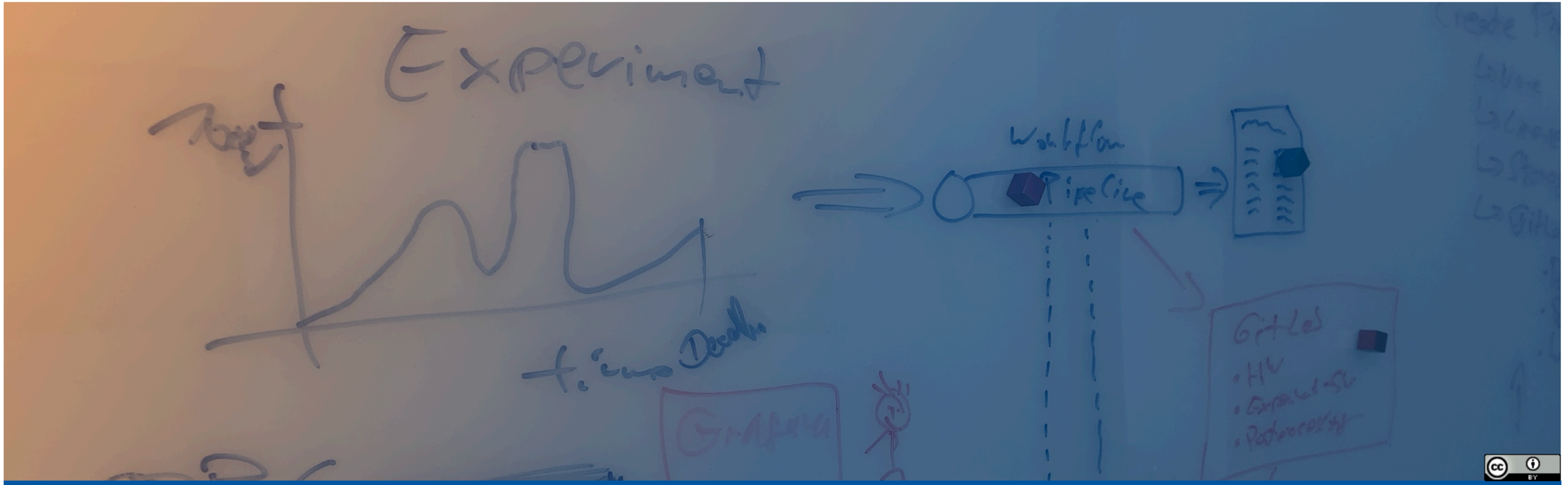
HELMHOLTZ ZENTRUM
DRESDEN ROSSENDORF



Overarching Data Management Ecosystem HELIPORT

From Small Experiments to Large-Scale Research Facilities

1st Workshop on Open Science at GSI/FAIR, 19. October 2023

Oliver Knodel, Thomas Gruber, Martin Voigt, Jeffrey Kelling, Mani Lokamani, Stefan E. Müller, David Pape 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. 1,470. Thereof 670 scientists.

— HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

Our Research Facility and our Large Scale Research Infrastructures

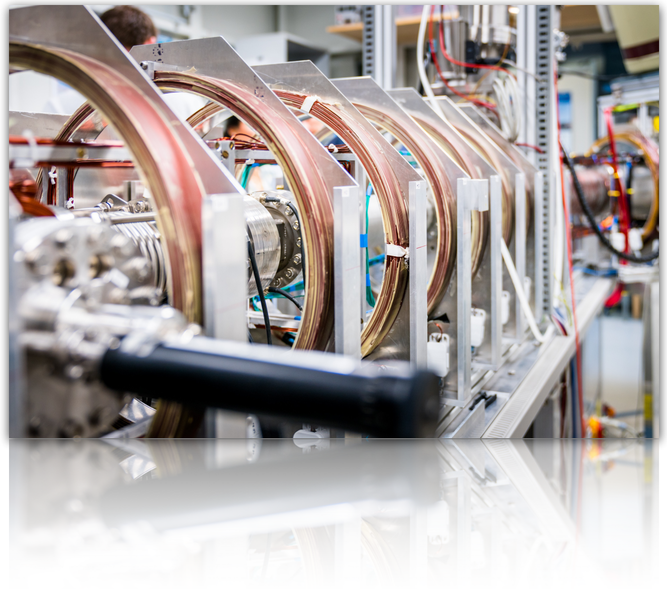
The Helmholtz-Zentrum Dresden - Rossendorf

— Employees approx. 1,470. Thereof 670 scientists.

— HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

Research Fields

— Energy, Health and Matter.



HZDR

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

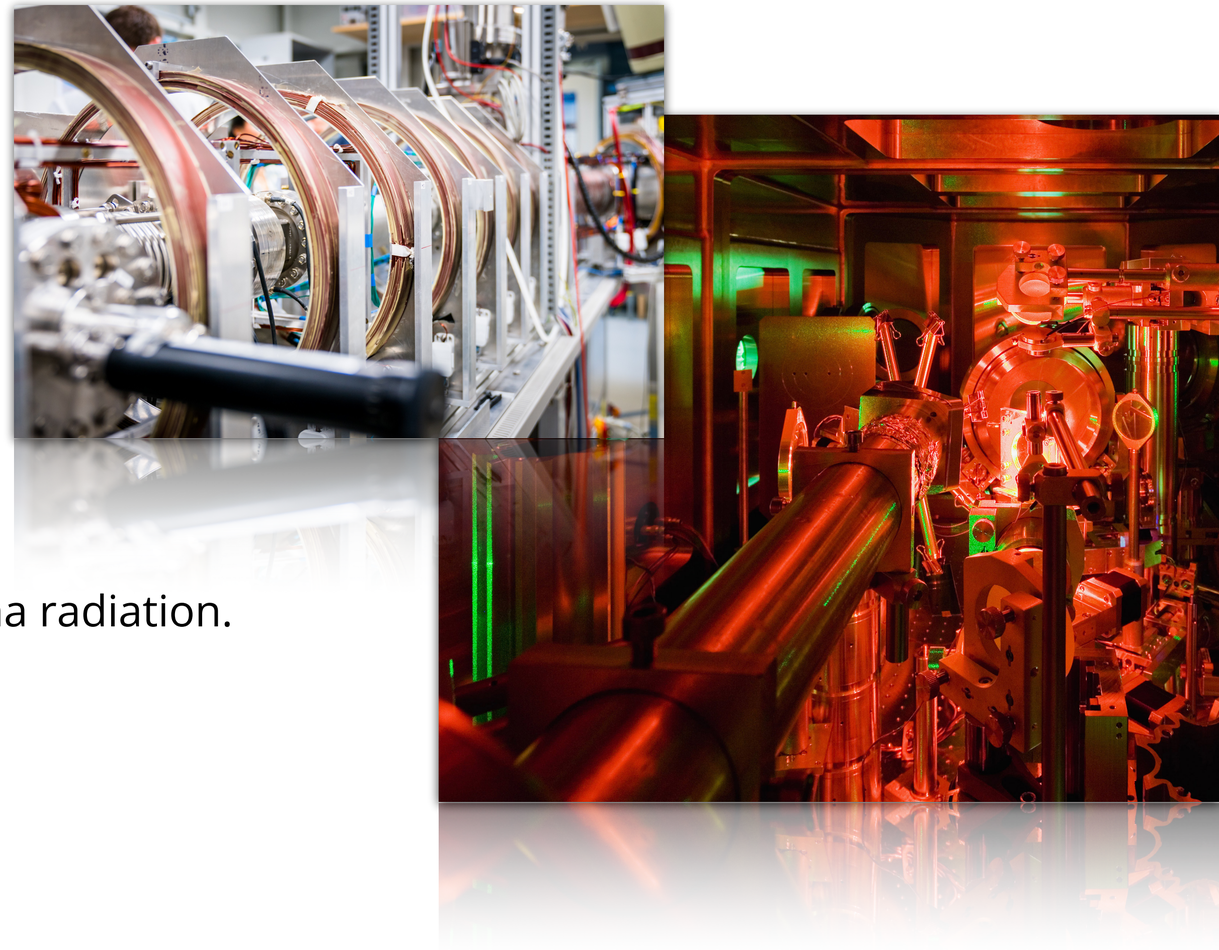
Research Fields

— 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.



HZDR

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.

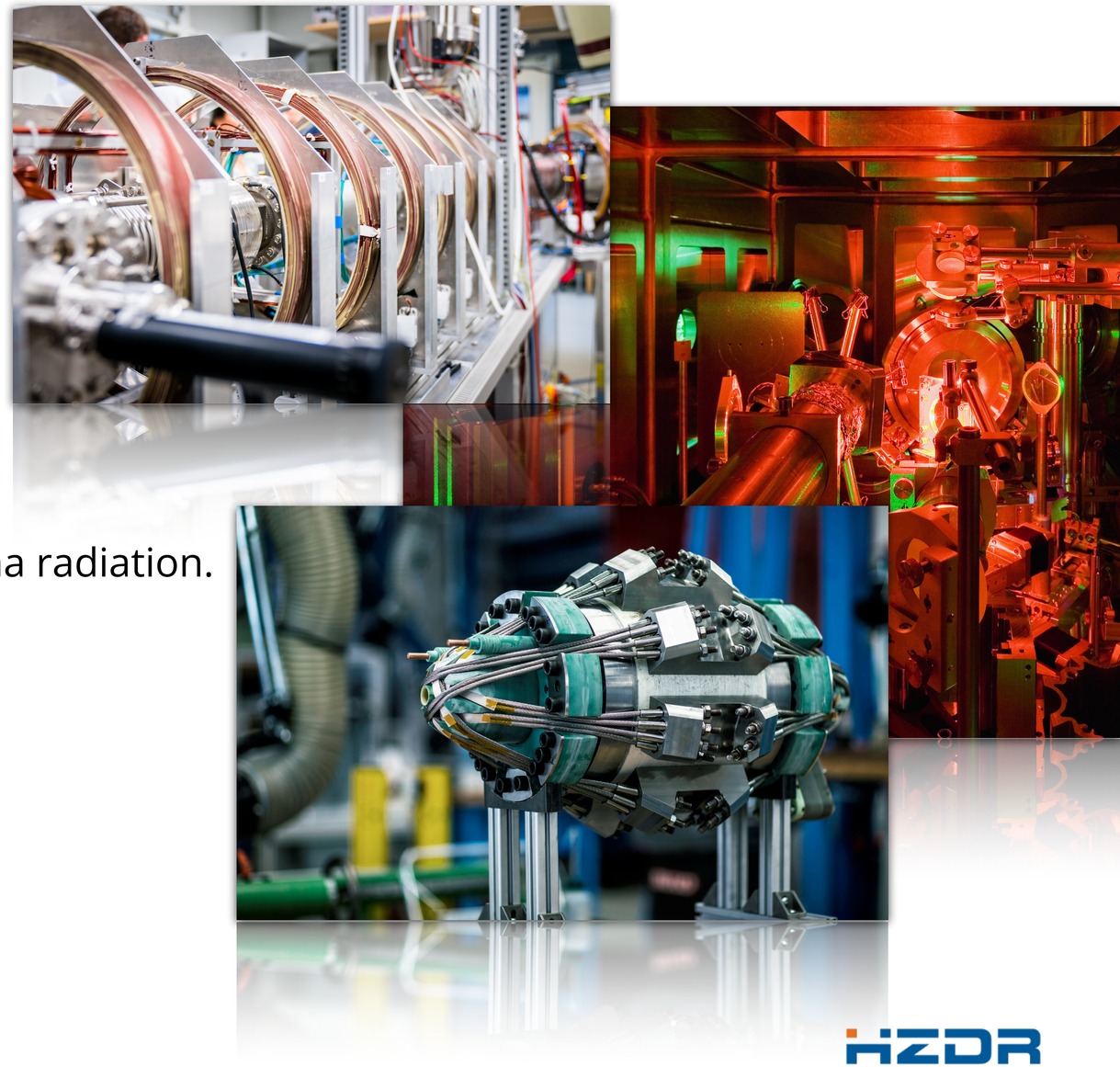
ELBE – Center for High-Power Radiation Sources

— 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.



HZDR

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.

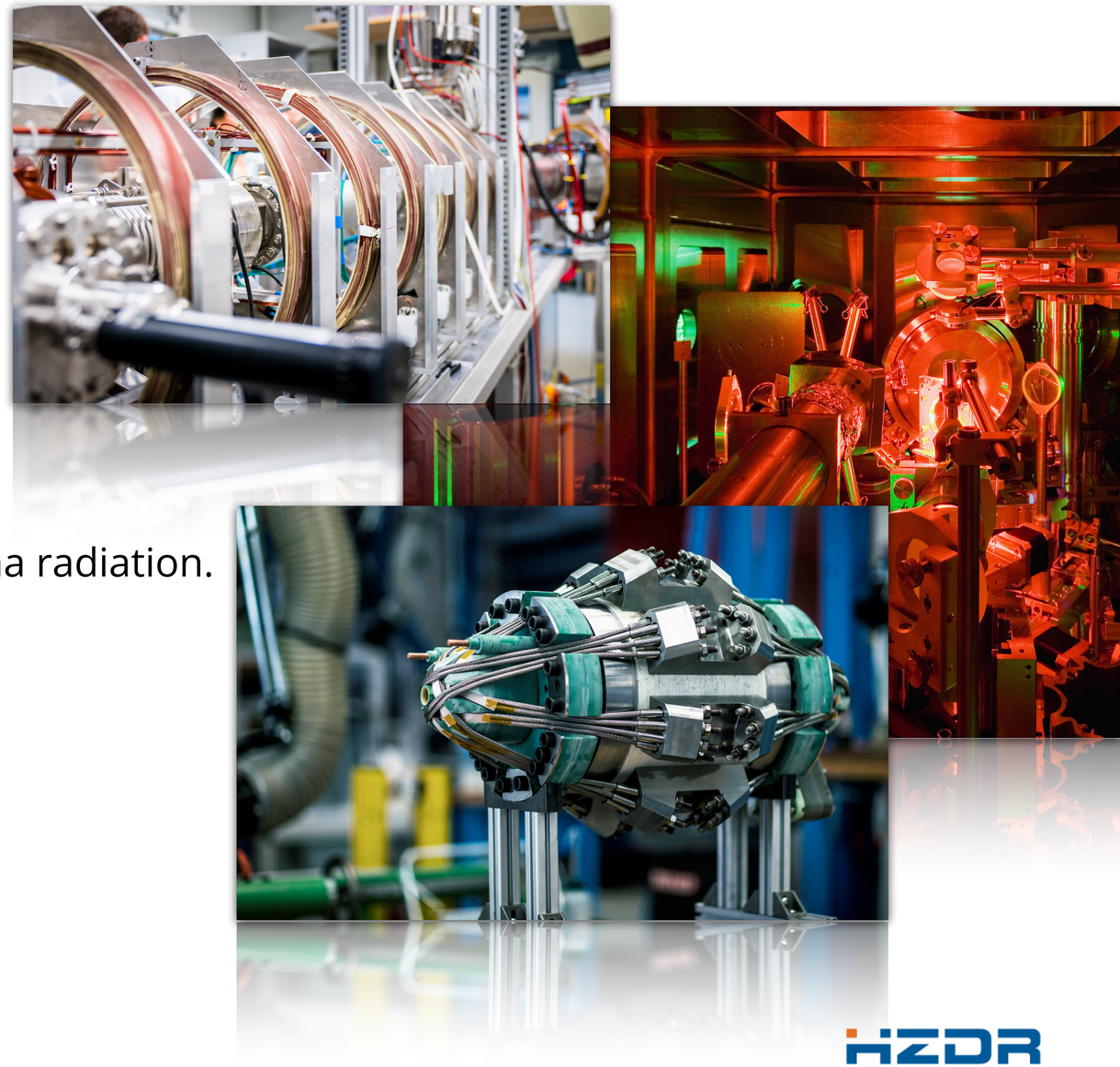
— Positrons, protons, neutrons as well as X-ray and gamma radiation.

Dresden High Magnetic Field Laboratory (HLD)

— Europe’s highest pulsed magnetic fields.

Ion Beam Center (IBC)

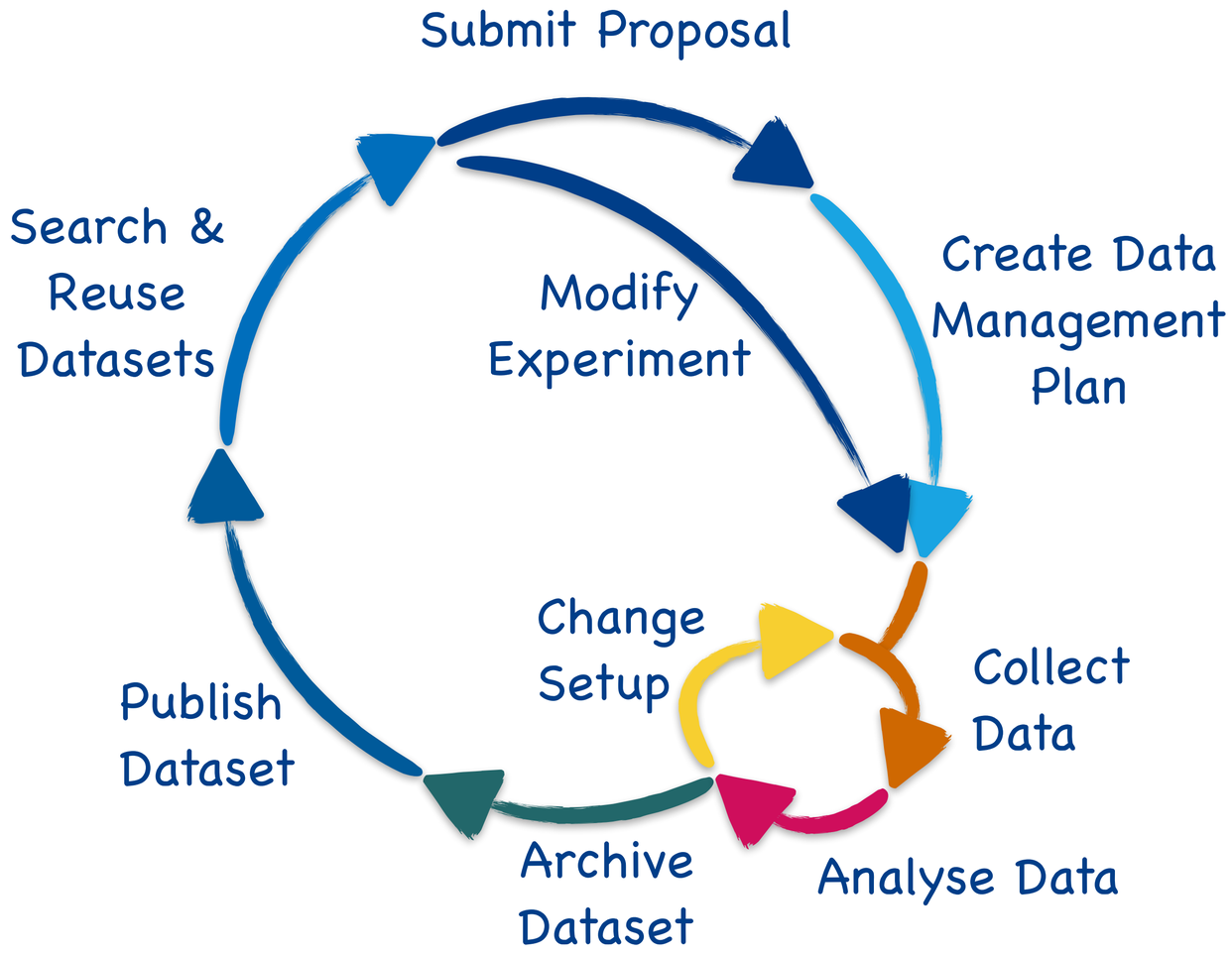
— Nanoscale surface analysis and modification.



HZDR

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

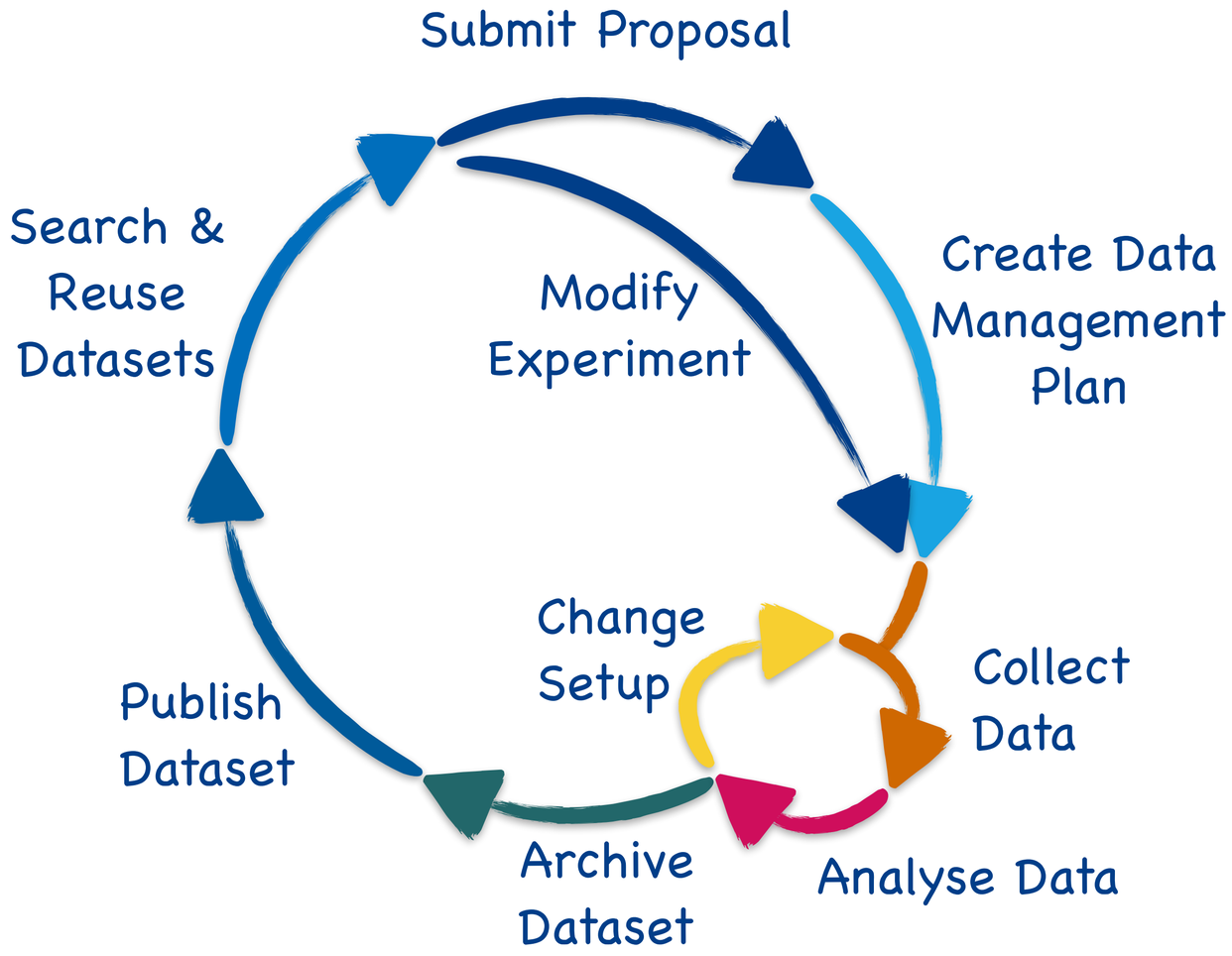
— We support many steps of our different research experiment (matter, energy and health) with tools:
  - electronic lab books,
  - interactive analysis,
  - publication of datasets,
  - scientific workflow management,
  - Handle generation and management.



