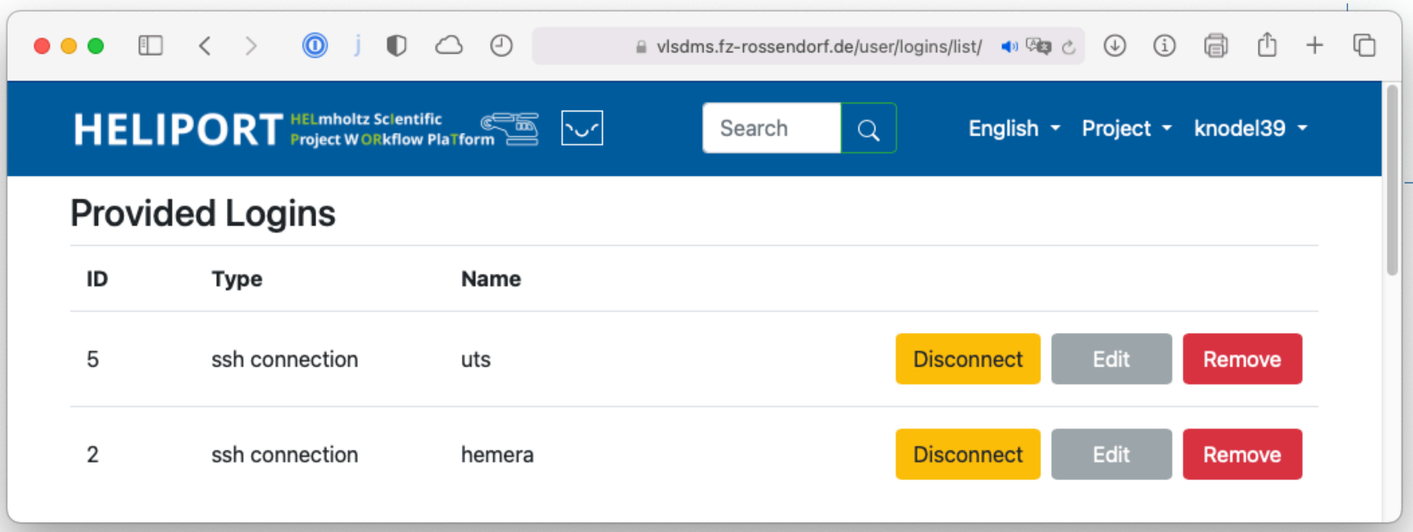


Heliport System Infrastructure

— The Heliport 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 a metadata scheme based on DataCite.