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.

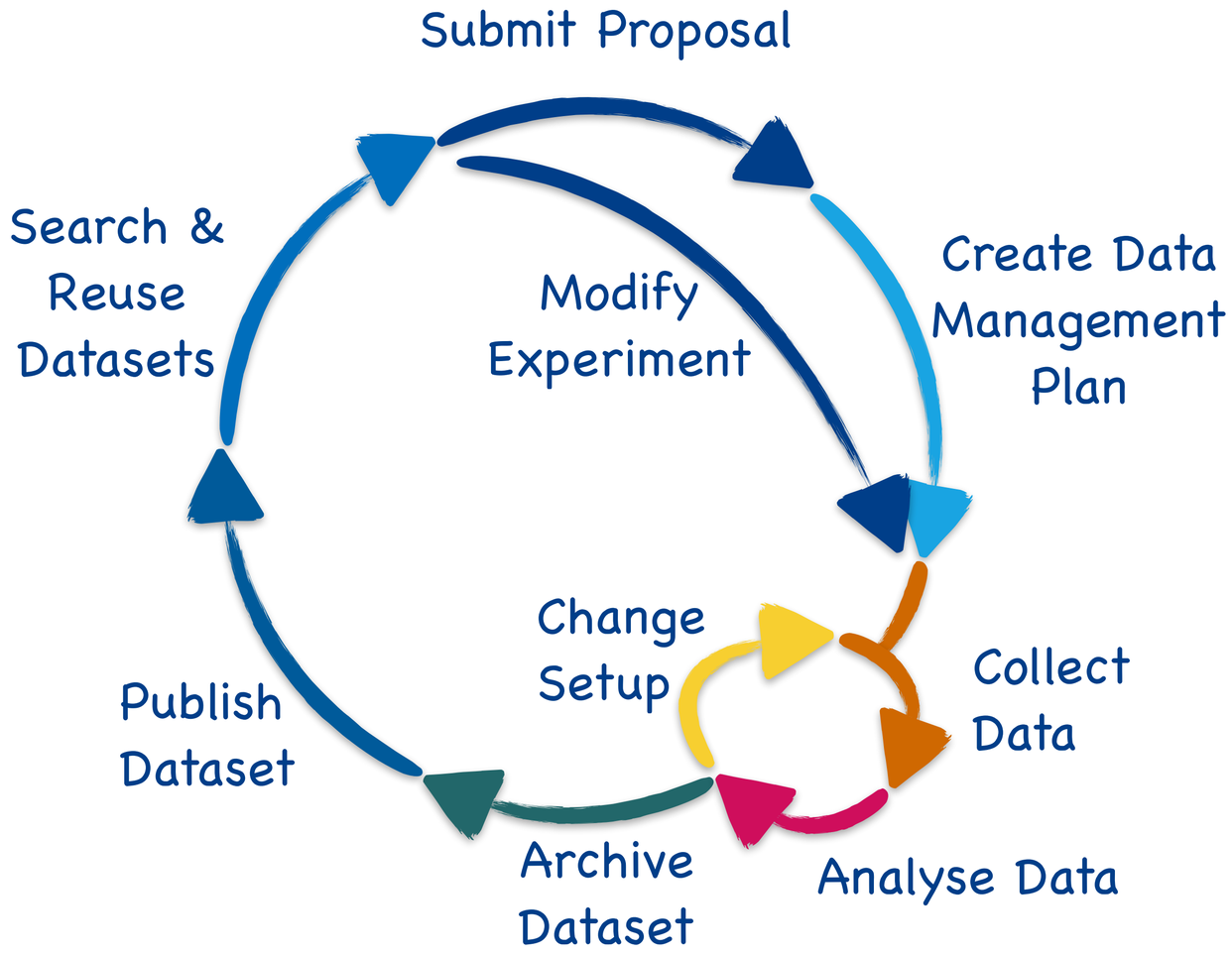


Our Challenge: An End-to-End Digital Data Lifecycle

— 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.

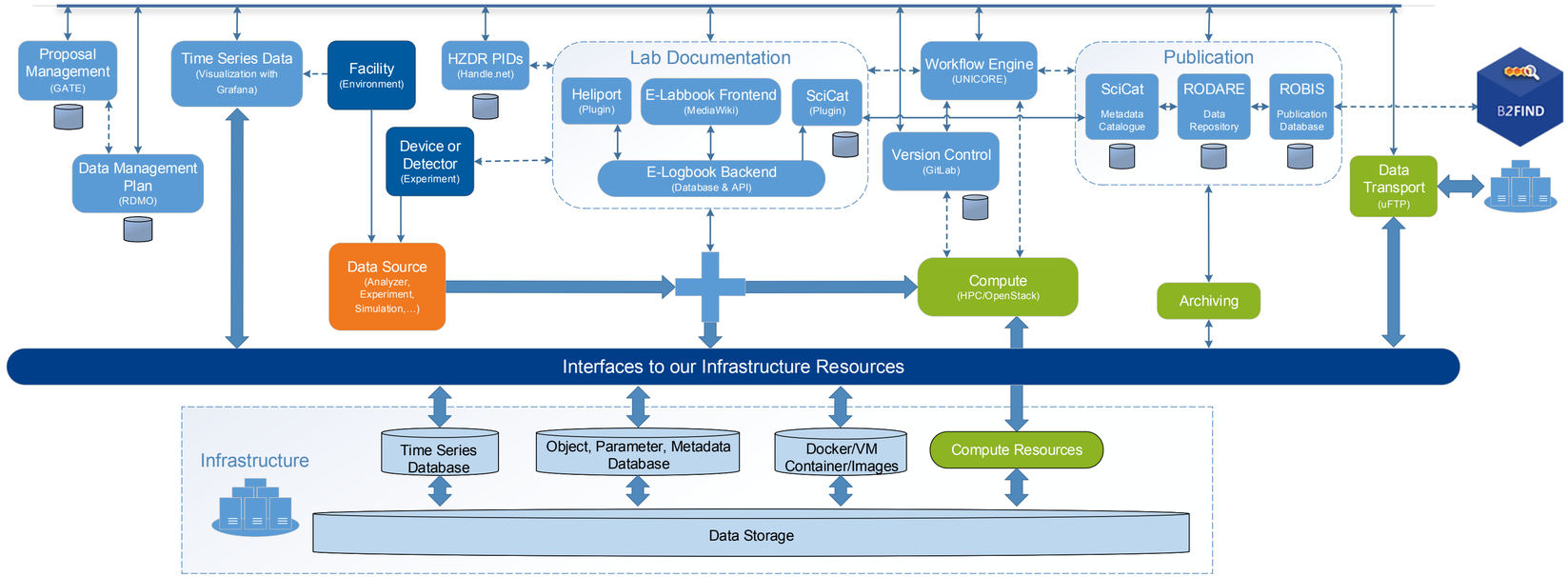
— A uniform and smooth access to and between all services and systems in our ecosystem is necessary.

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



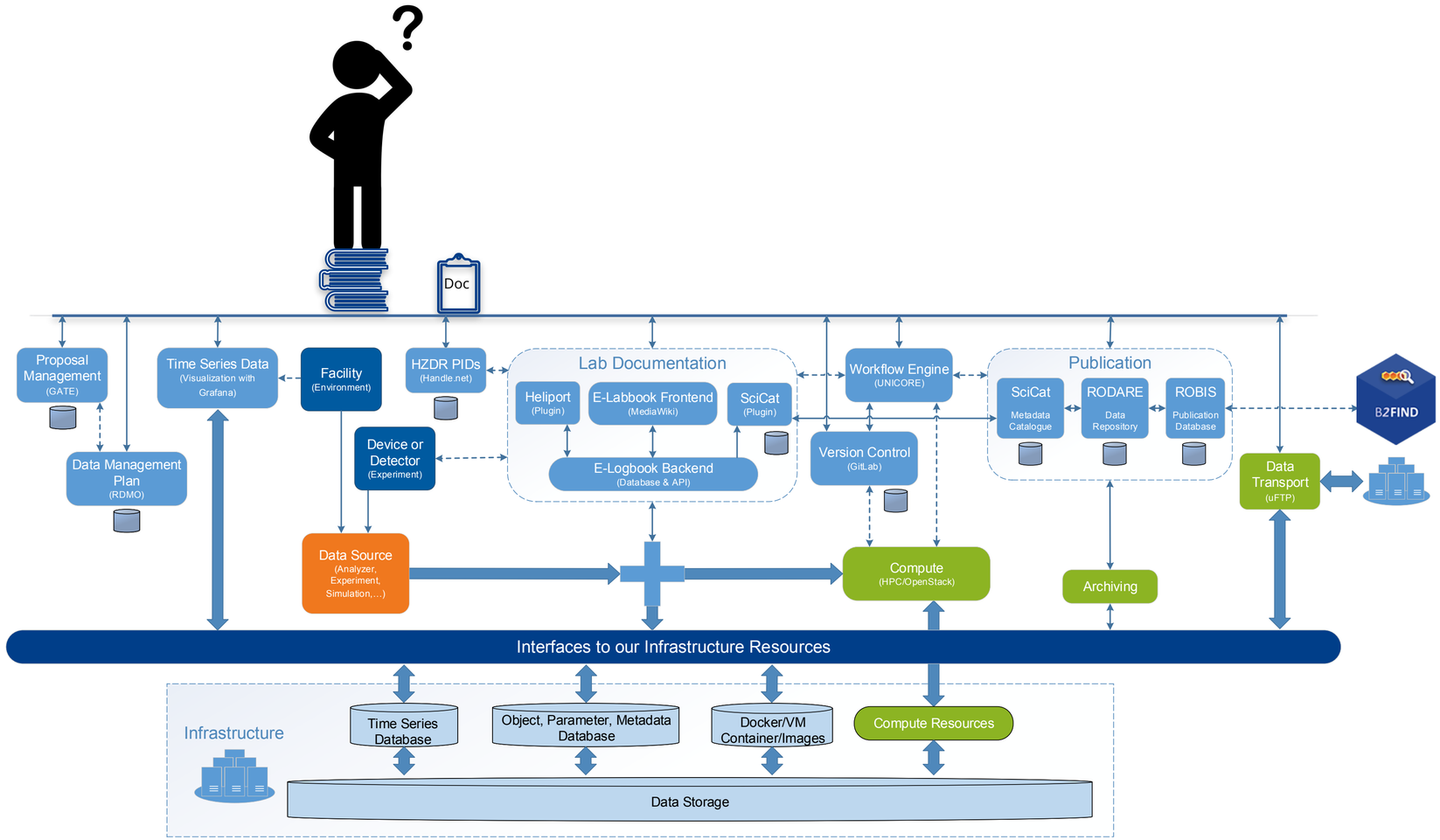
Our Observations and Experiences

— Our IT infrastructures can support various experiments, but they are complex…



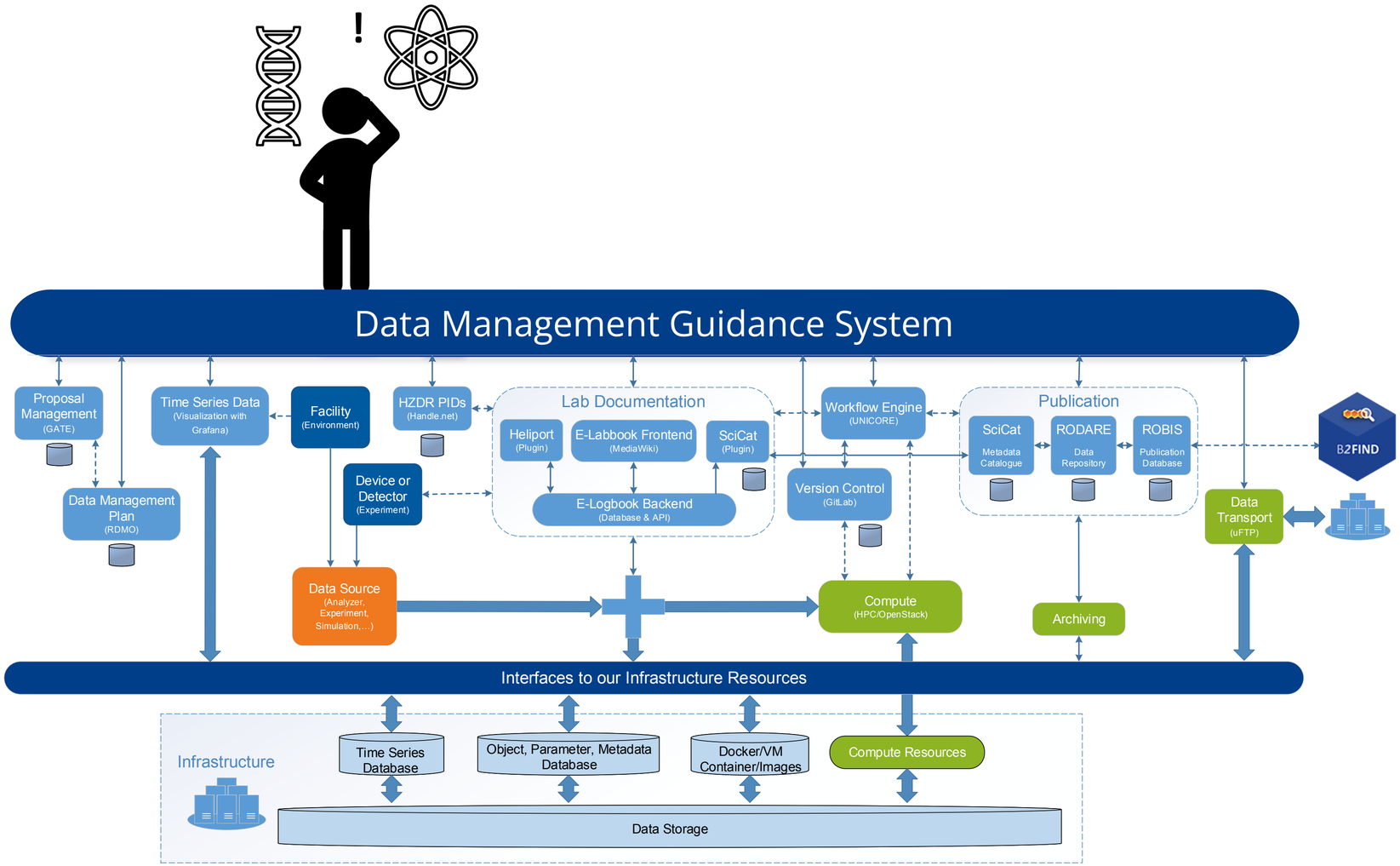
Our Observations and Experiences

- 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.



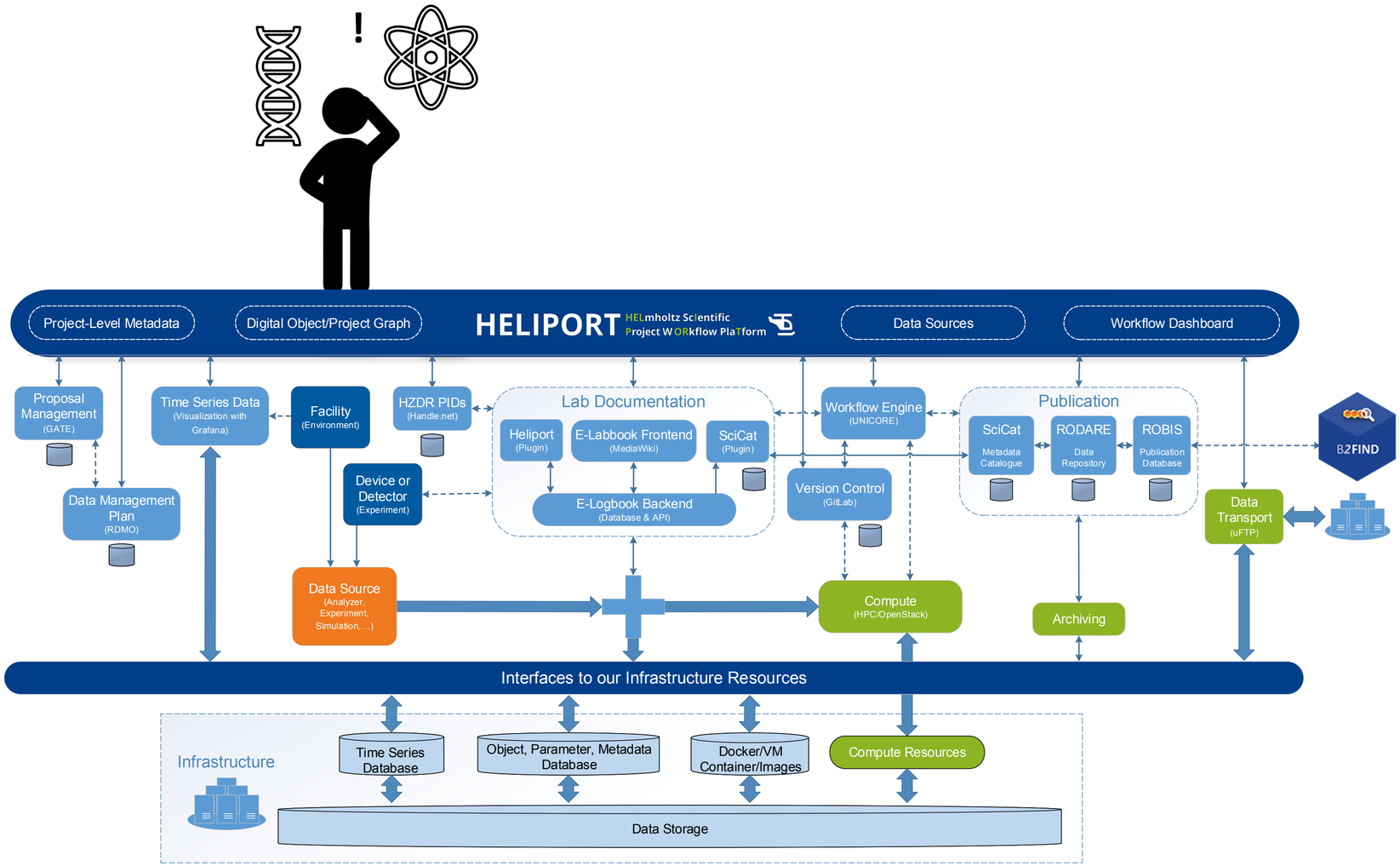
Our Observations and Experiences

- 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.



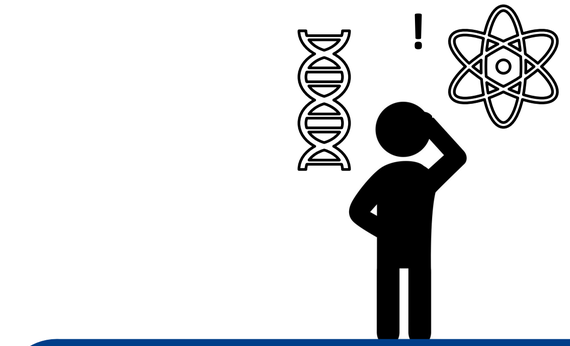
Our Observations and Experiences

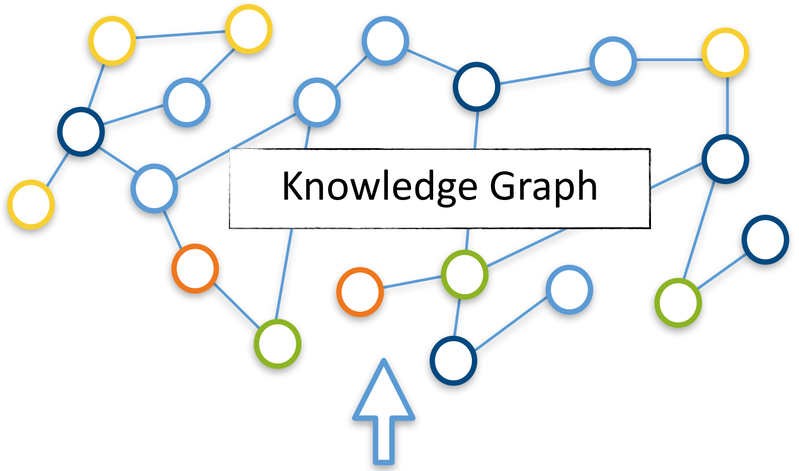
- Our IT infrastructures can support various experiments, but they are complex…
- Scientists often don’t know which services are available and how to use them.
- An overarching system guiding our scientists (and visitors) through the lifecycle of their research project is essential.

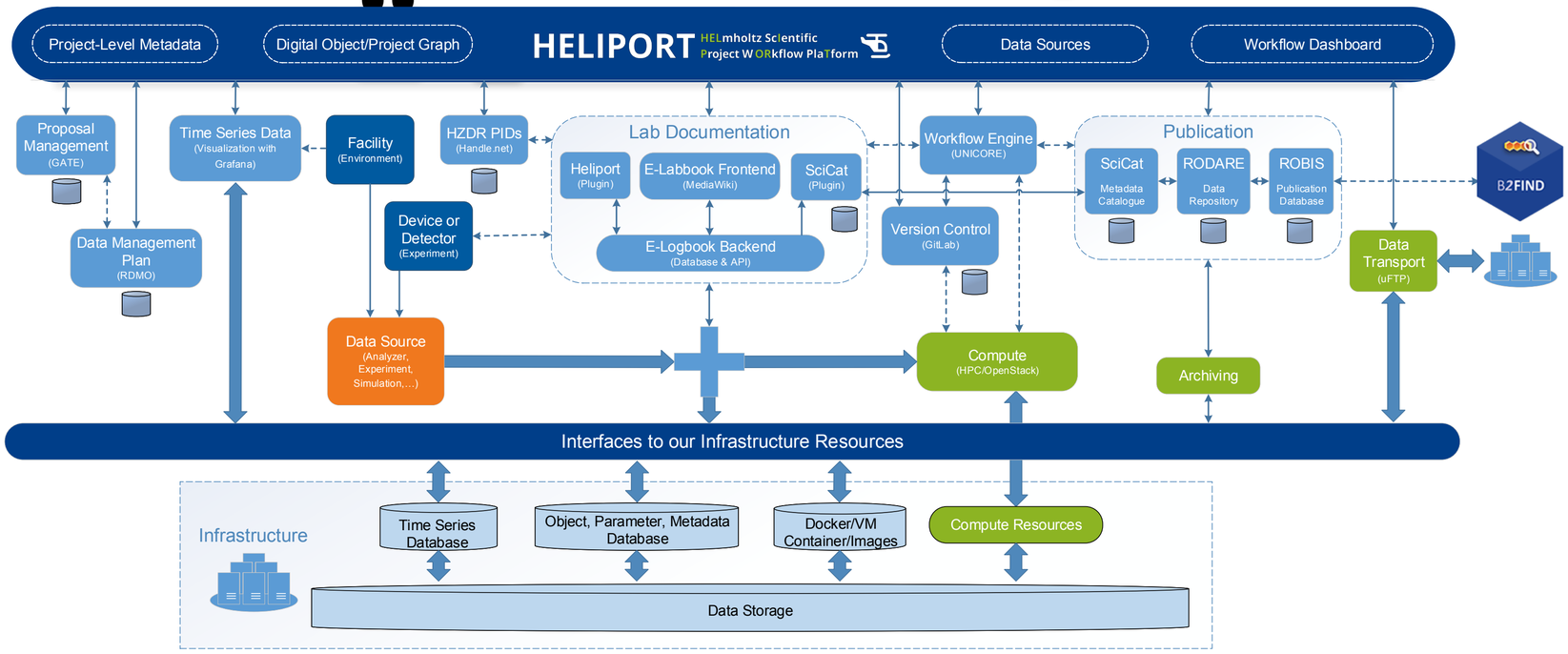


Our Observations and Experiences

- 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 next step we can provide an overall knowledge graph!







Requirements and Conditions

— HELIPORT was intended to provide only the **proposal's metadata**, from internal and external scientists, to allow the assignment of resources.

Requirements and Conditions

— HELIPORT was 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 HELIPORT to answer the most important questions of our scientists and :

Requirements and Conditions

— HELIPORT was 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 HELIPORT to answer the most important questions of our scientists and :

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

Requirements and Conditions

— HELIPORT was 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 HELIPORT to answer the most important questions of our scientists and :

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

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

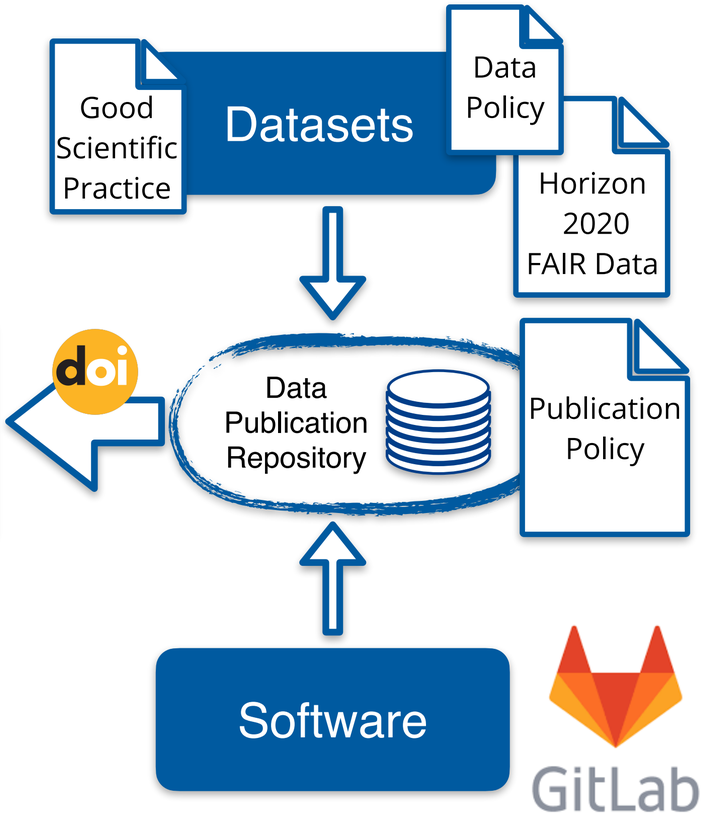
Requirements and Conditions

— HELIPORT was 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 HELIPORT to answer the most important questions of our scientists and :

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

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



Requirements and Conditions

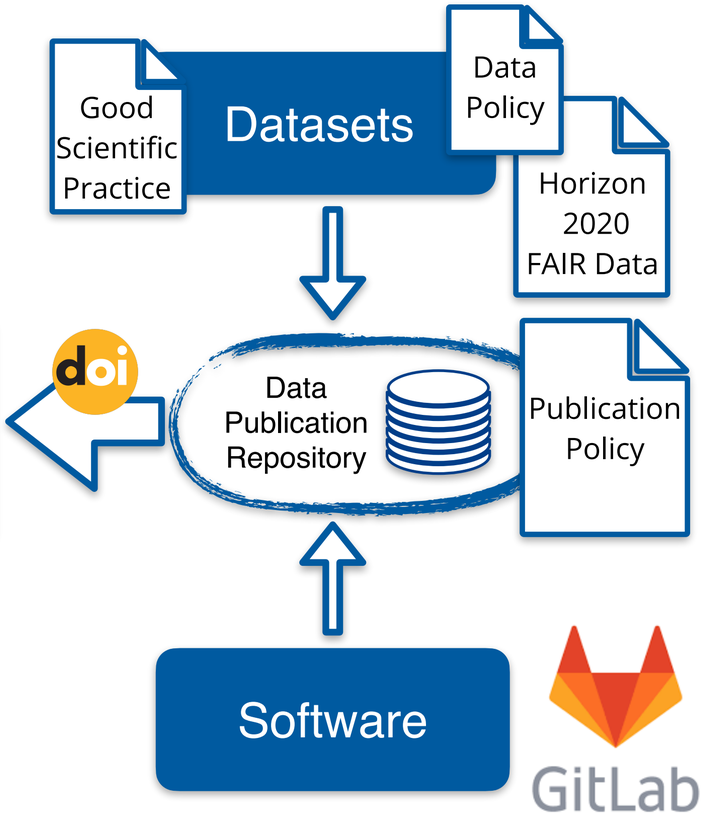
— HELIPORT was intended to provide only the **proposal's metadata**, from internal and external scientists, to allow the assignment of resources.

— Over the time we decided to use HELIPORT to answer the most important questions of our scientists and :

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

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

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



Requirements and Conditions

— HELIPORT was intended to provide only the **proposal's metadata**, from internal and external scientists, to allow the assignment of resources.

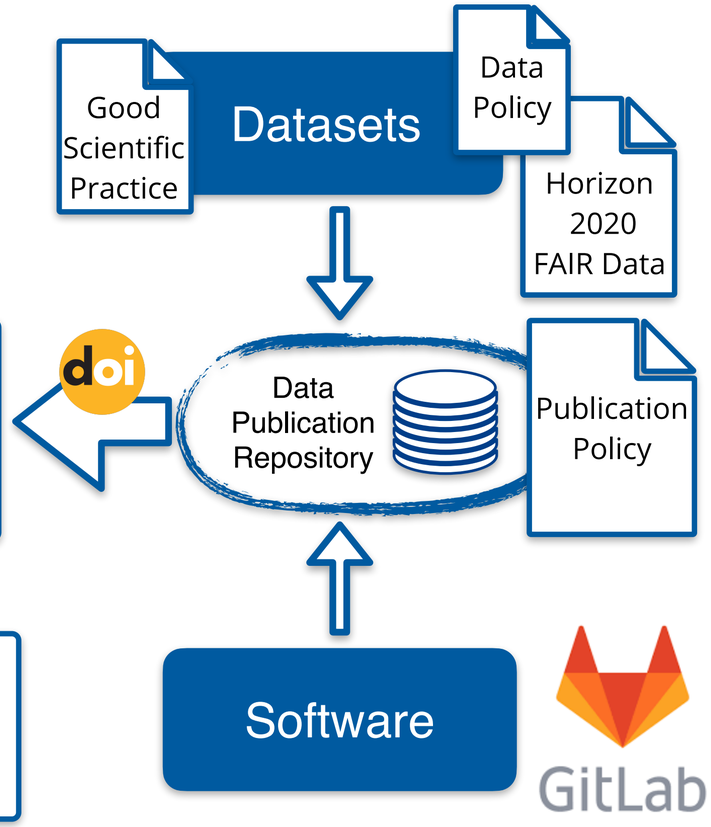
— Over the time we decided to use HELIPORT to answer the most important questions of our scientists and :

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 **visiting/external scientists** into our project lifecycle and all associated tools?

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



Requirements and Conditions

— HELIPORT was 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 HELIPORT to answer the most important questions of our scientists and :

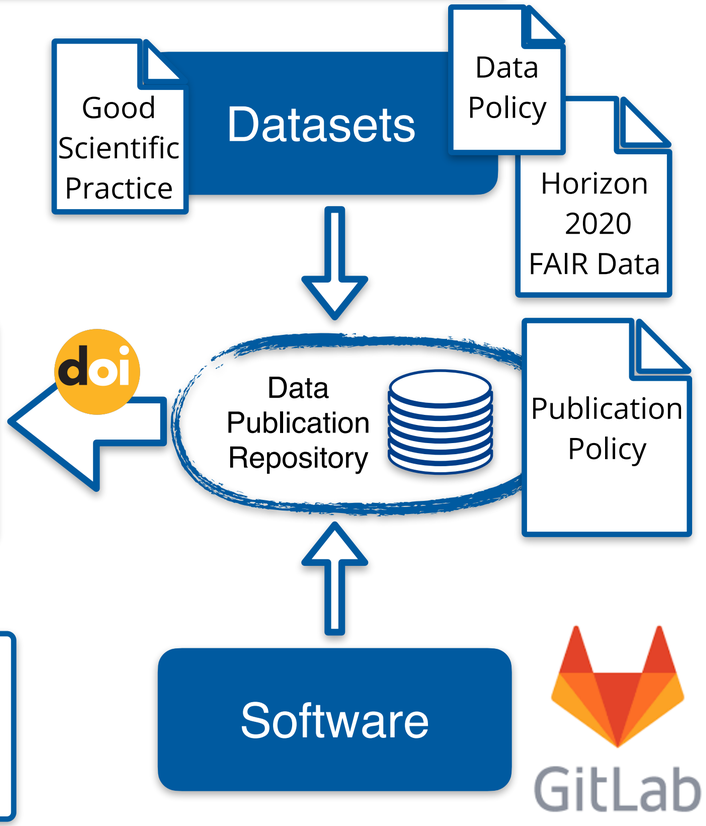
What are the necessary steps towards a full comprehensible and FAIR research experiment ensuring data provenance?

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 **visiting/external scientists** into our project lifecycle and all associated tools?

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



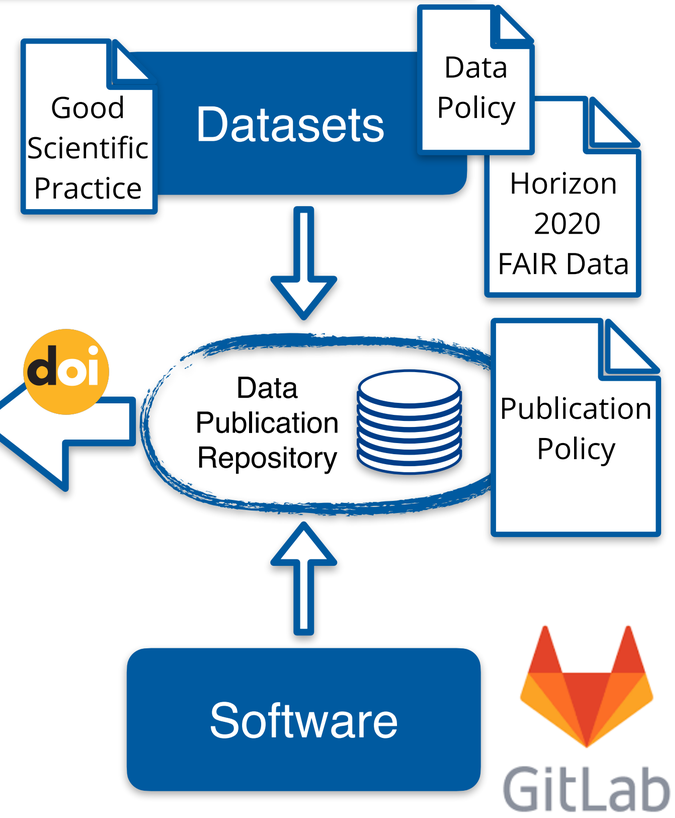
Requirements and Conditions

— HELIPORT was 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 HELIPORT to answer the most important questions of our scientists and :

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 **visiting/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

HELmholtz Sc**I**entific
Project W**OR**kflow Pla**T**form

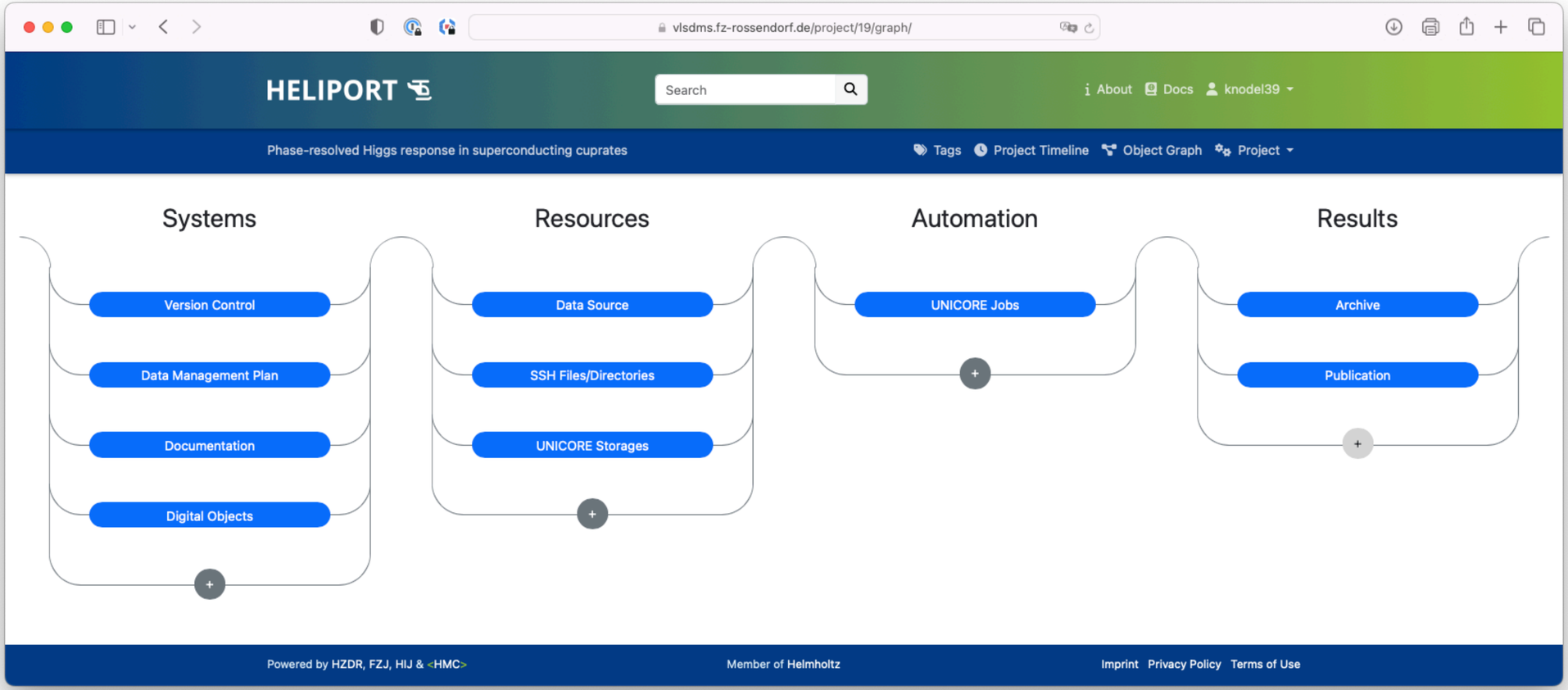
> The HELIPORT project aims at developing a platform which accommodates the **complete life cycle** of a scientific project and links all corresponding programs, systems and workflows to create a more **FAIR** and comprehensible project description.

Project Members:

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

Funded by:

<HMC> HELMHOLTZ Metadata Collaboration



HZDR

HELIPORT

HELmholtz Sc**I**entific
Project W**OR**kflow 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:

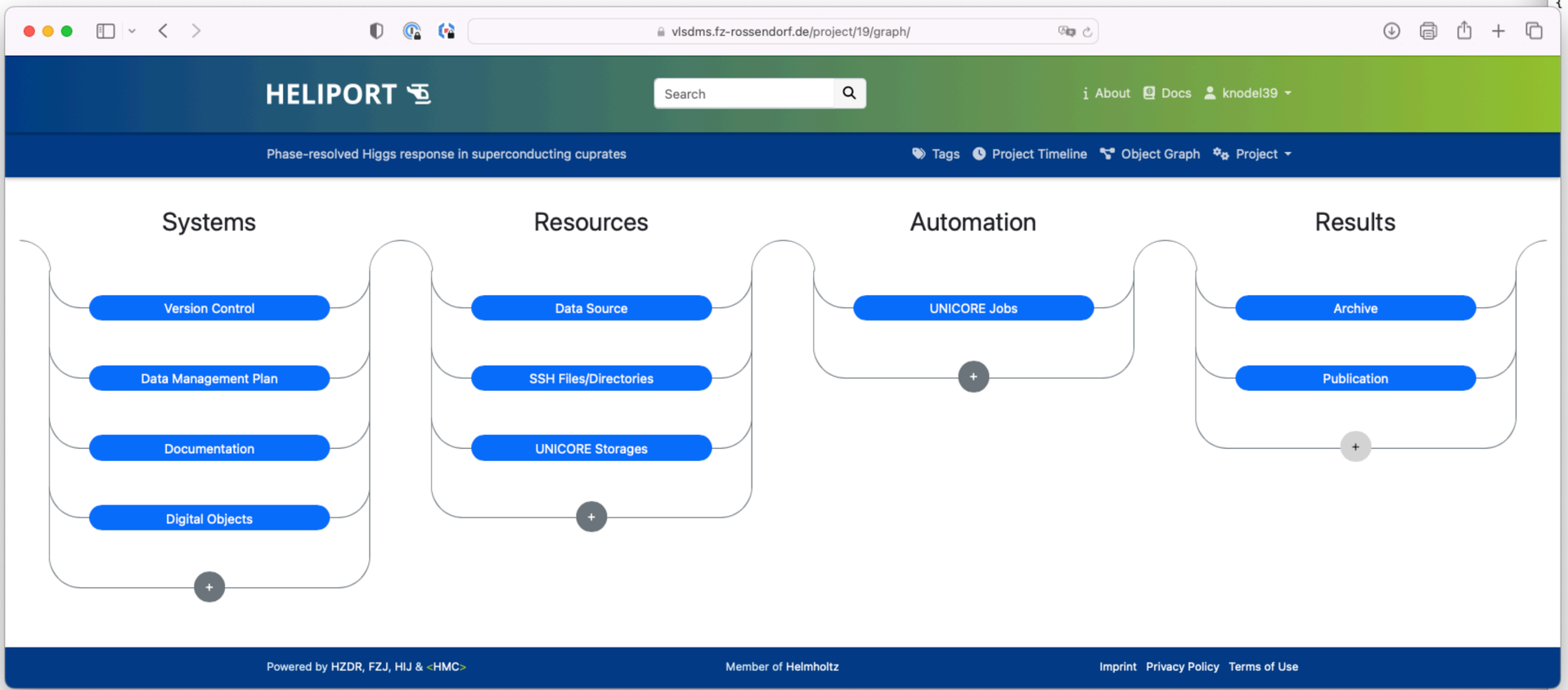
HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

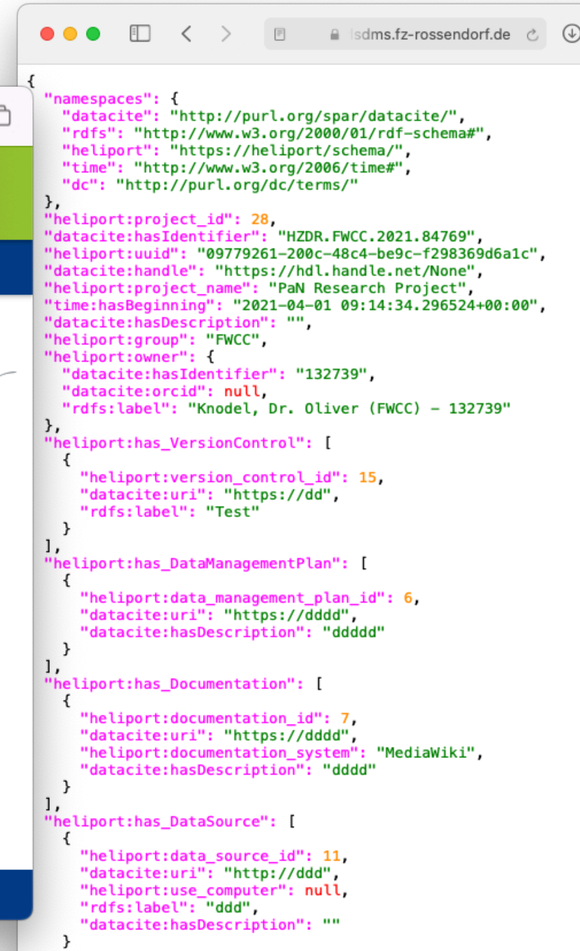
HI JENA Helmholtz Institute Jena

JÜLICH FORSCHUNGSZENTRUM

Funded by:

<HMC> HELMHOLTZ Metadata Collaboration





HELIPORT

HELmholtz Sc**I**entific
Project W**OR**kflow 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:

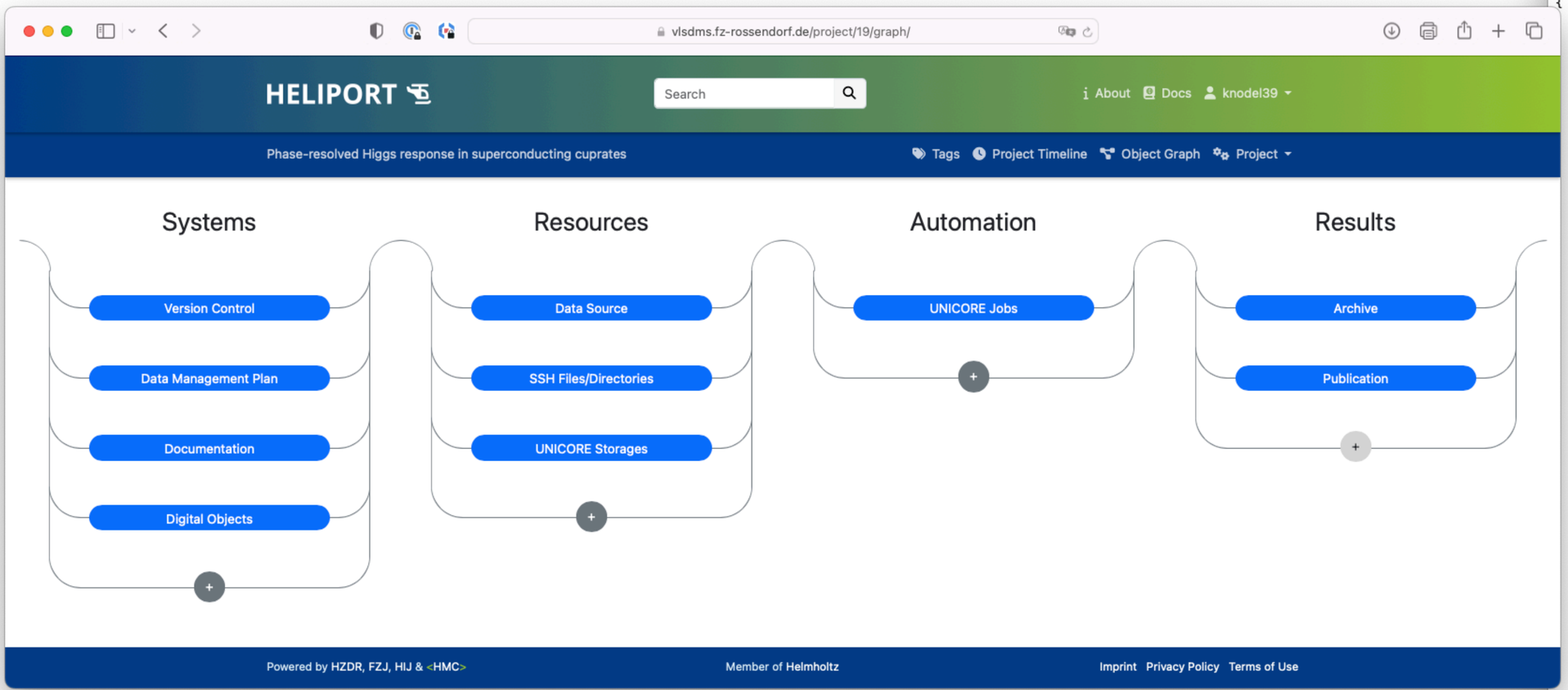
HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