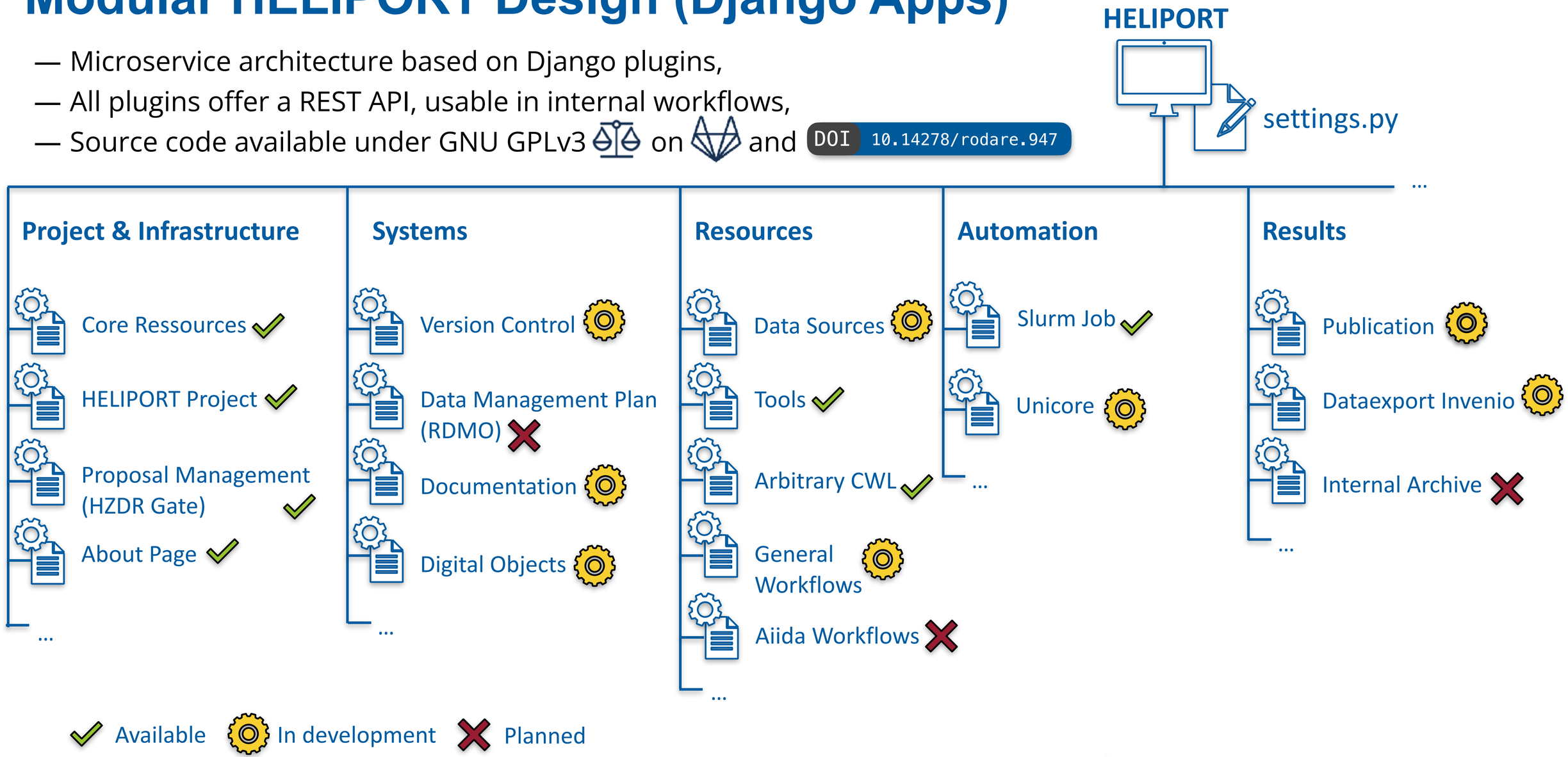
— The CWL workflows are managed in Heliport, but executed on our cluster using UNICORE.