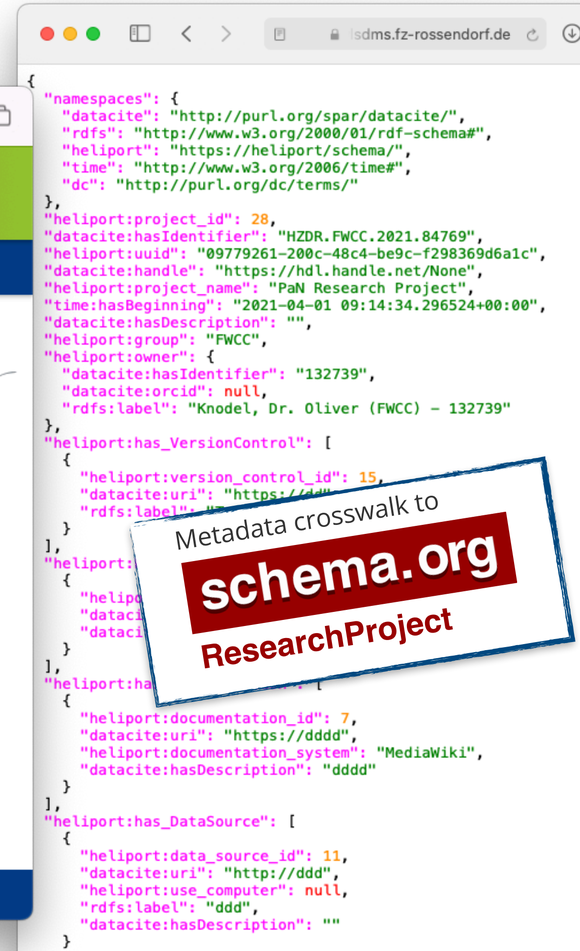
HI JENA Helmholtz Institute Jena

JÜLICH FORSCHUNGSZENTRUM

Funded by:

<HMC> HELMHOLTZ Metadata Collaboration





HELIPORT

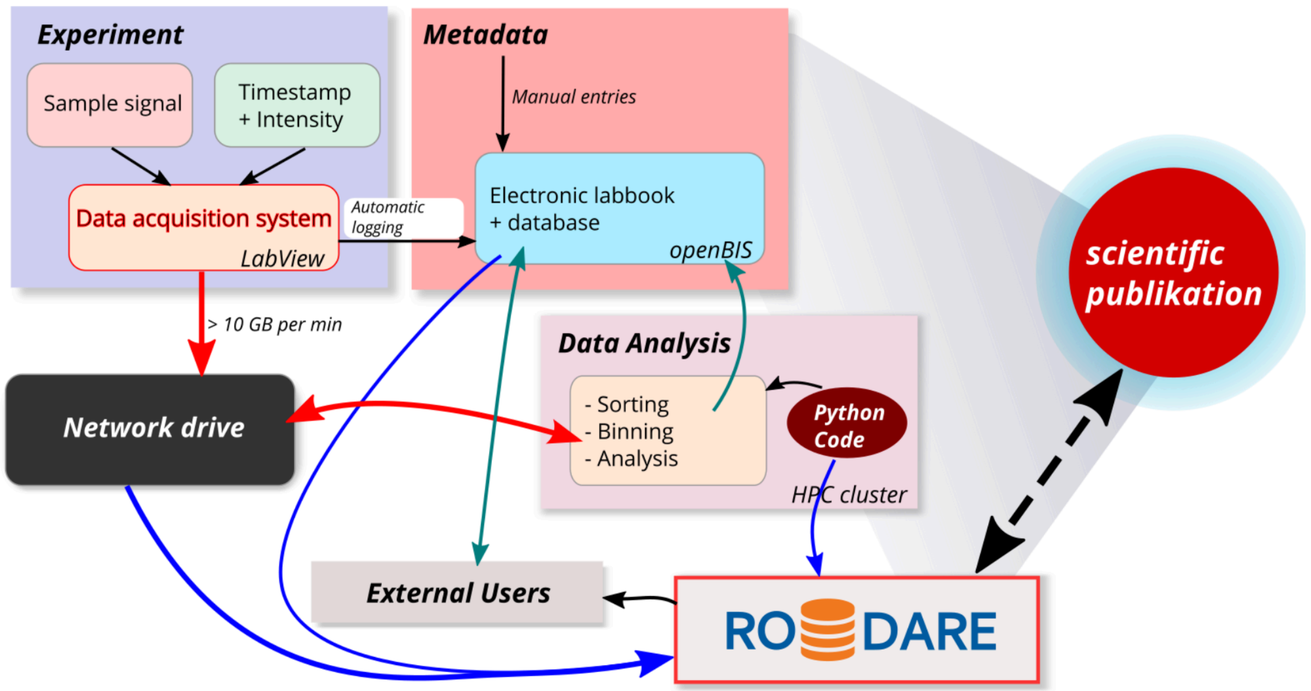
HZDR
HELMHOLTZ ZENTRUM
DRESDEN ROSSENDORF

TELBE in HELIPORT

DRESDEN concept

Data Management View of the TELBE Experiment

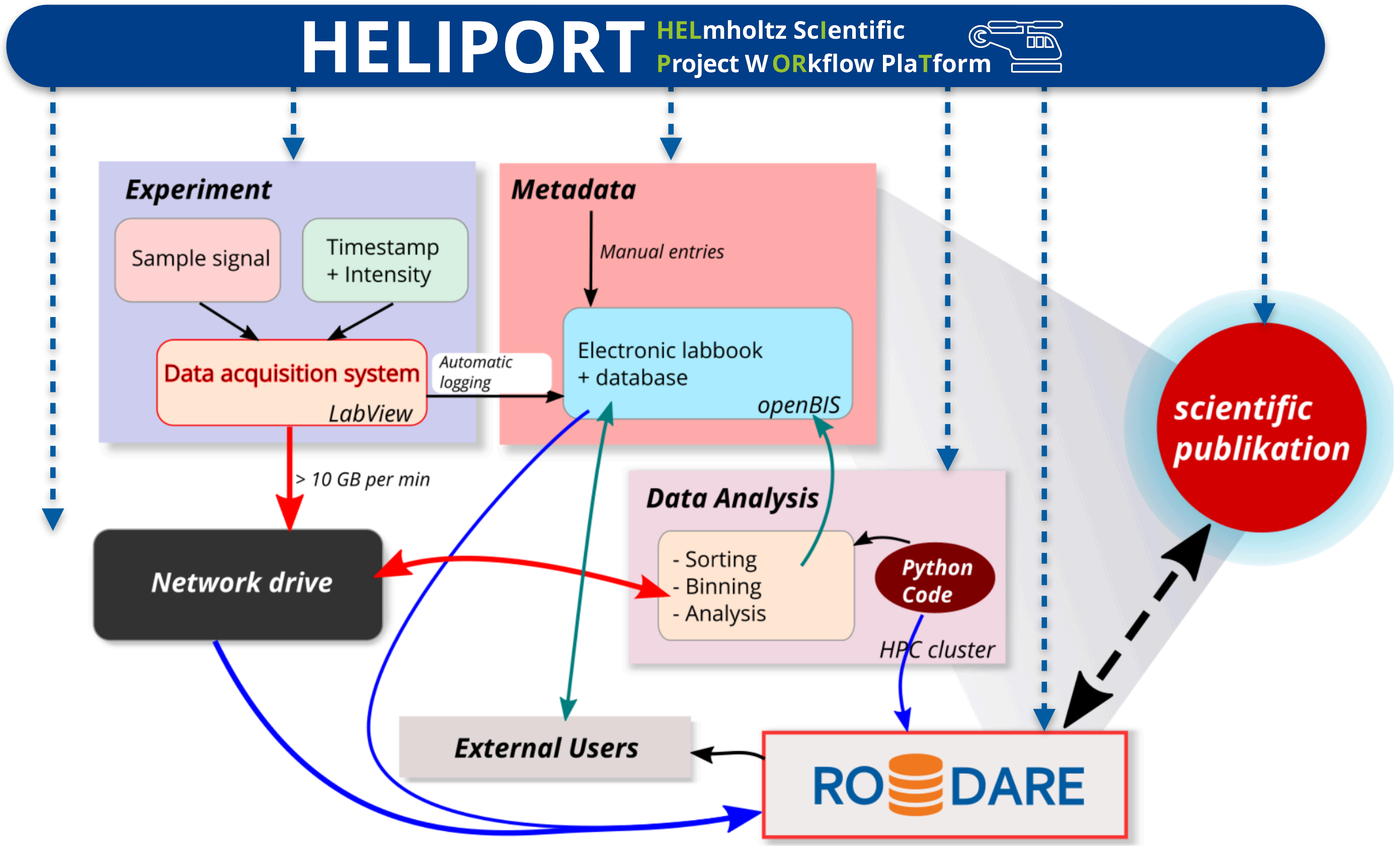
— Terahertz facility at the ELBE center for High-Power Radiation Sources.



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

Data Management View of the TELBE Experiment

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

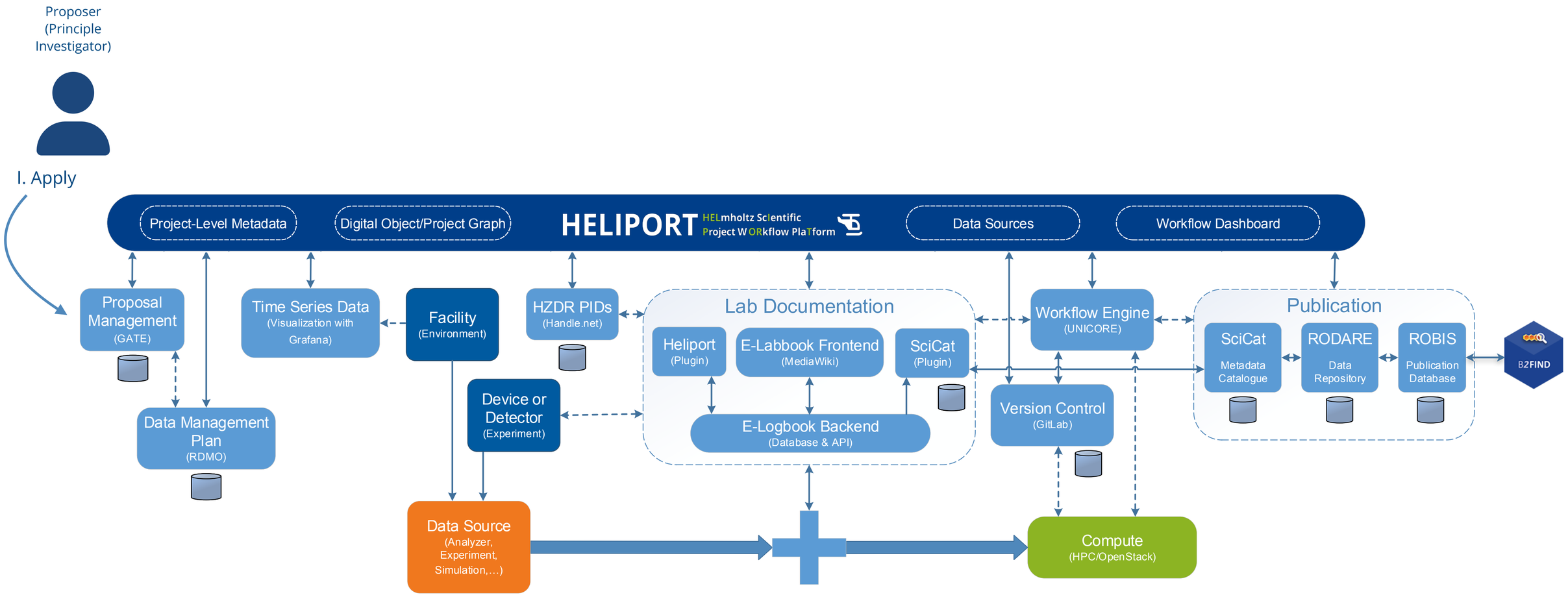


Deinert, Jan-Christophi. (2021, November 18). TELBE data analysis workflow and the PaN training platform UX — Booklet of presentations from the PaN EOSC Symposium 2021. Zenodo. https://doi.org/10.5281/zenodo.5636331

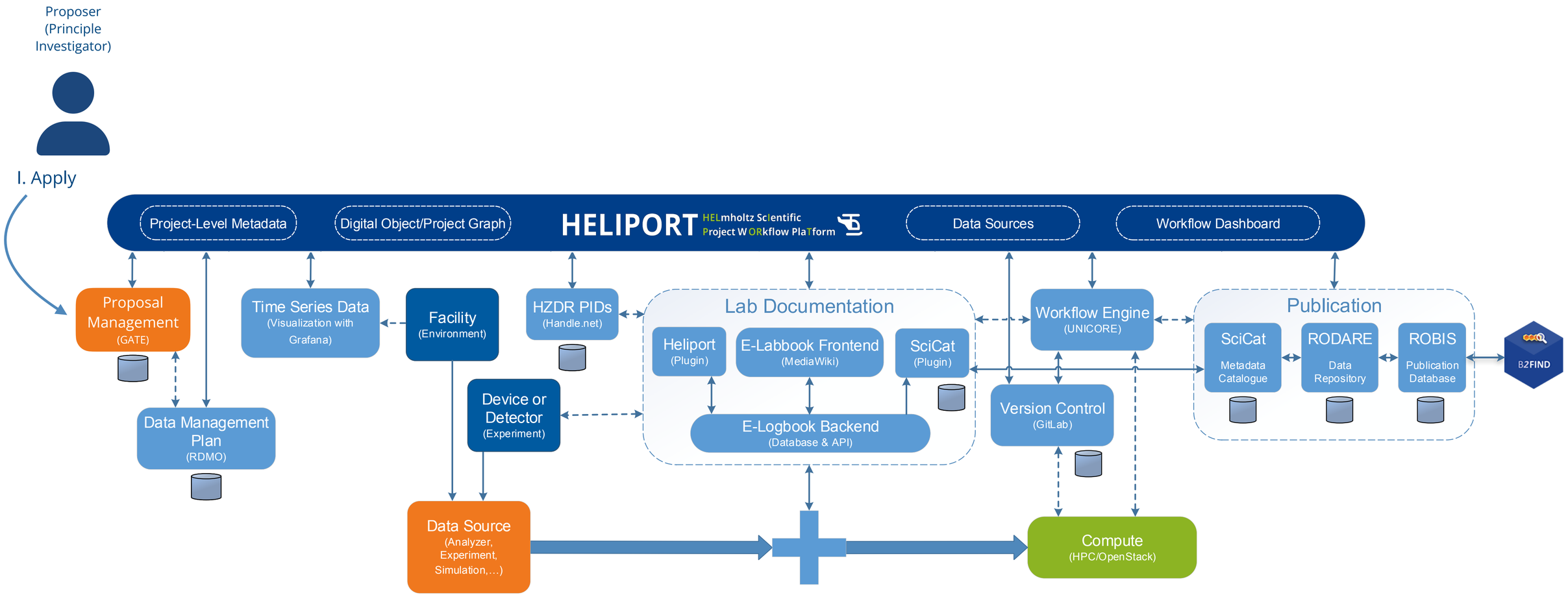
Mapping of the TELBE Resources to HELIPORT

Mapping of the TELBE Resources to HELIPORT

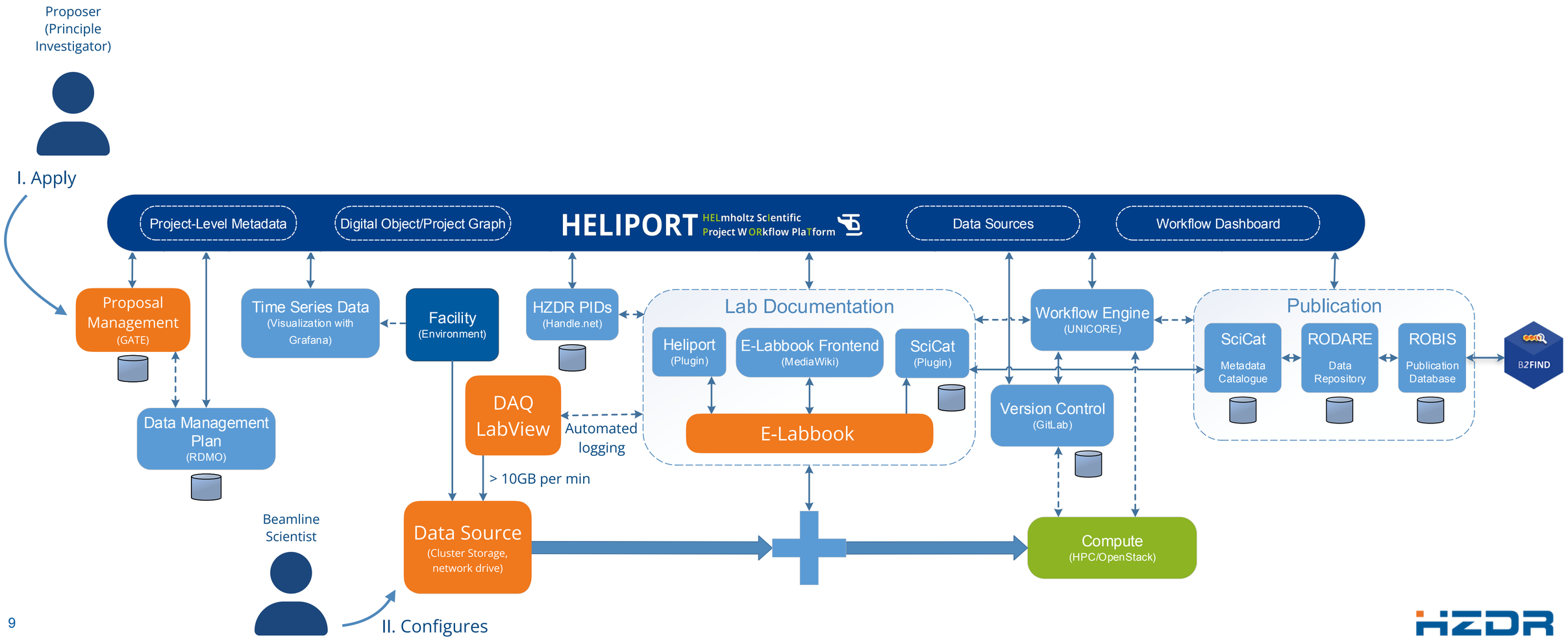
Mapping of the TELBE Resources to HELIPORT



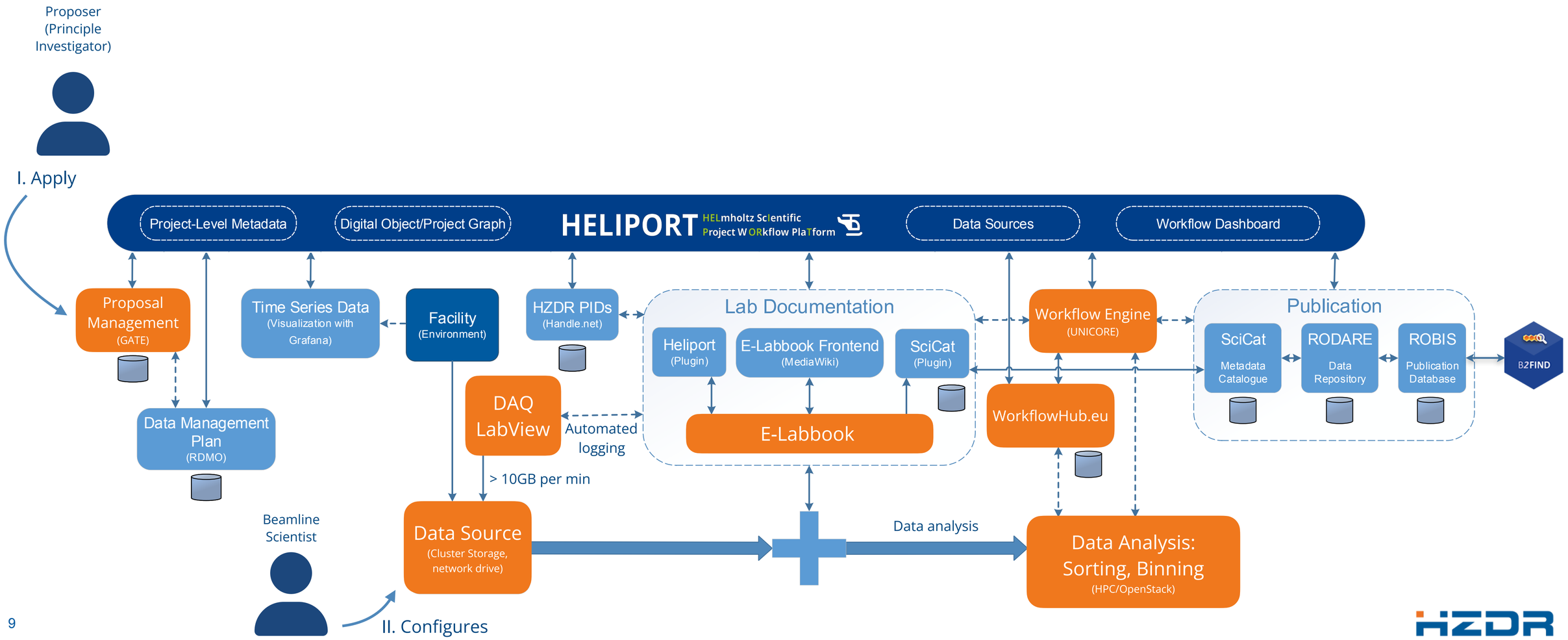
Mapping of the TELBE Resources to HELIPORT