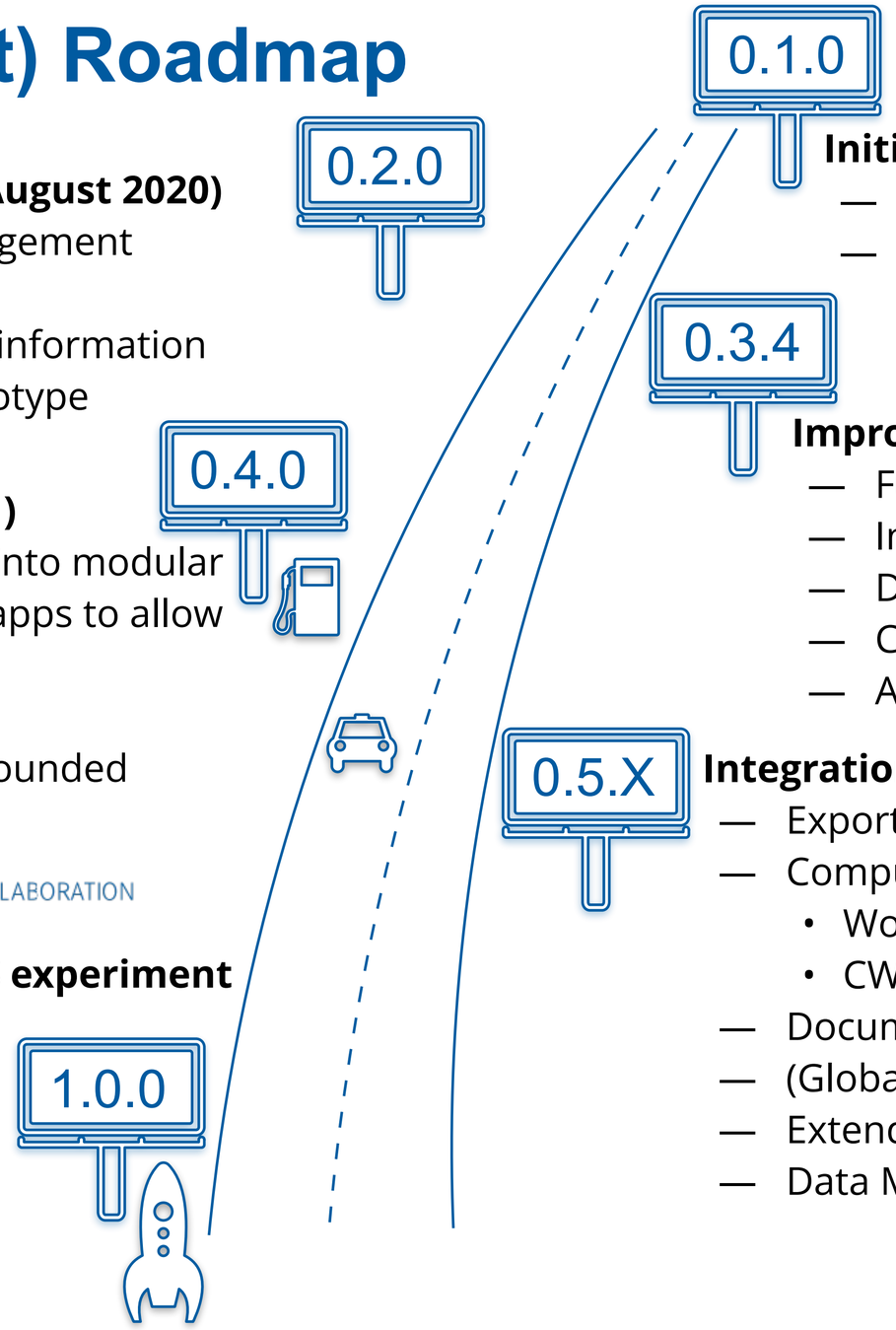
Modular HELIPORT Design (Django Apps)

— 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

Heliport (Project) Roadmap



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

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)
- Fully configurable stages and modules
- Infrastructure and database updates
- Daily proposal database update
- CI pipeline for test and deployment
- Advanced logging and monitoring

Modular Structure (July 2021)
- Subdivision of the stages into modular and configurable Django apps to allow individual extensions
- Refactoring of the project
- Official start of the HMC founded Heliport project:

HMC | HELMHOLTZ METADATA COLLABORATION

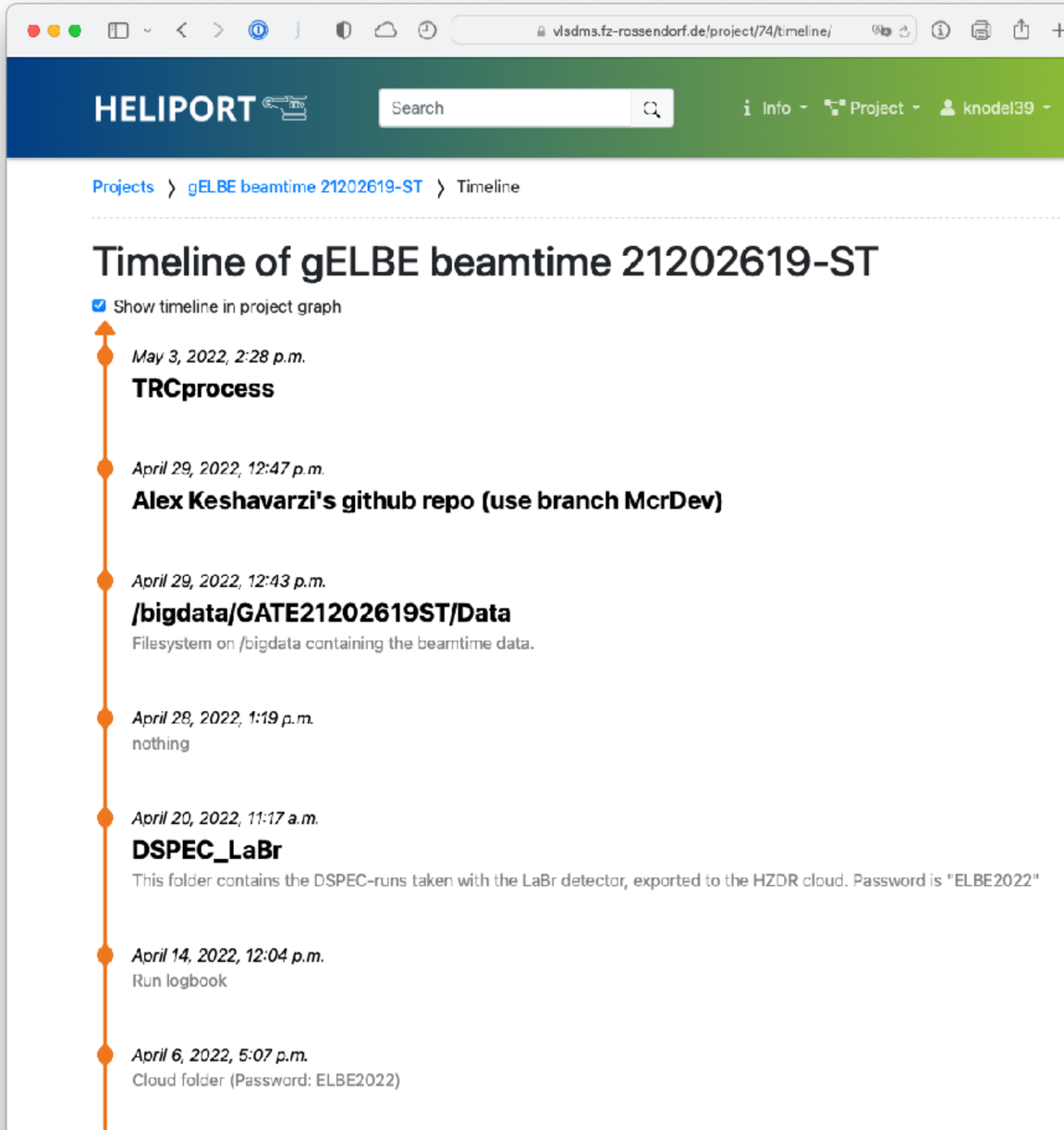
Integration of various Apps and Features
- Export for (different) Metadata Schemas
- Computational/Scientific workflow execution
  - Workflow management and monitoring
  - CWL support
- Documentation using GitLab pages
- (Global) **Handle management**
- Extended **Support for Digital Twins**
- Data Management Plan Export for RDMO

Documentation TELBE and POLARIS experiment
- Integration of all related data sources
- Automated workflow initiation
- Publication of all data products