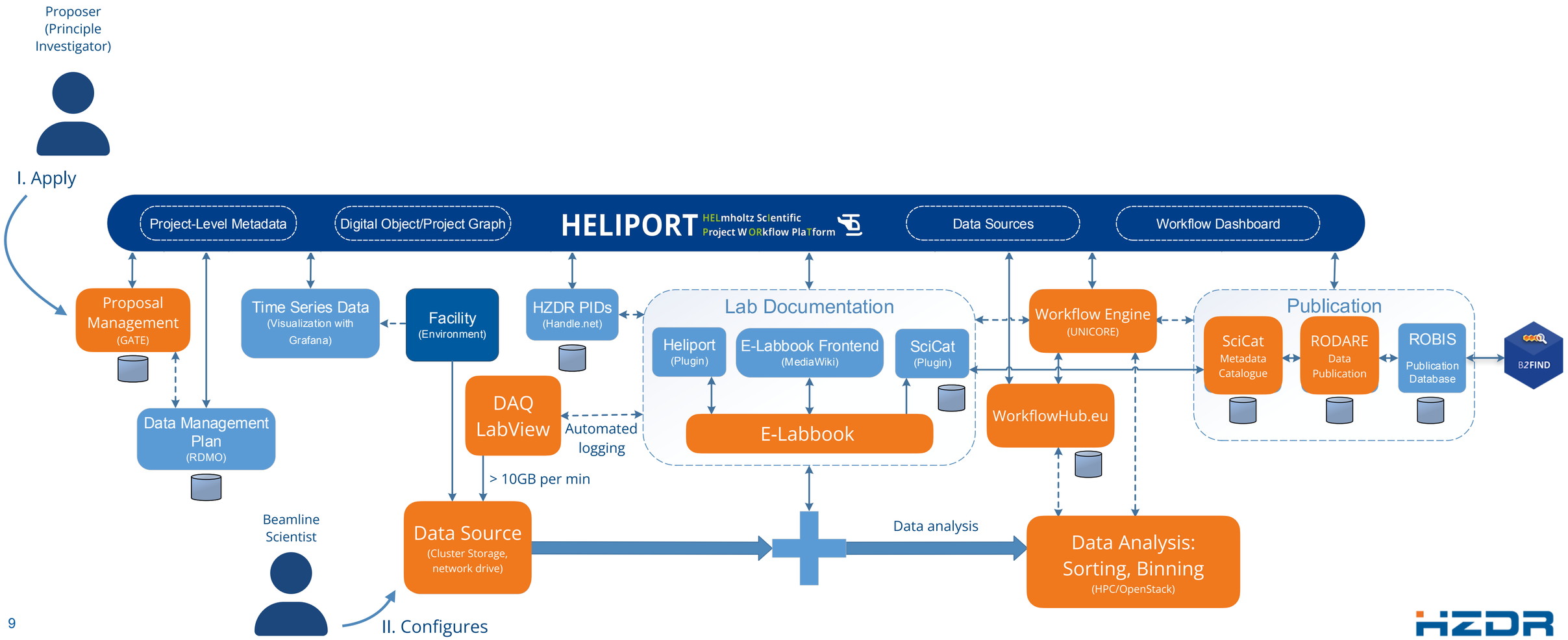
Mapping of the TELBE Resources to HELIPORT



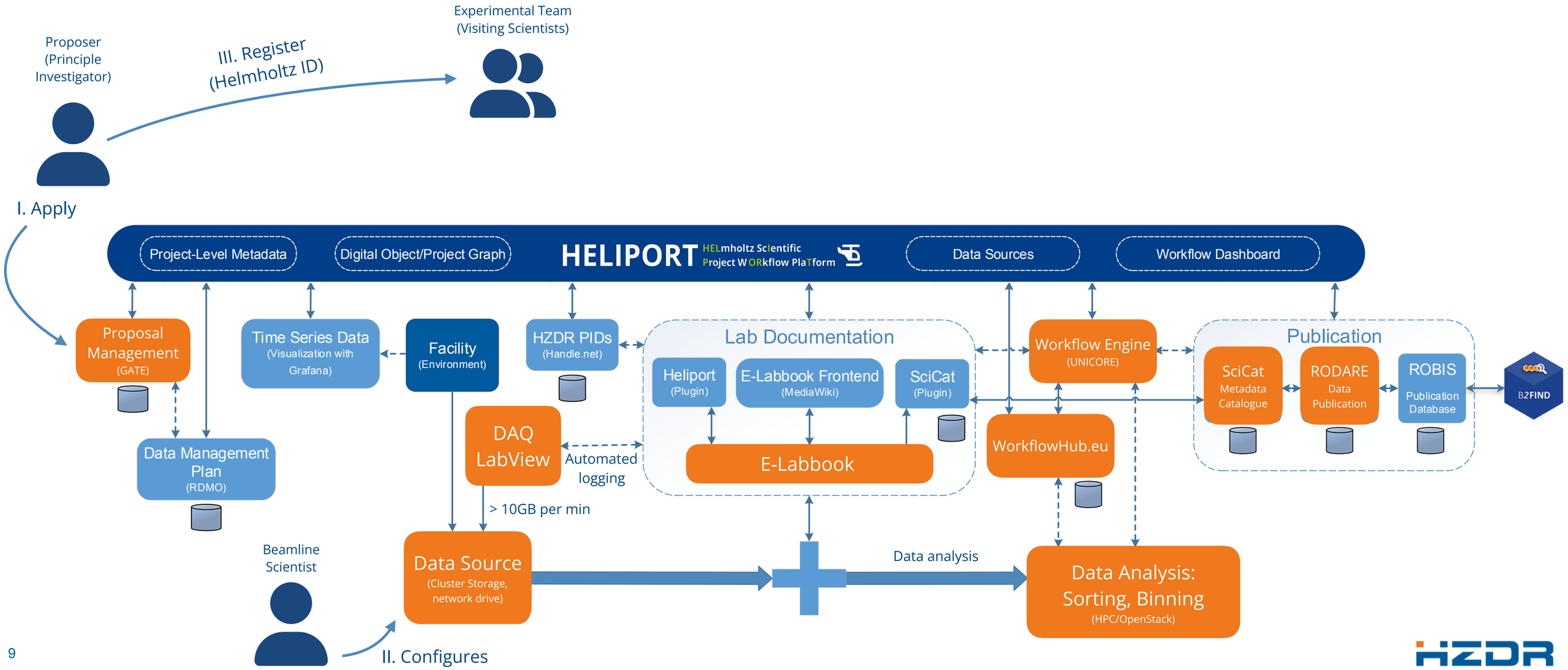
Mapping of the TELBE Resources to HELIPORT



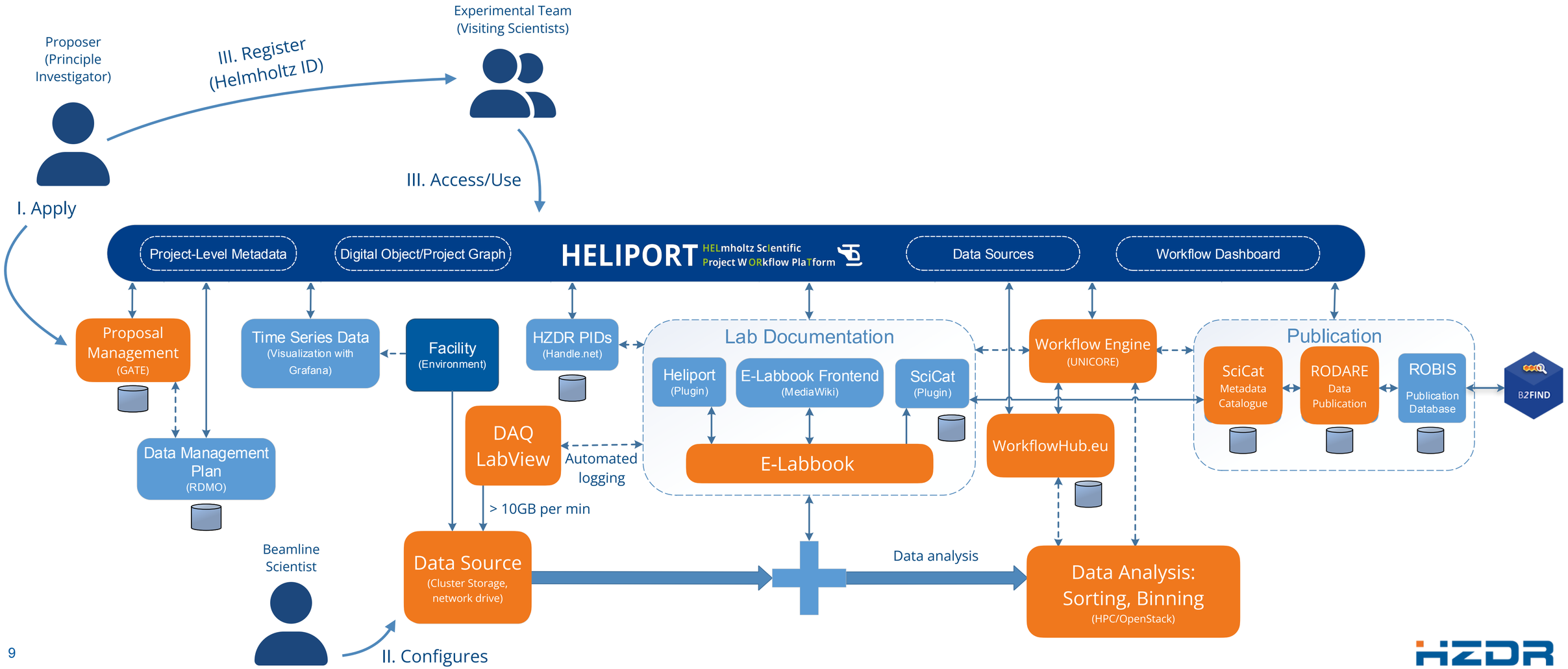
Mapping of the TELBE Resources to HELIPORT



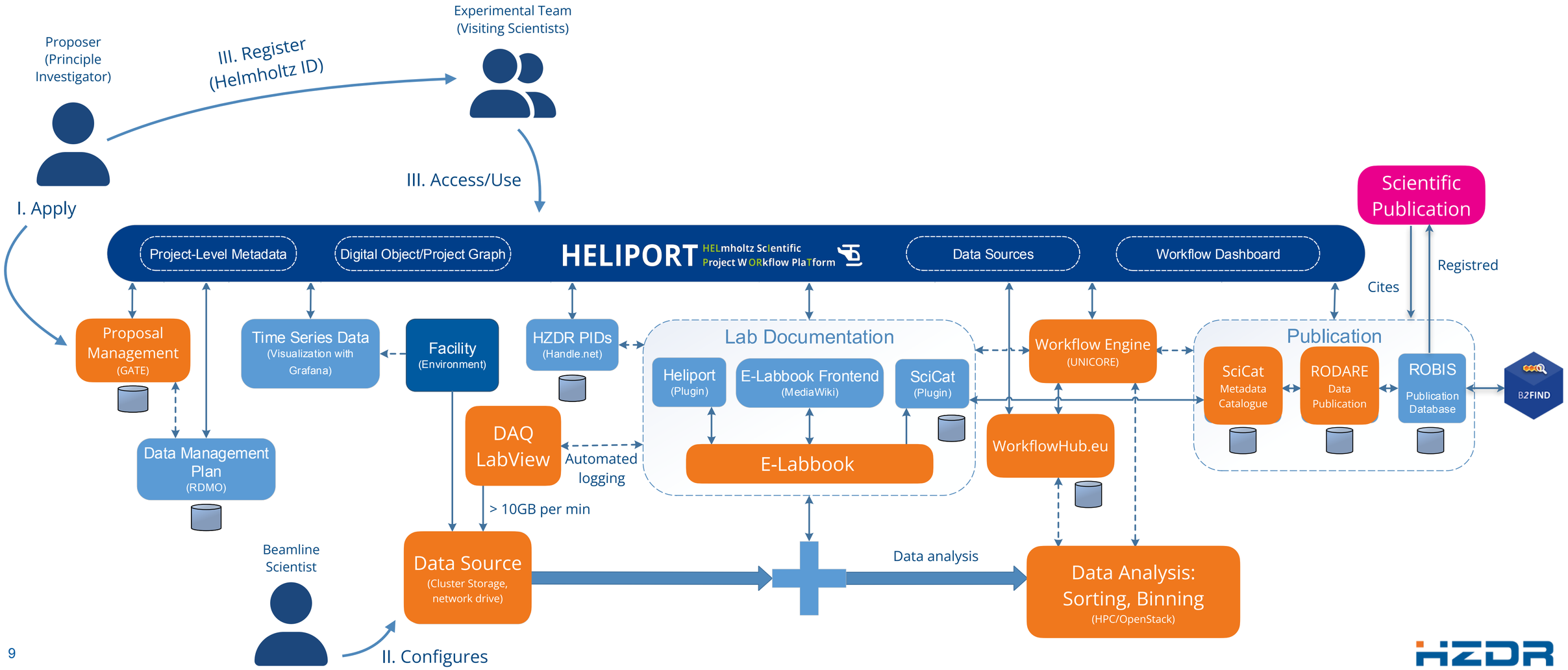
Mapping of the TELBE Resources to HELIPORT



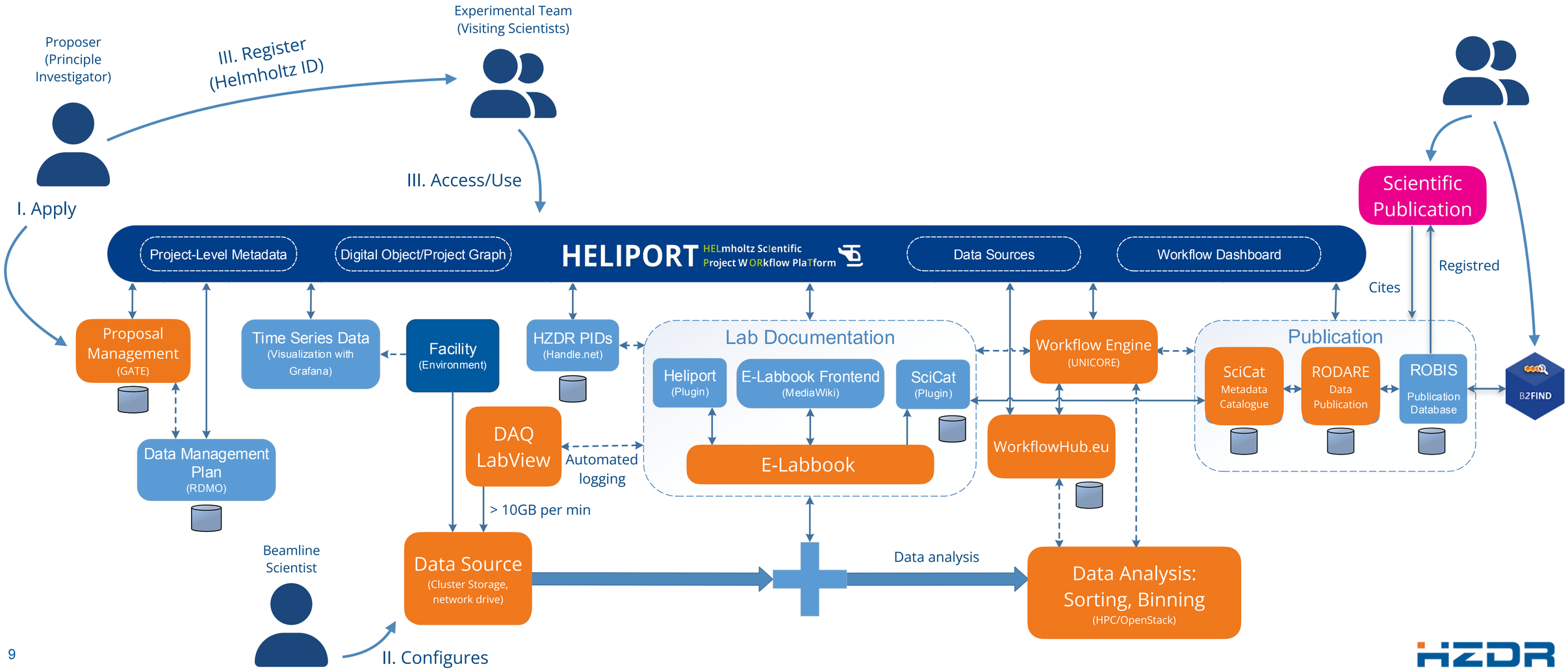
Mapping of the TELBE Resources to HELIPORT



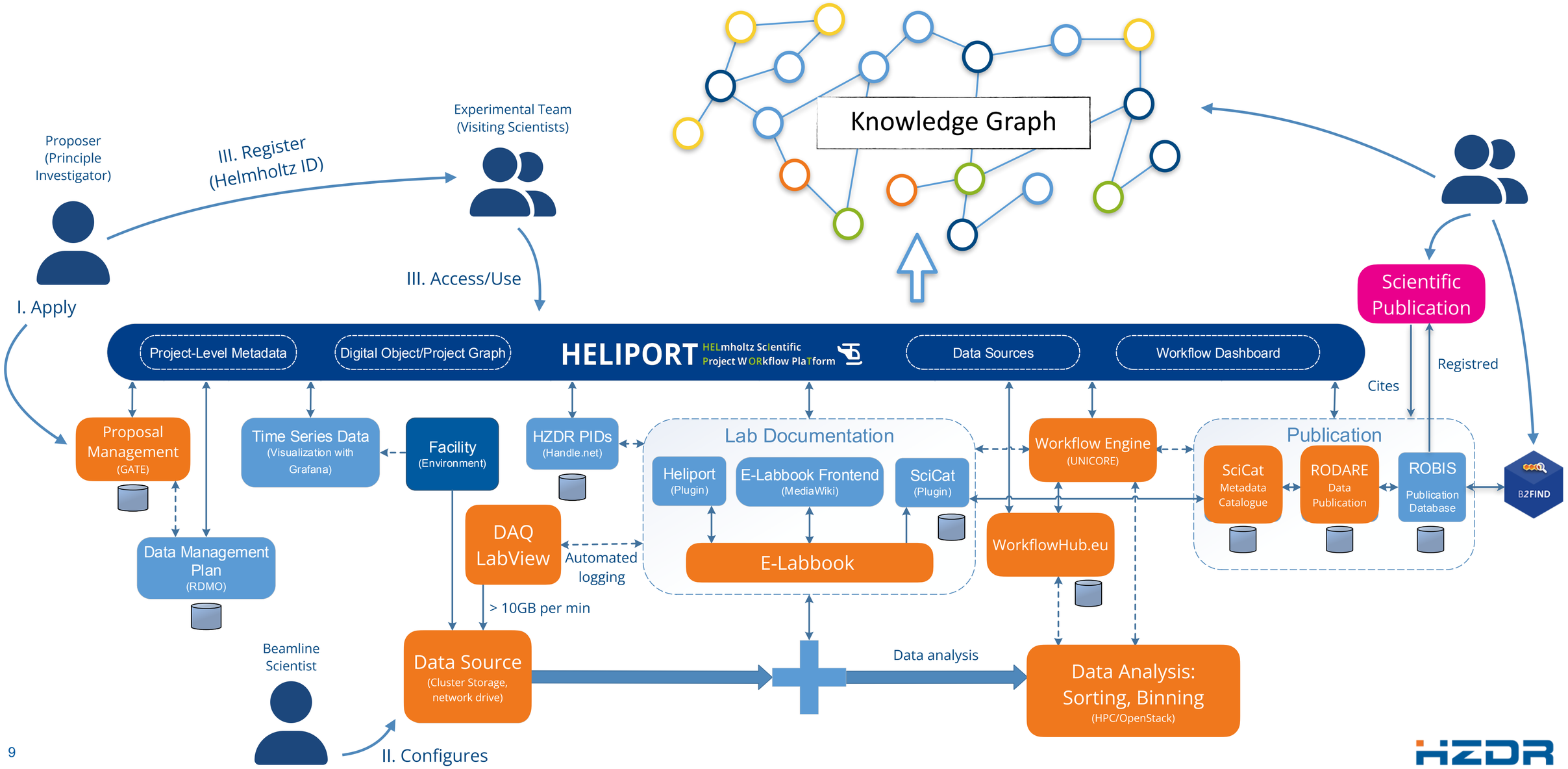
Mapping of the TELBE Resources to HELIPORT



Mapping of the TELBE Resources to HELIPORT



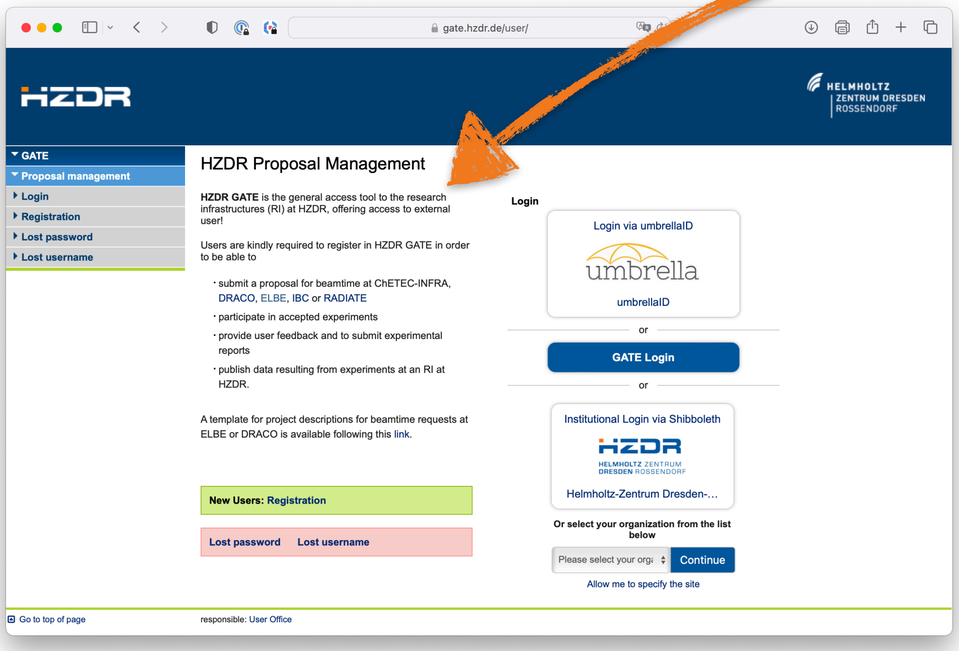
Mapping of the TELBE Resources to HELIPORT