DRESDEN concept HZDR

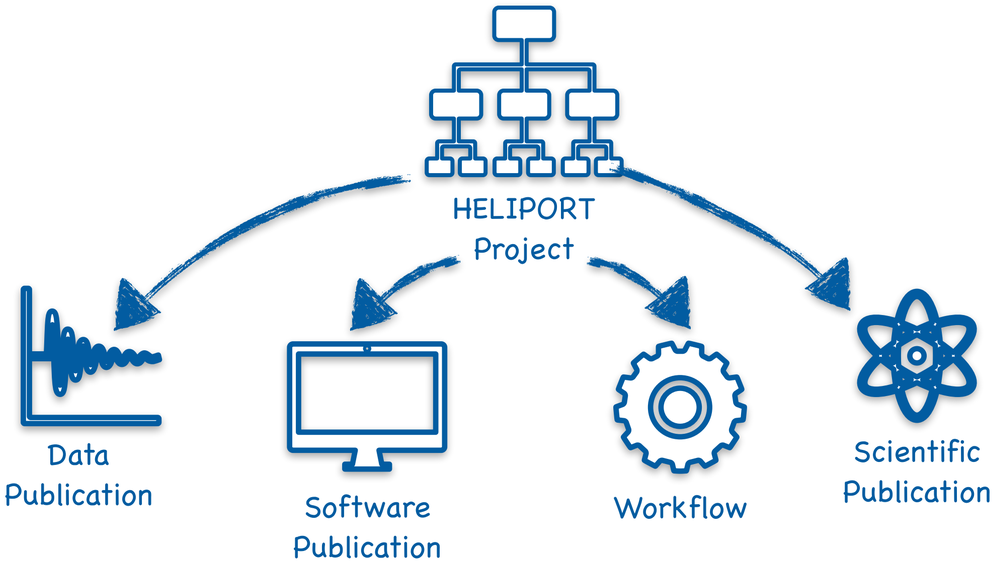
Upcoming Extensions: Data provenance and Comprehensibility

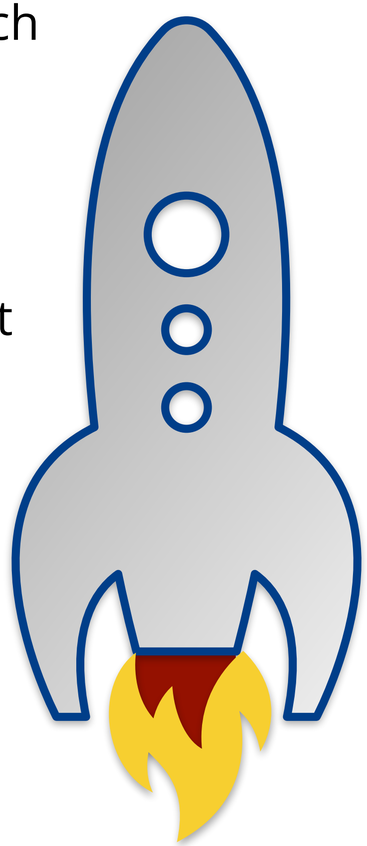
- For many systems and services we still have to develop necessary plug-ins for the integration into Heliport.
- The versioning of an experiment lifecycle is unavoidable and we are still discussing how we can present the feature in our web frontend:
  - A Git project with all metadata to restore a lifecycle,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



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.