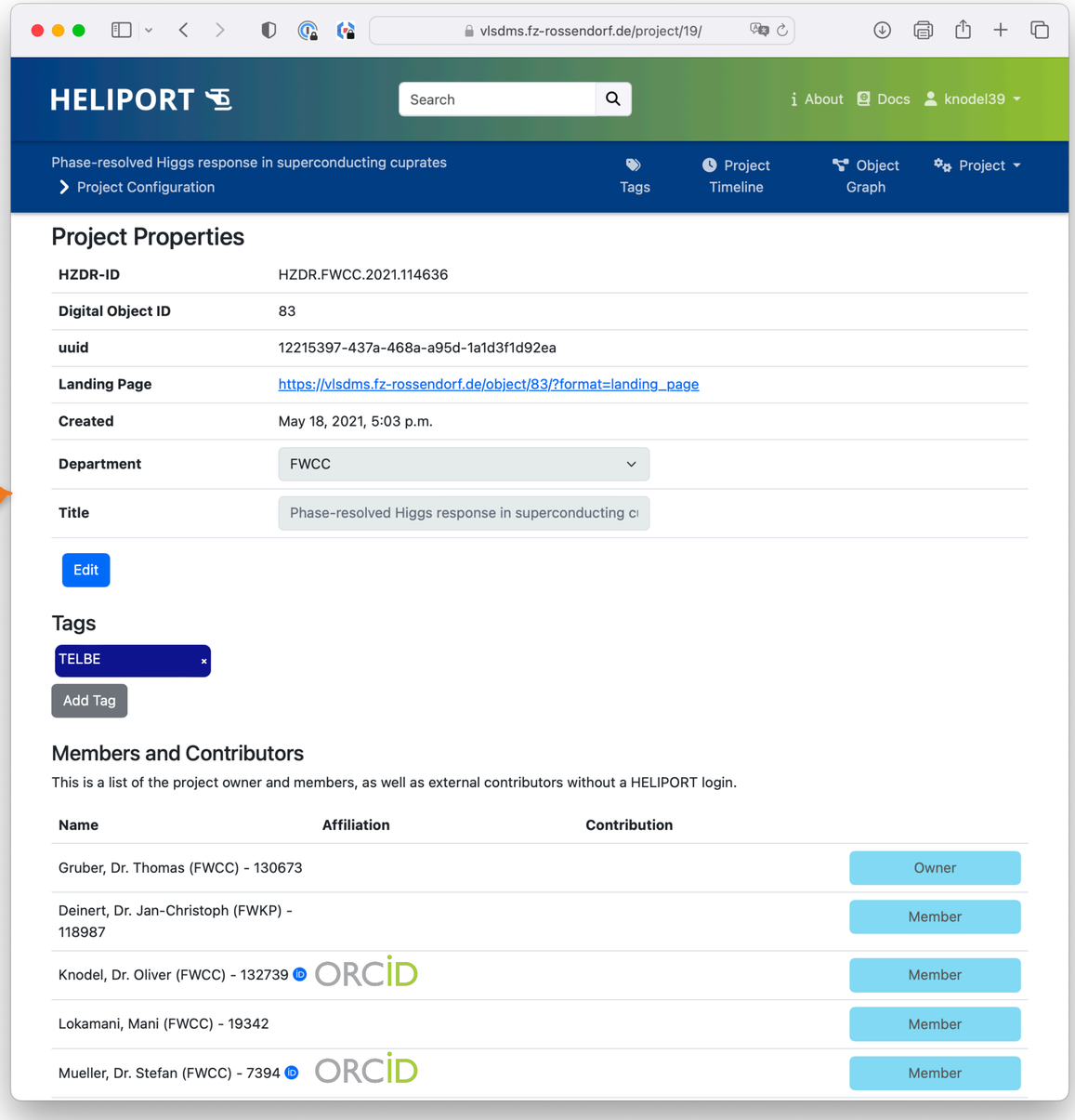


I. Proposal Submission

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

- Title, Authors, Description,
- Beamtime schedule,
- Large-scale facility used,
- Scientific method (PaNET)

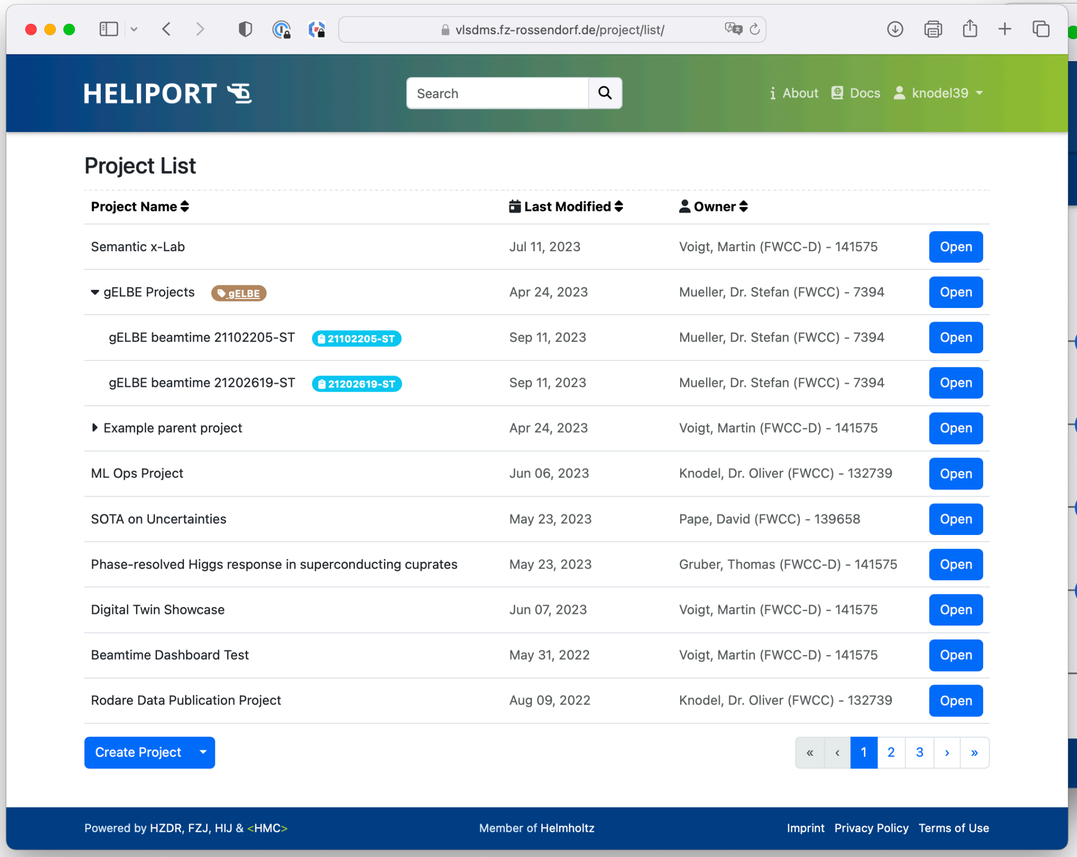


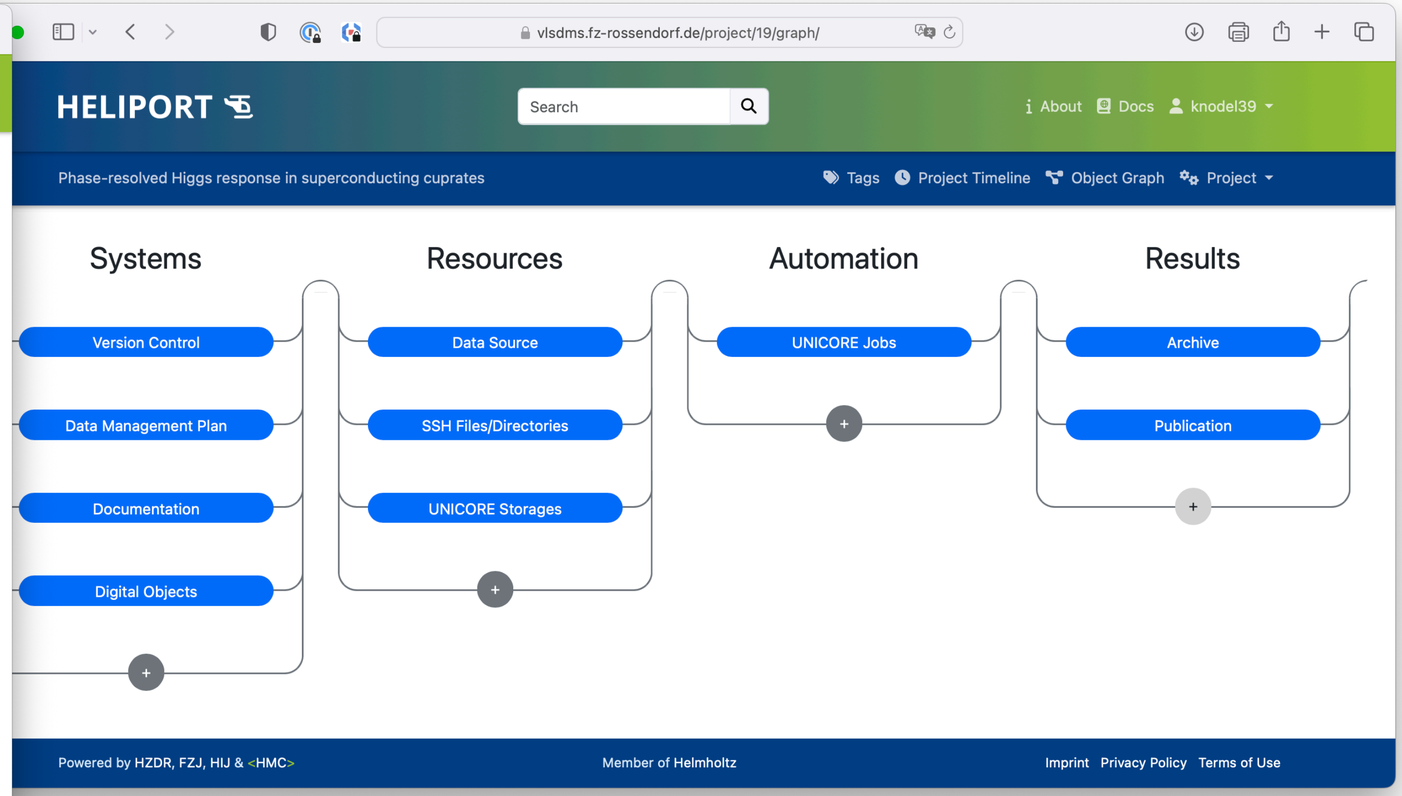


II. Project List and Dashboard

— 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.

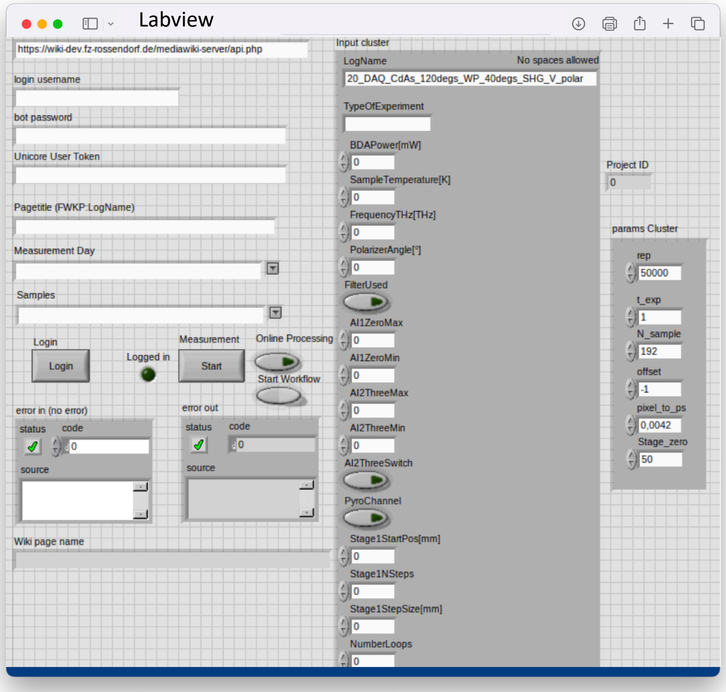
— Tags and sub-projects including inheritance are possible in the project list.





III. Detector Control and Workflows

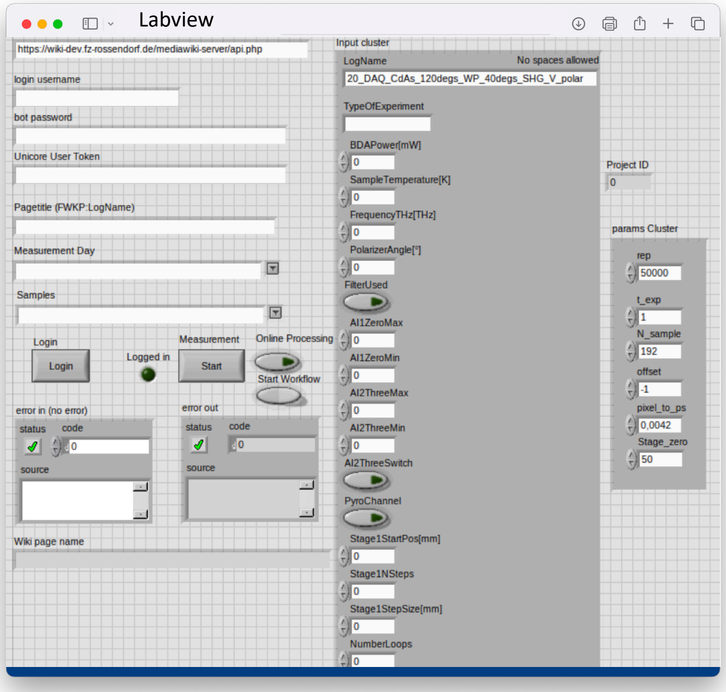
— The HELIPORT REST-API enables the transfer of metadata between HELIPORT and external systems (e.g. the detector control in LabView).



III. Detector Control and Workflows

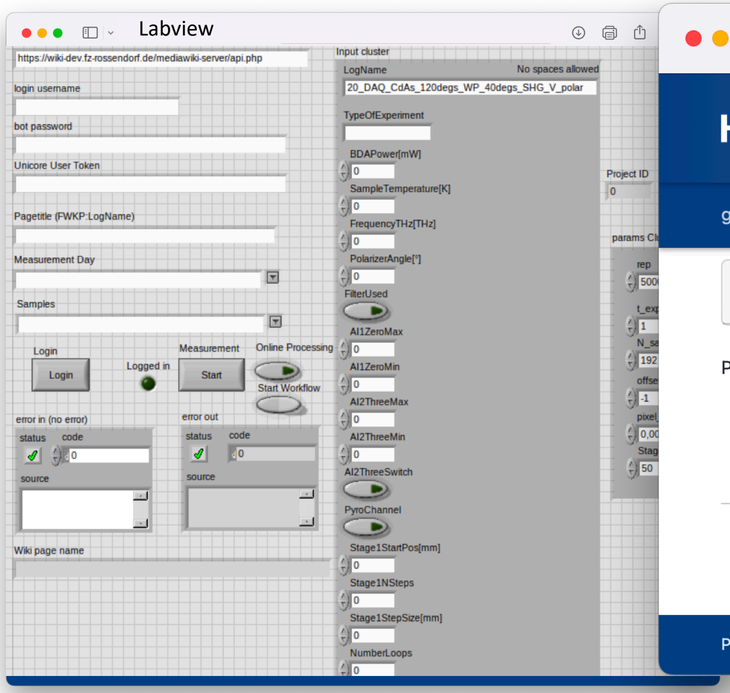
— The HELIPORT REST-API enables the transfer of metadata between HELIPORT and external systems (e.g. the detector control in LabView).

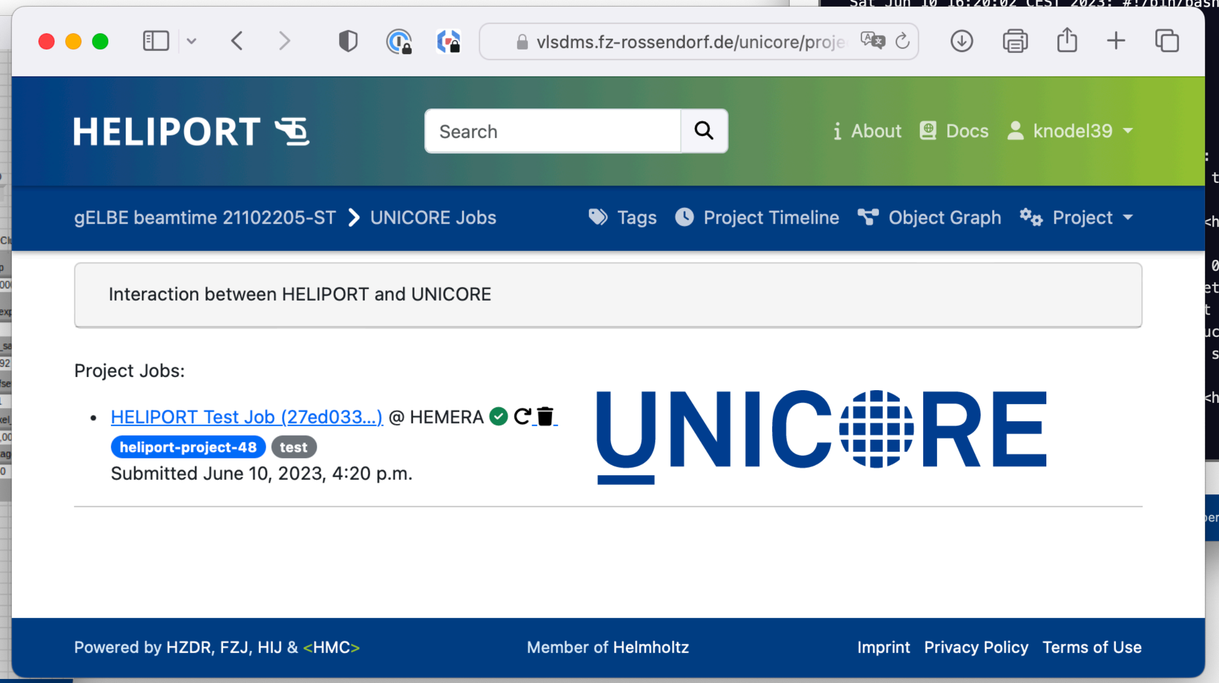
— Via the API, LabView can control the integrated workflow management (e.g. UNICORE), which provides metadata for provenance information.

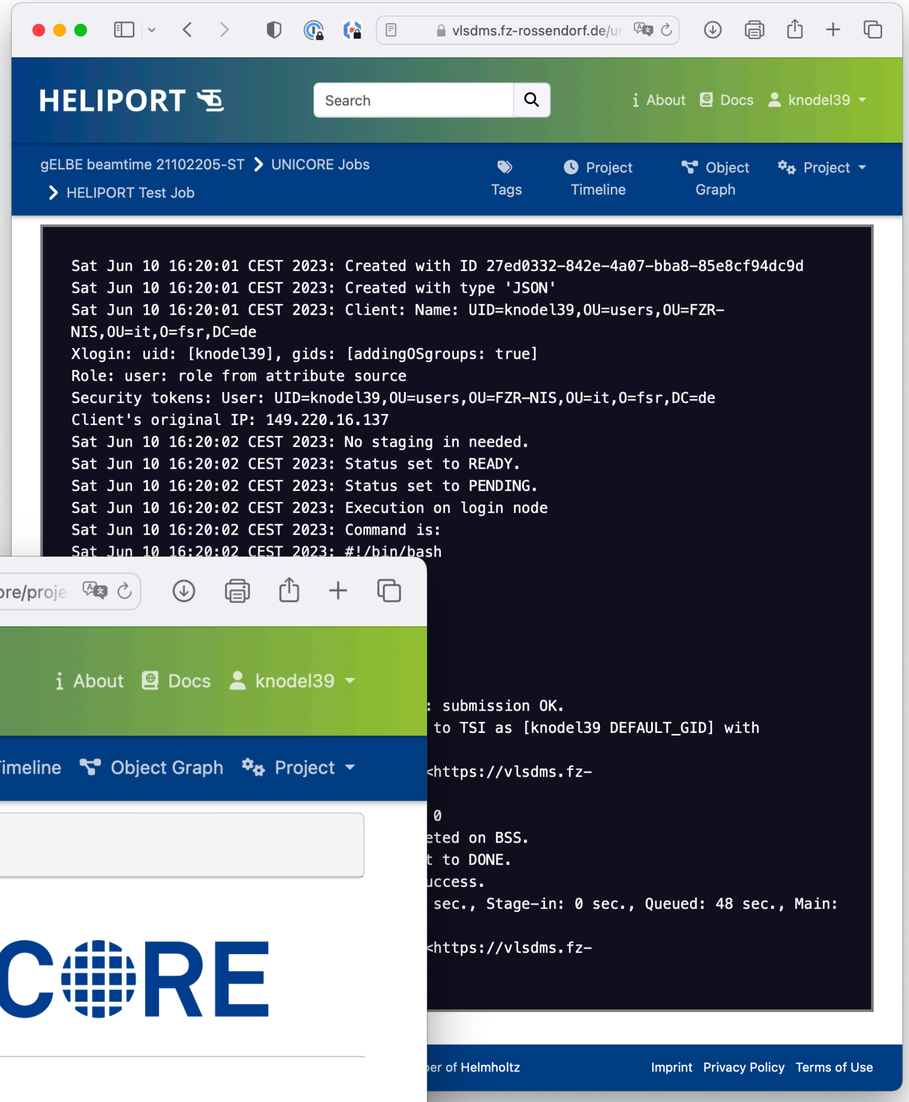


III. Detector Control and Workflows

- The HELIPORT REST-API enables the transfer of metadata between HELIPORT and external systems (e.g. the detector control in LabView).
- Via the API, LabView can control the integrated workflow management (e.g. UNICORE), which provides metadata for provenance information.
- The output of workflows (on our HPC cluster) can be accessed by any project member directly in the HELIPORT web frontend.

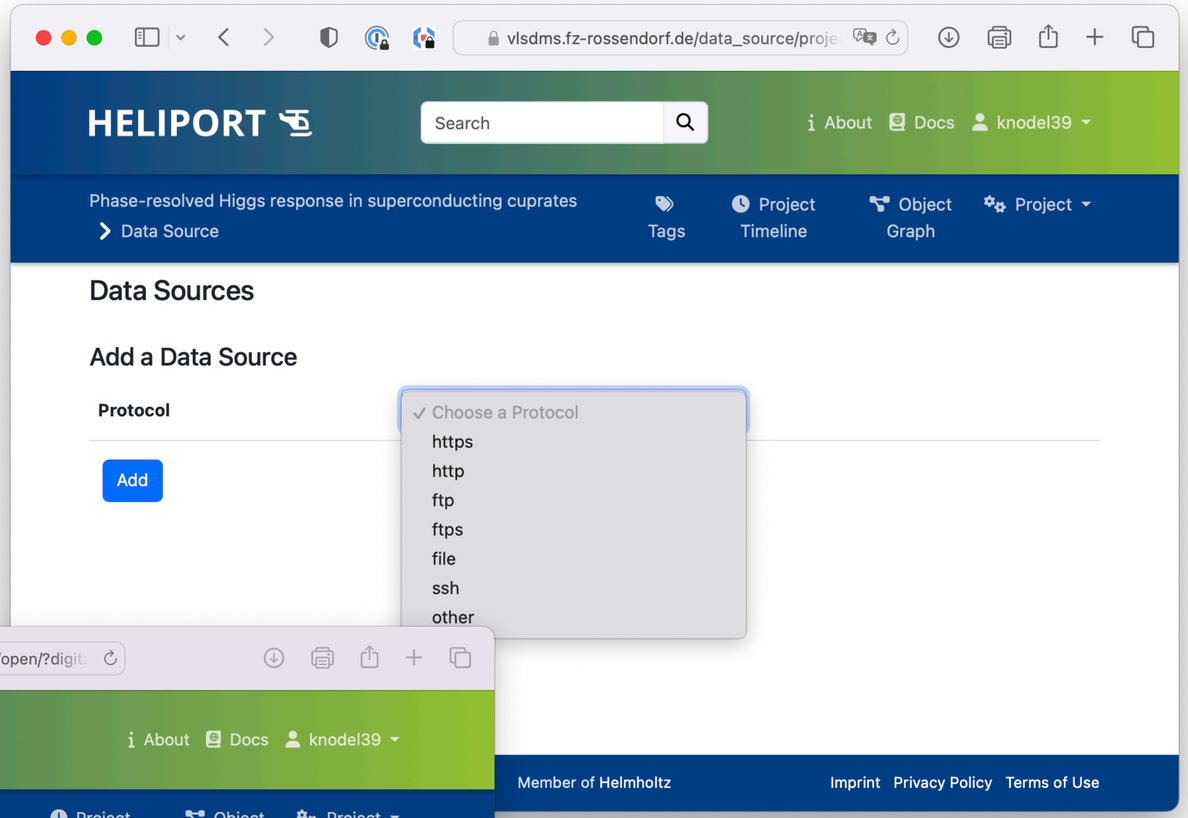


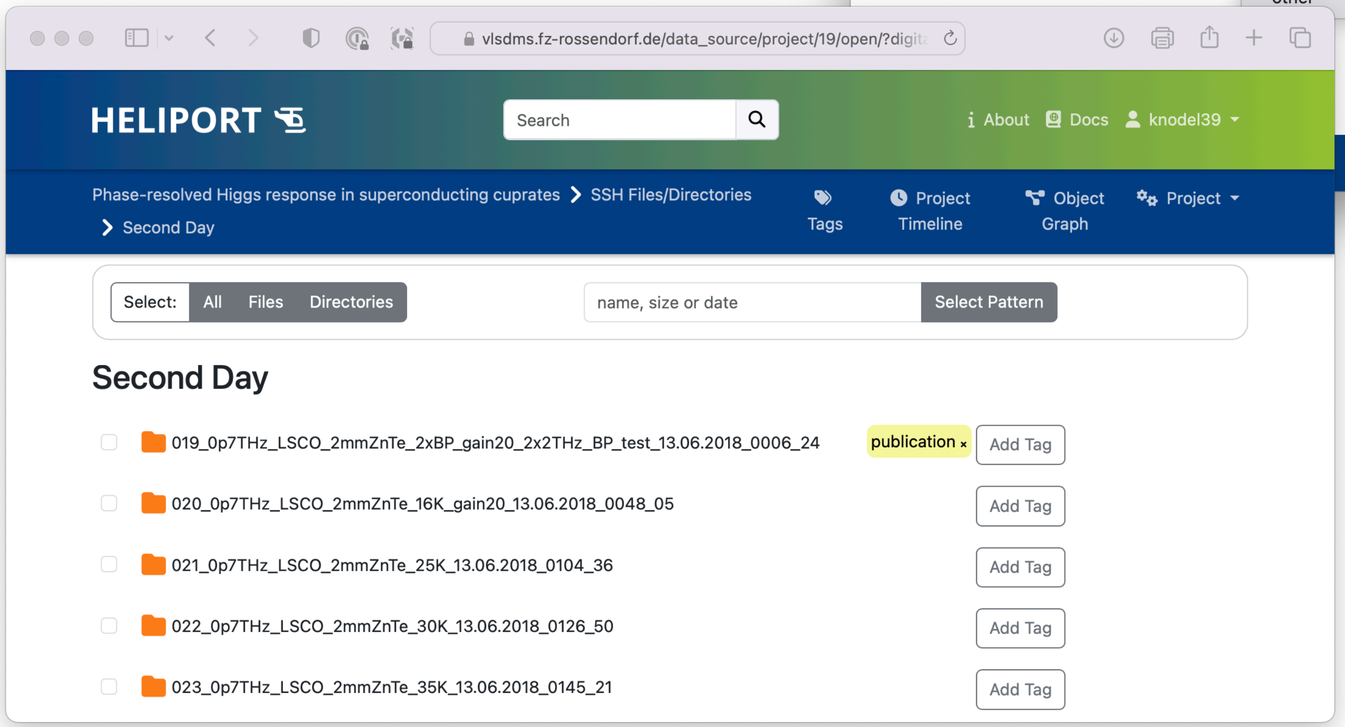




IV. Data Sources

- Folders and Files in our internal filesystem can be registered in HELIPORT as data source.
- Each member of a HELIPORT project (Login via Helmholtz ID) has access to the files and folders.
- The provenance of the data sets generated from an experiment is entirely comprehensible.

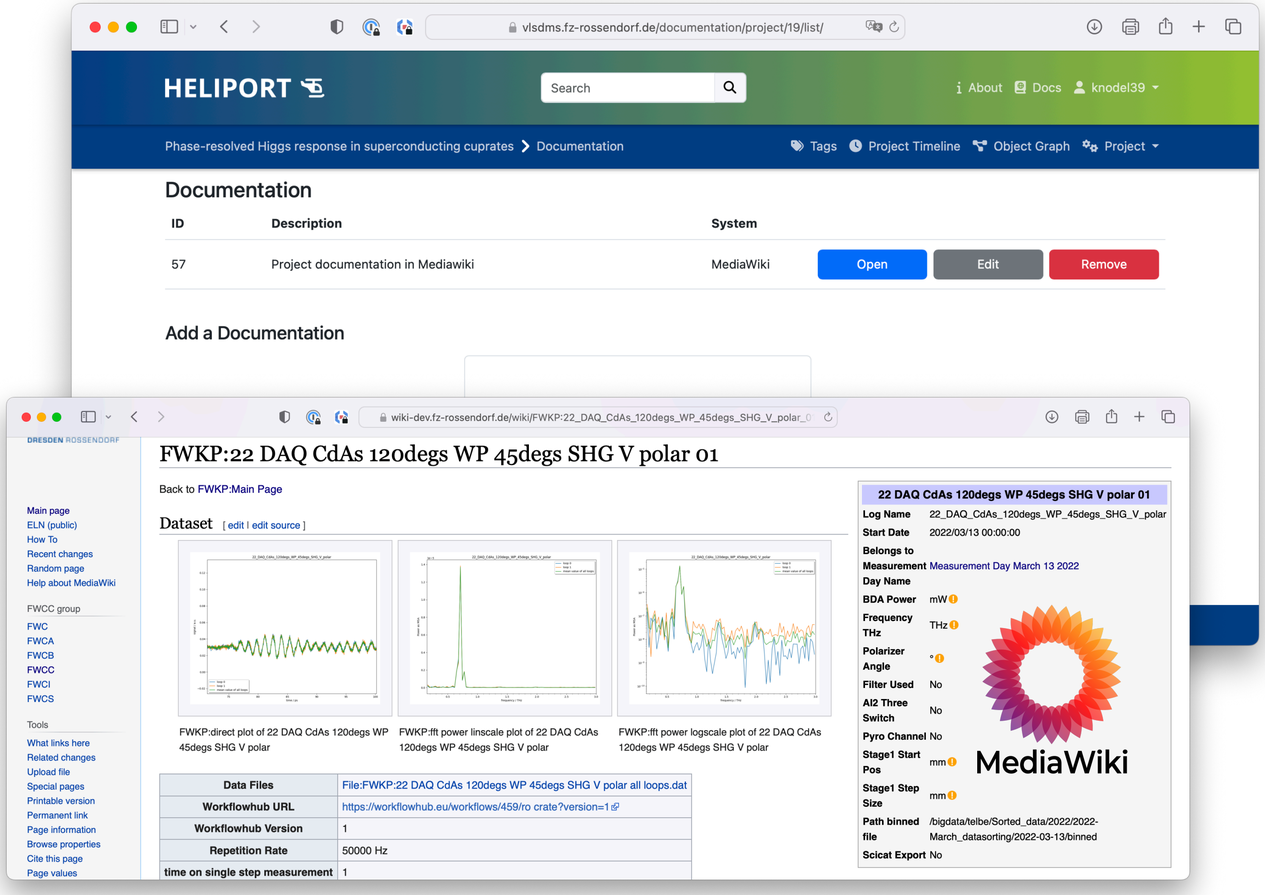




V. Resources: Documentation and Repositories

The documentation section is typically used to refer to all internal and external systems or services used:

— E-Labbook (Mediawiki),

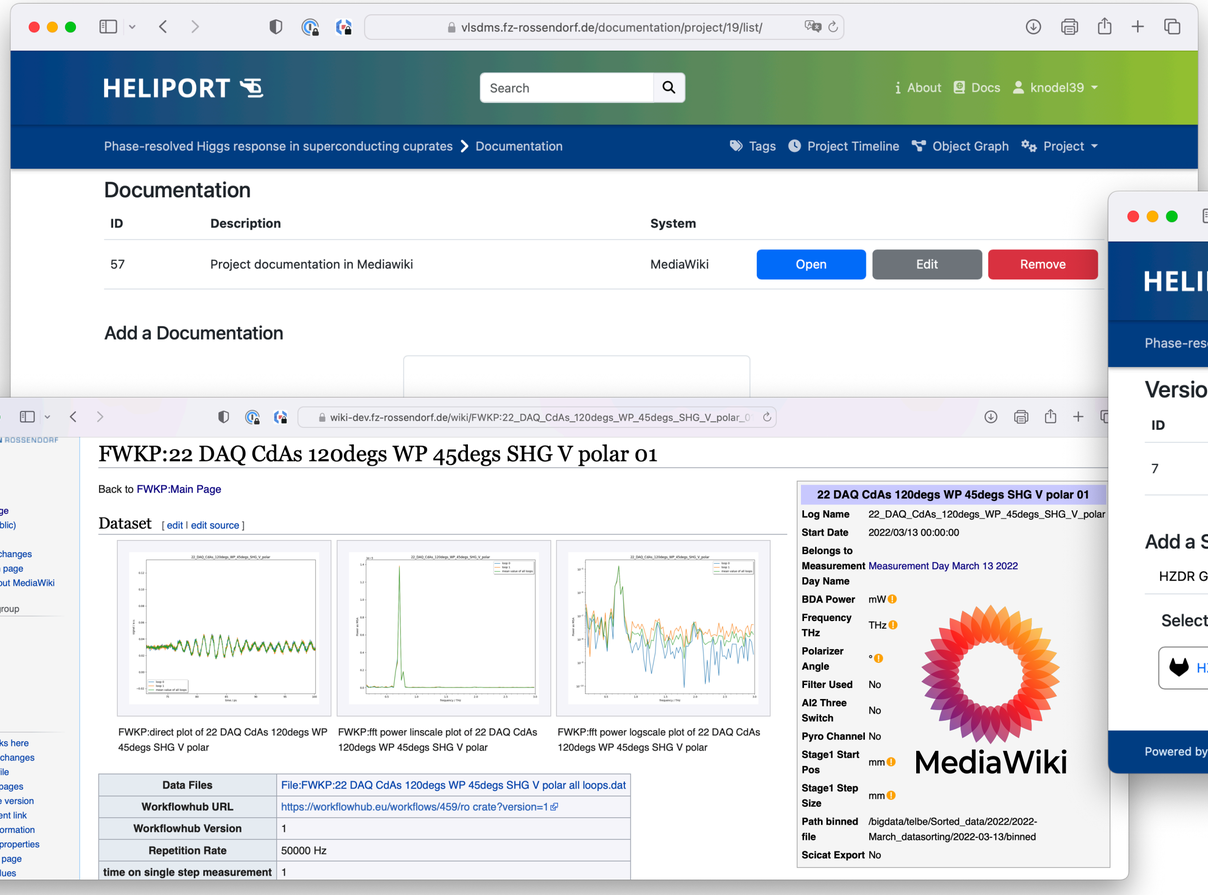


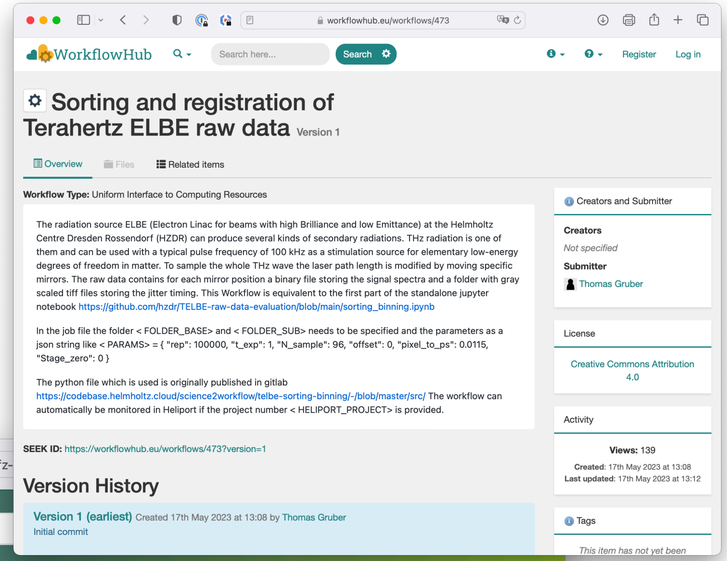
V. Resources: Documentation and Repositories

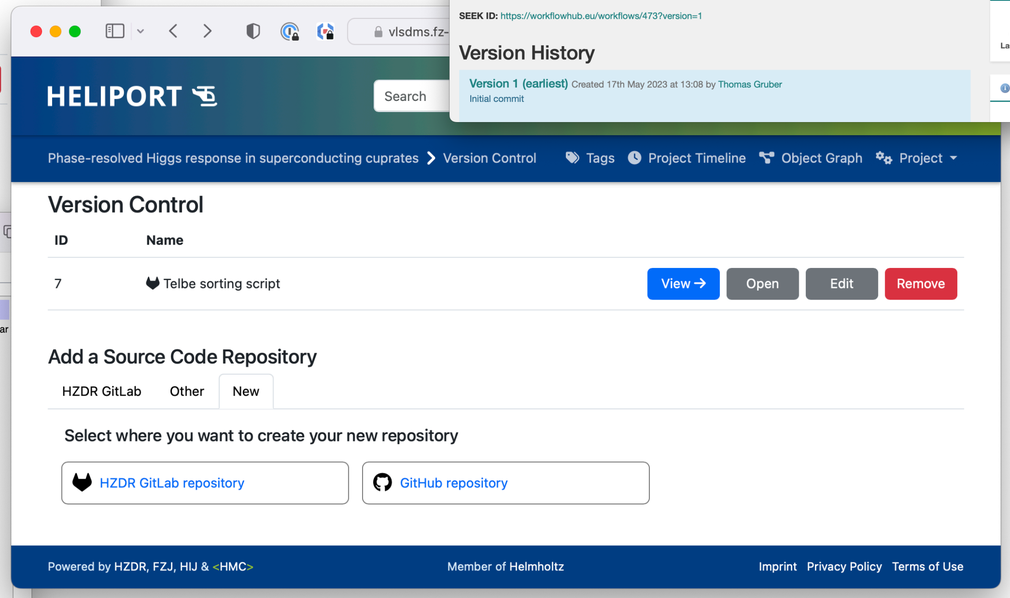
The documentation section is typically used to refer to all internal and external systems or services used:

— E-Labbook (Mediawiki),

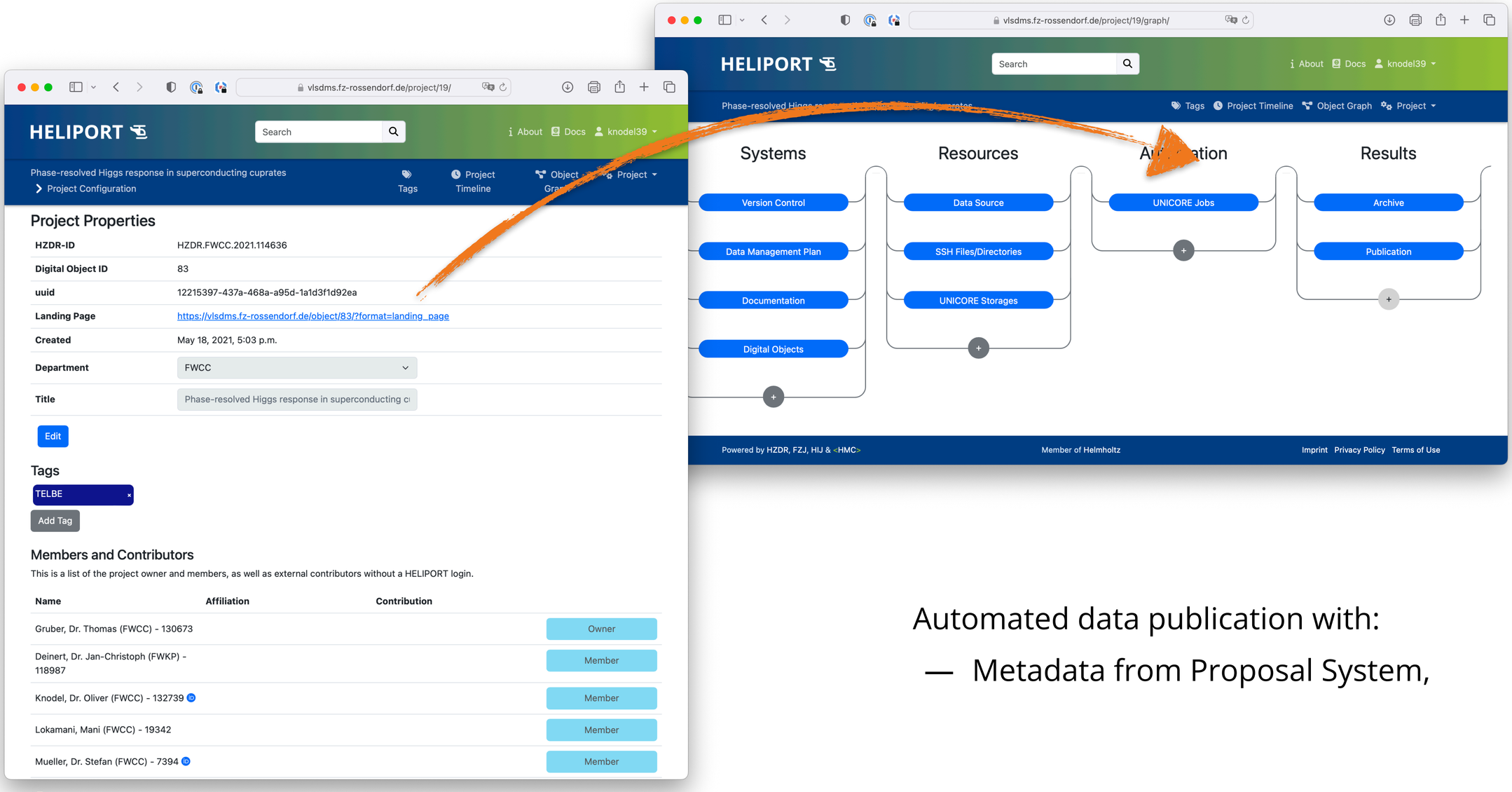
— GitLab, Github, Workflowhub, …







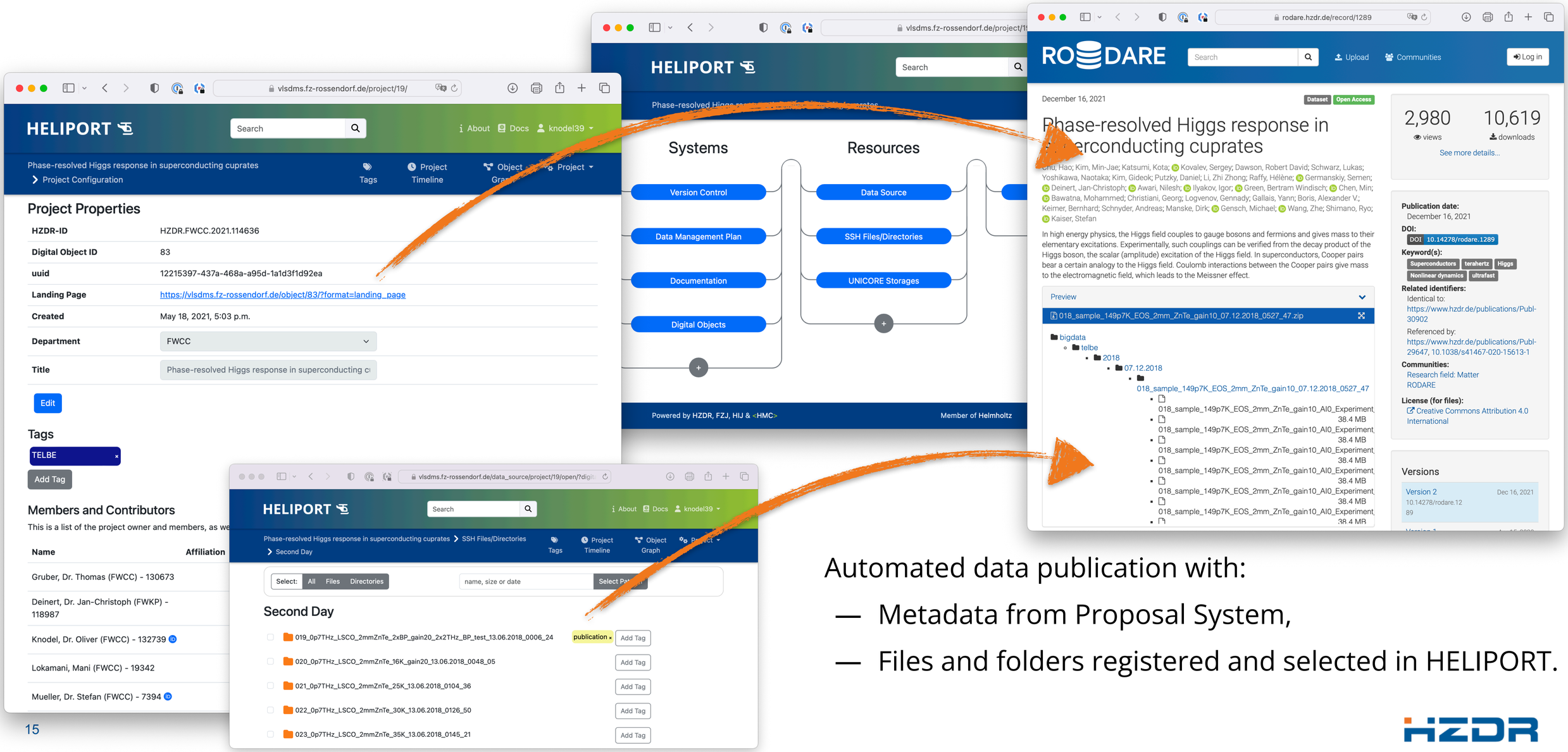
VI. Integration in an Overall Data Publication Workflow



Automated data publication with:

— Metadata from Proposal System,

VI. Integration in an Overall Data Publication Workflow

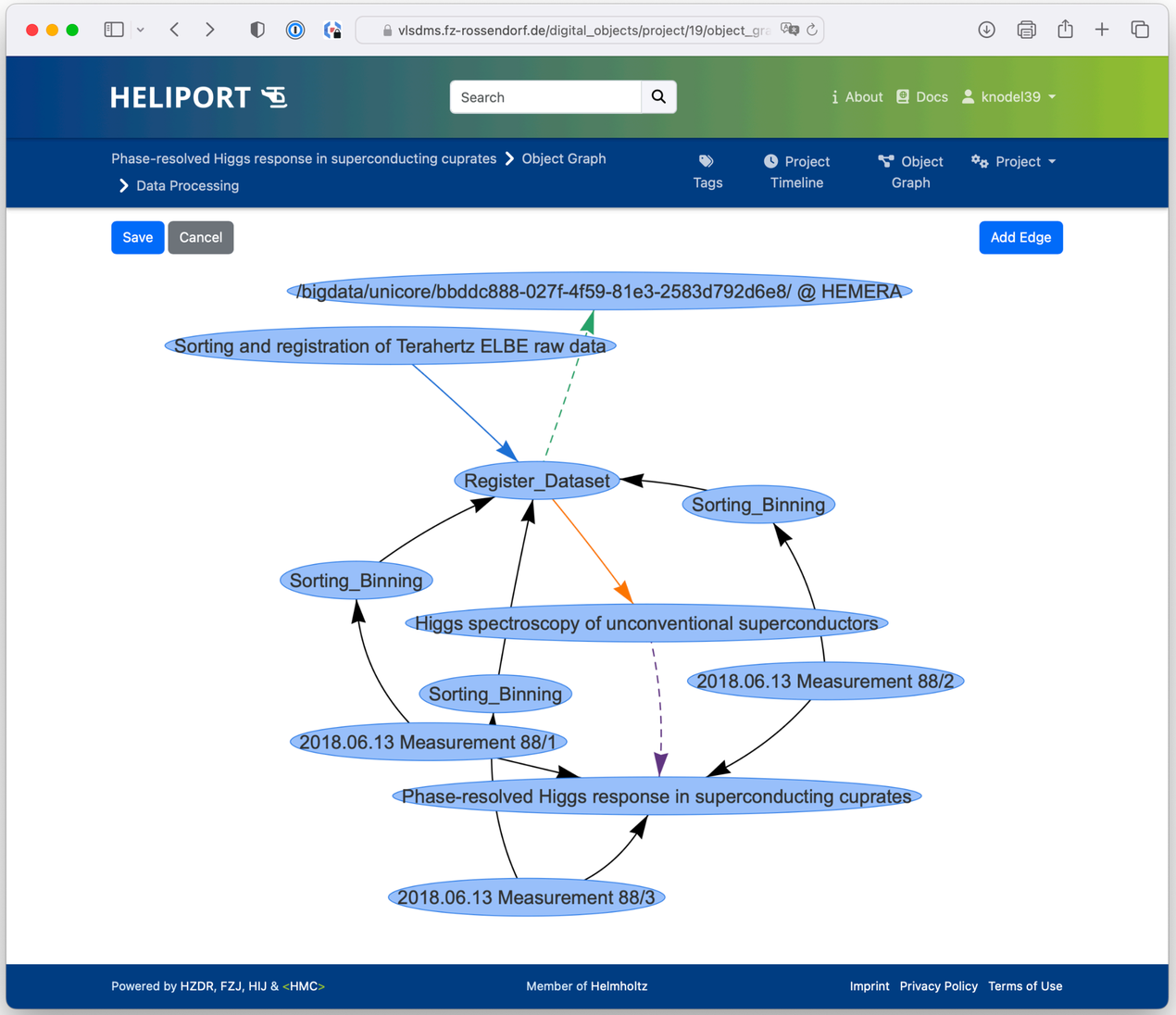


Automated data publication with:

- Metadata from Proposal System,
- Files and folders registered and selected in HELIPORT.