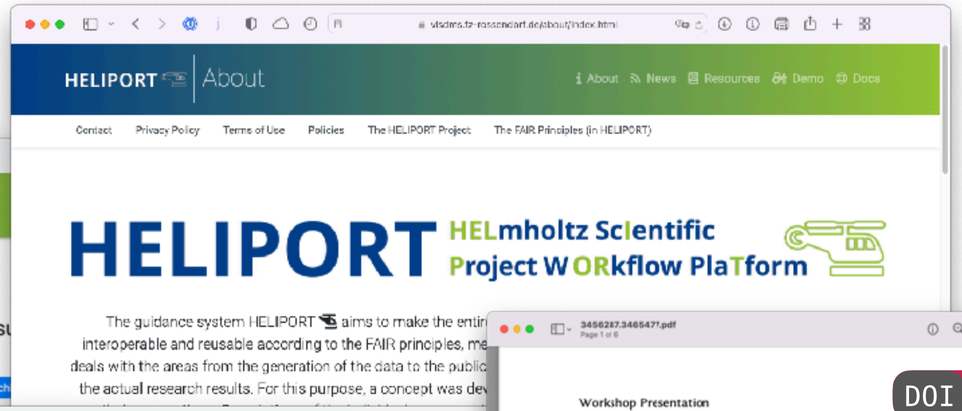


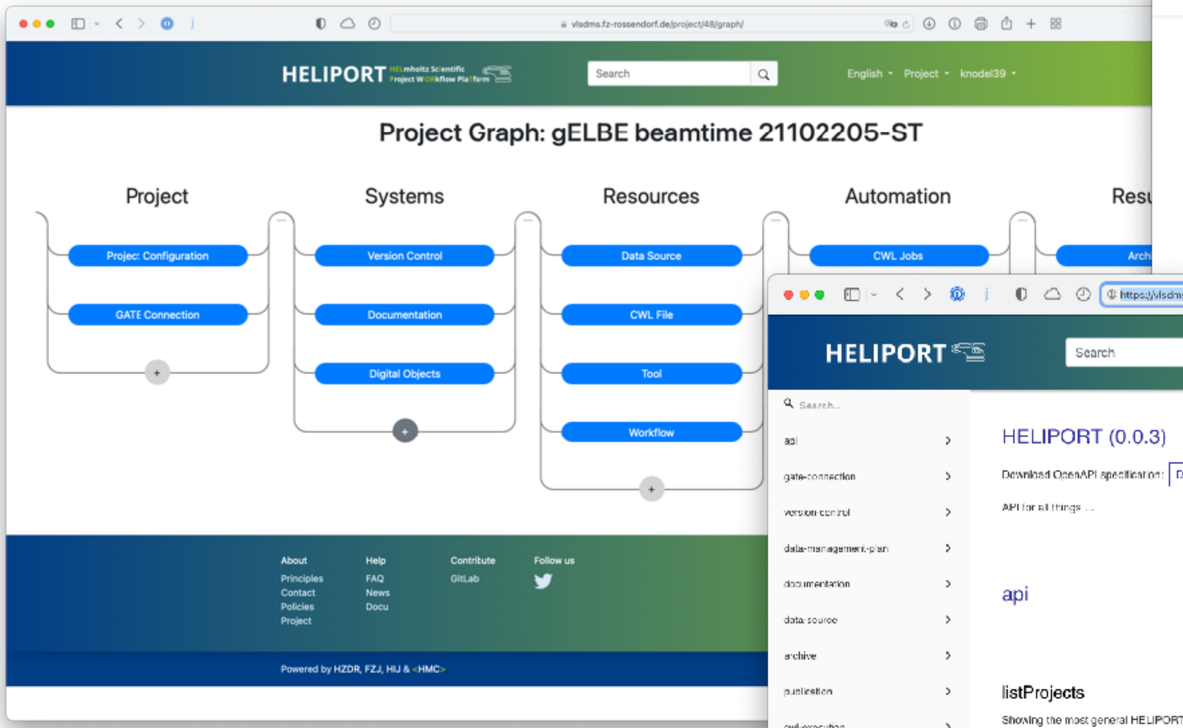


HELIPORT JÜLICH FORSCHUNGSZENTRUM HI JENA Helmholtz Institute Jena HZDR

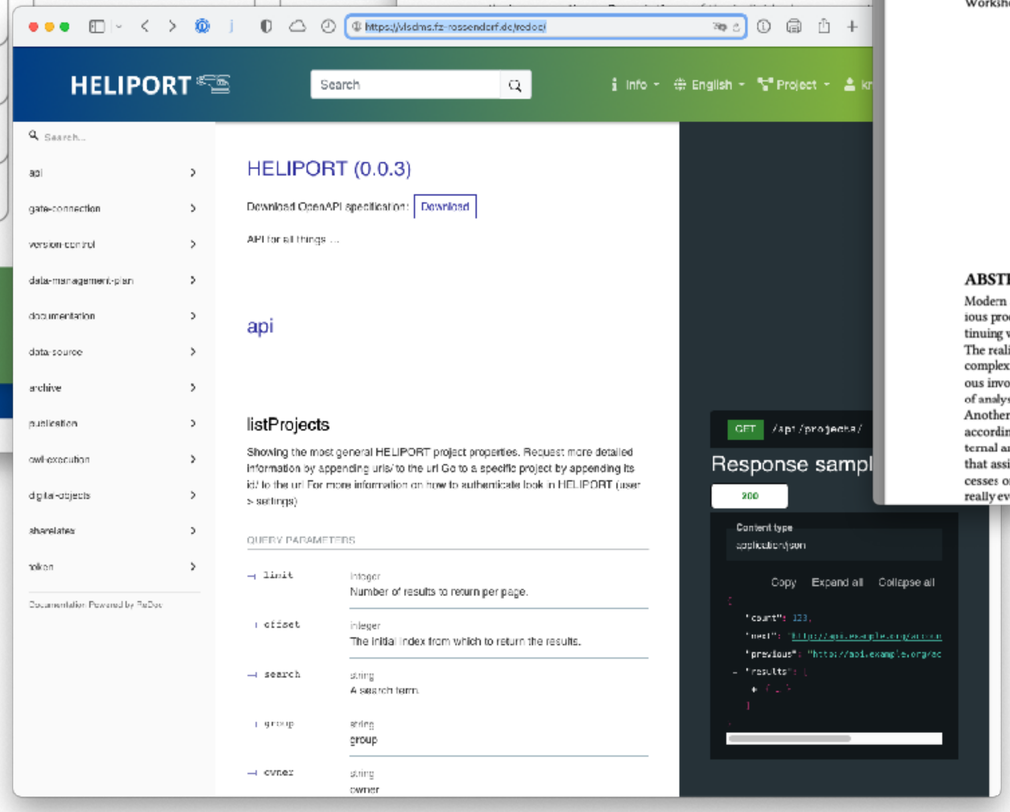
Resources

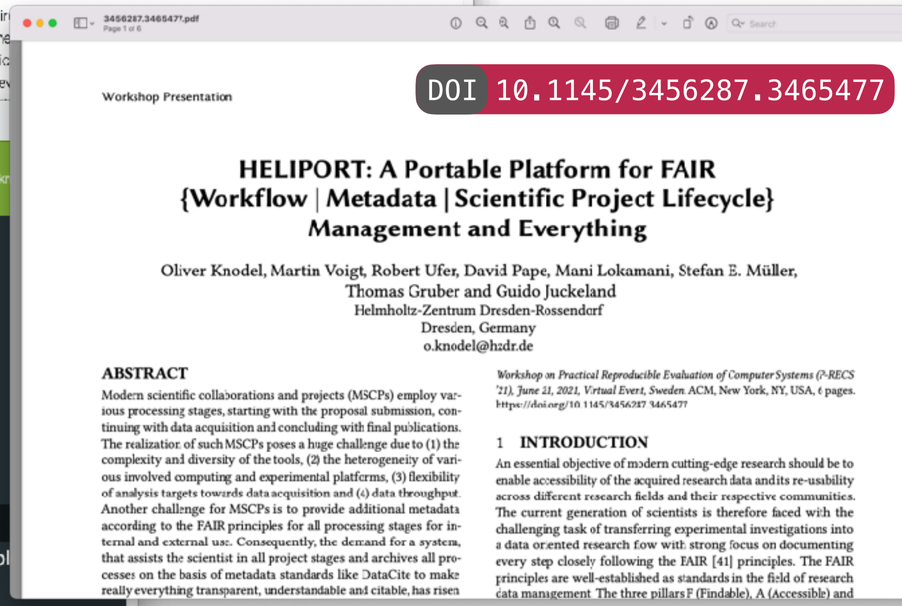
Website: heliport.hzdr.de





Demo: heliport.hzdr.de/app





API Doc: heliport.hzdr.de/redoc/