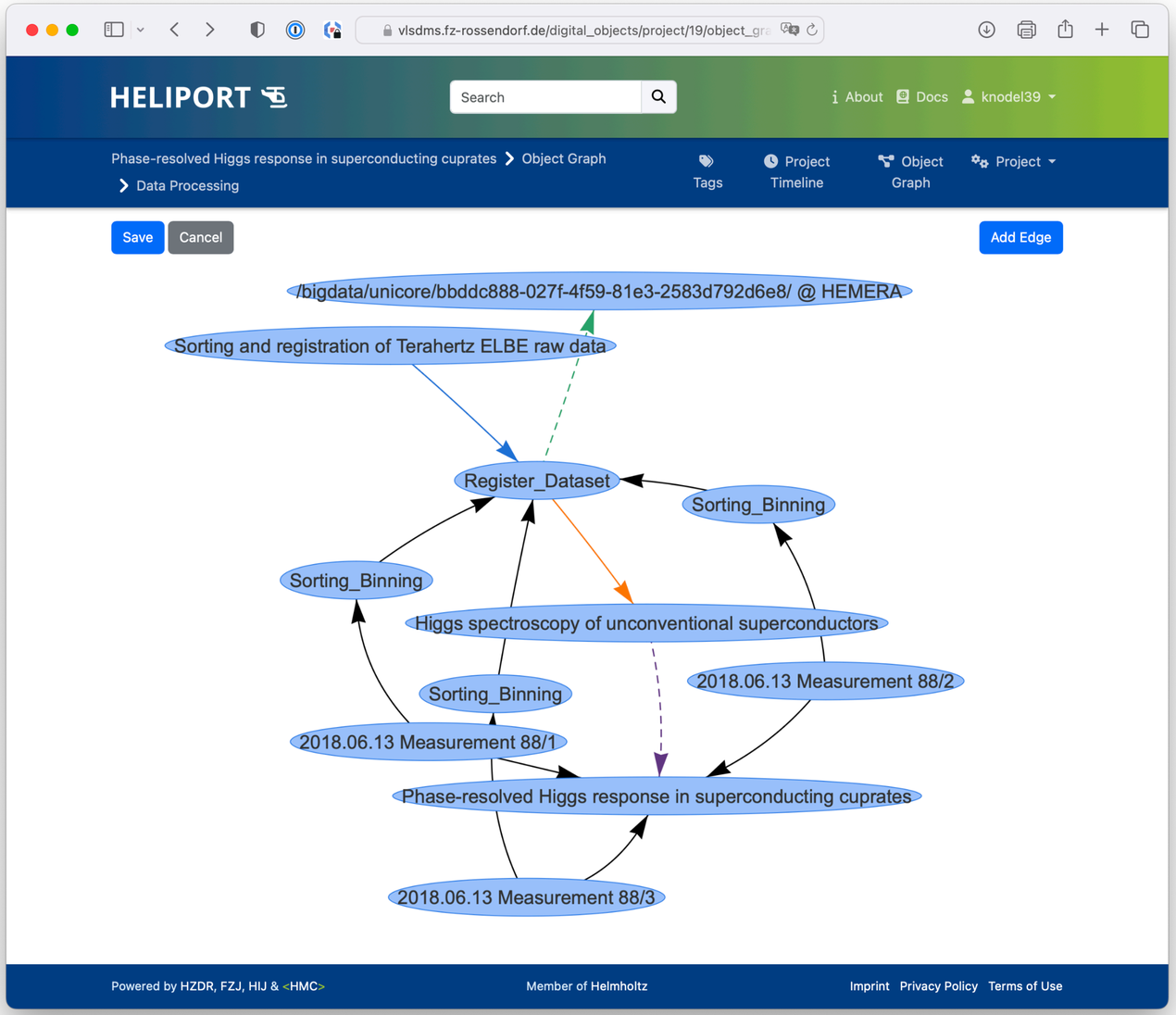
VII. Relations Between Digital Objects and

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



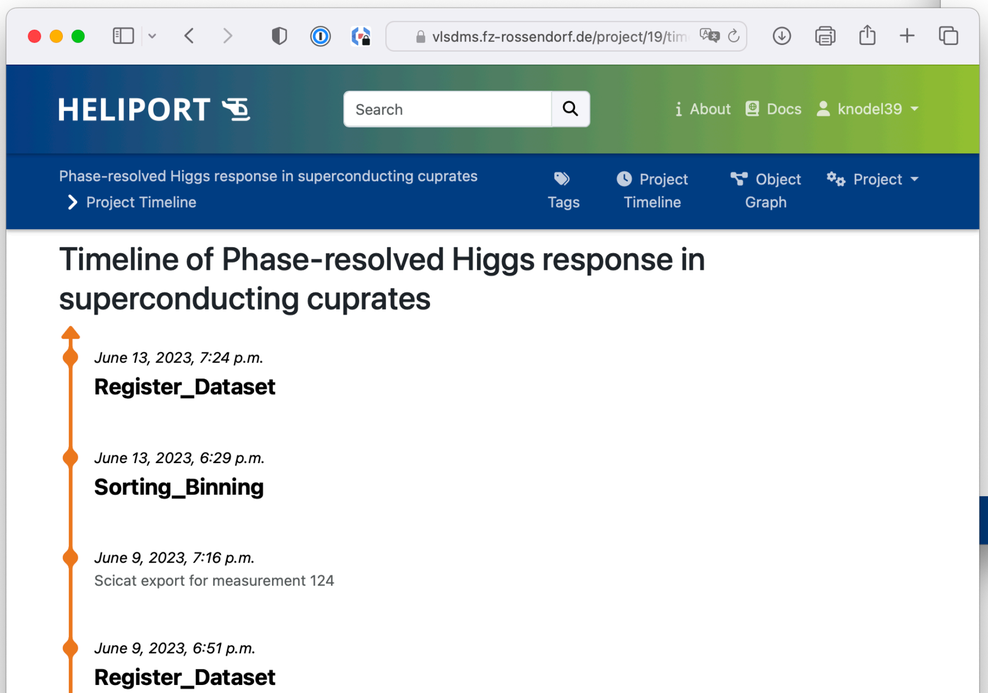
VII. Relations Between Digital Objects and

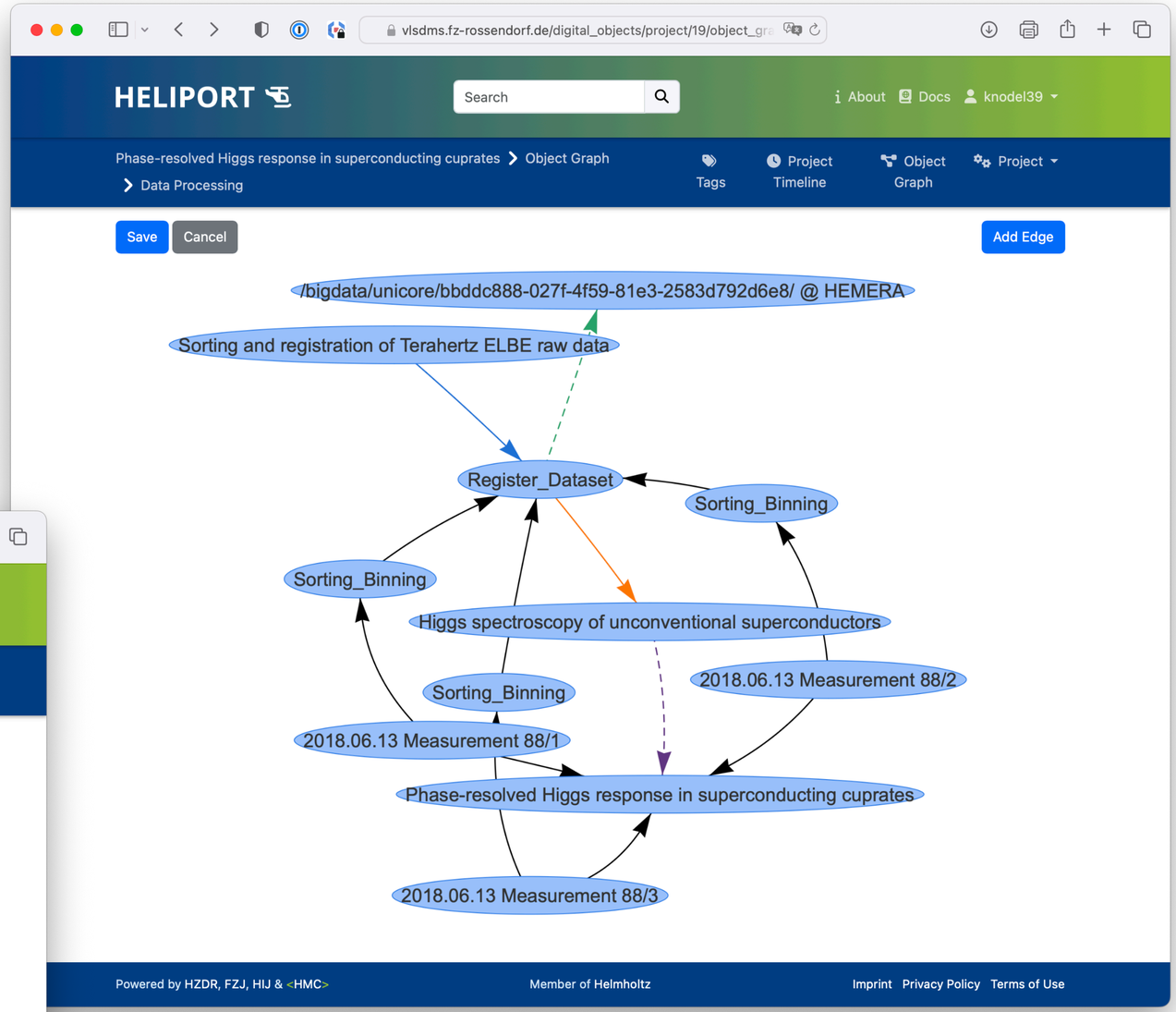
- Relations between digital objects are visualised to provide a top-level view on the project.
- The relationships between simulation (surrogate model) and experiment can also be illustrated.



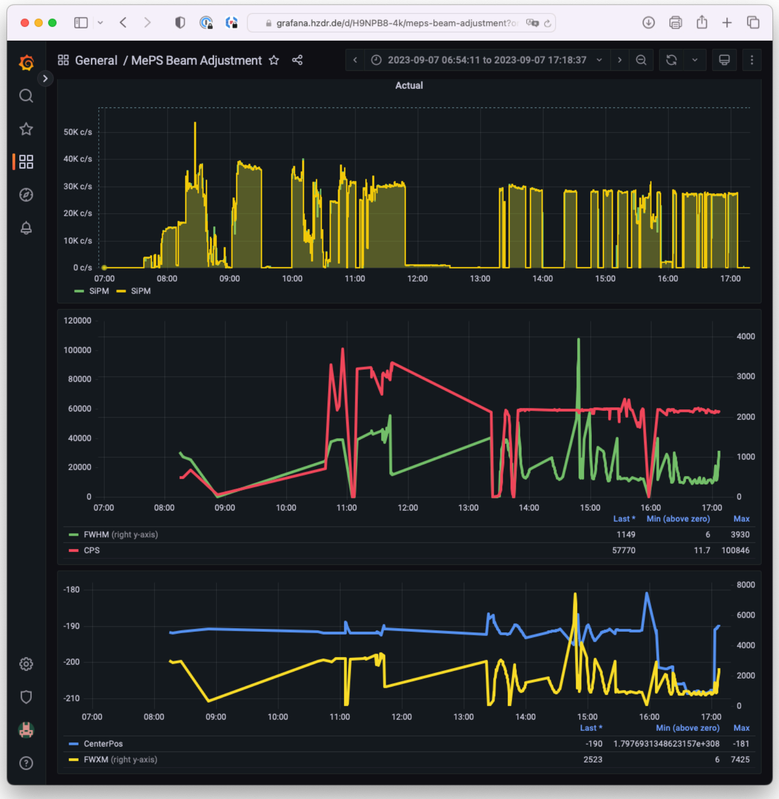
VII. Relations Between Digital Objects and

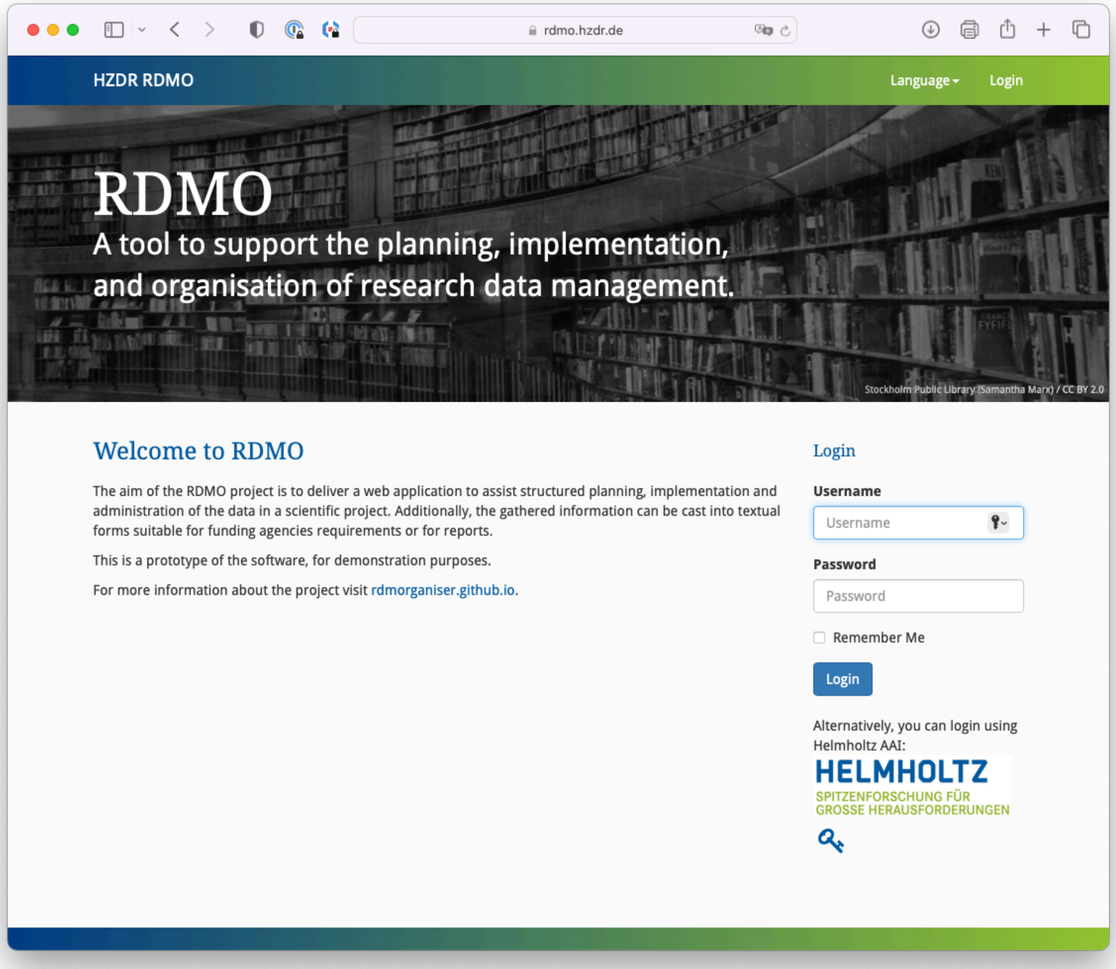
- Relations between digital objects are visualised to provide a top-level view on the project.
- The relationships between simulation (surrogate model) and experiment can also be illustrated.
- The versioning of an experiment is an essential extension, and first approaches via a timeline are being evaluated.





Additional Systems/Services: Grafana & RDMO

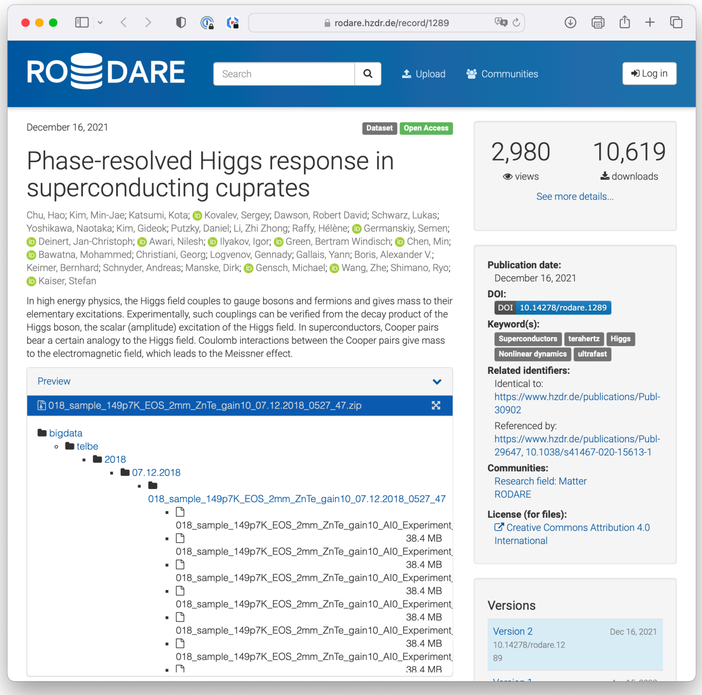




Next Steps: Integration of an Additional Metadata Catalogue

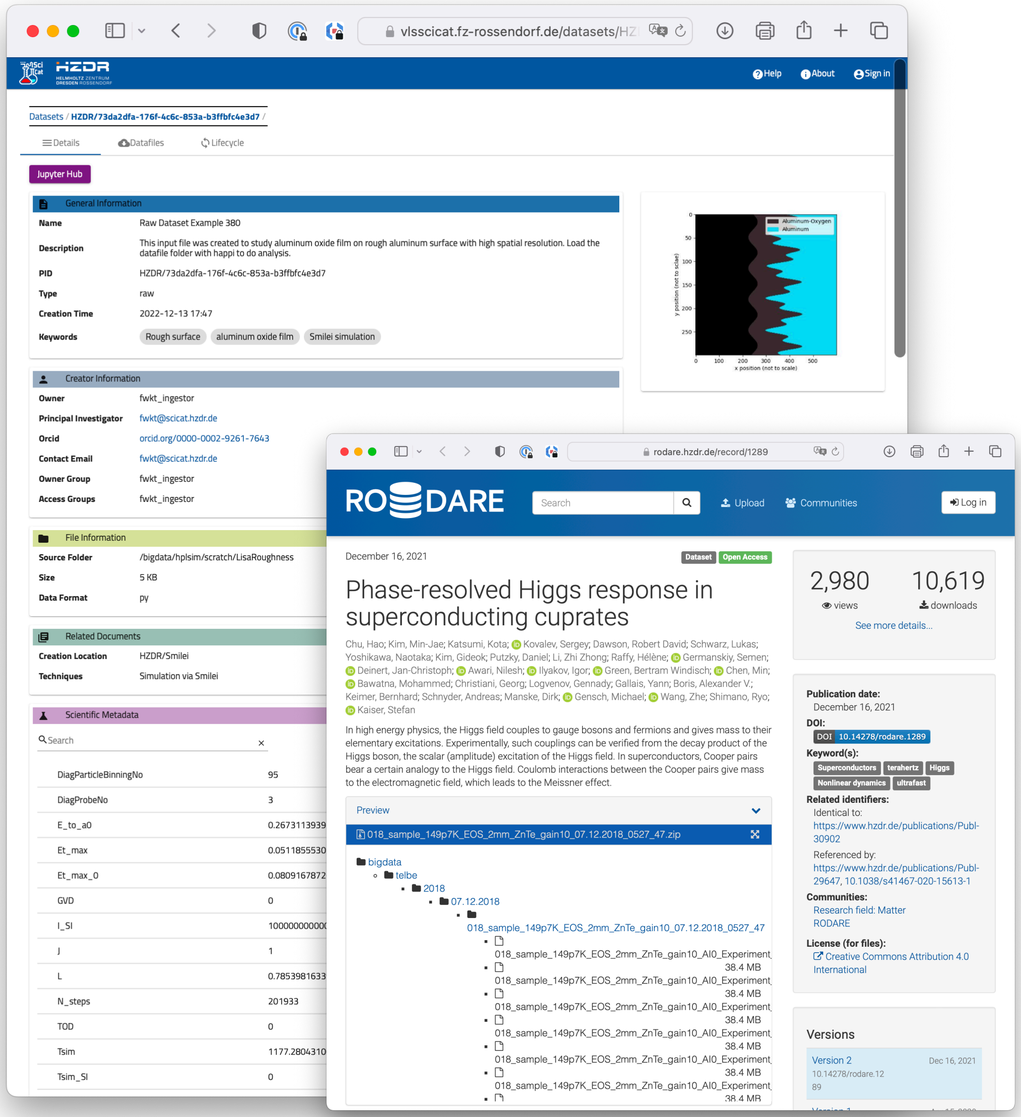
- With our data publication repository RODARE (based on InvenioRDM) we provide a platform for publishing datasets.
- In RODARE, bibliographic metadata is based on DataCite.

RODARE
ROSSENDORF DATA REPOSITORY



Next Steps: Integration of an Additional Metadata Catalogue

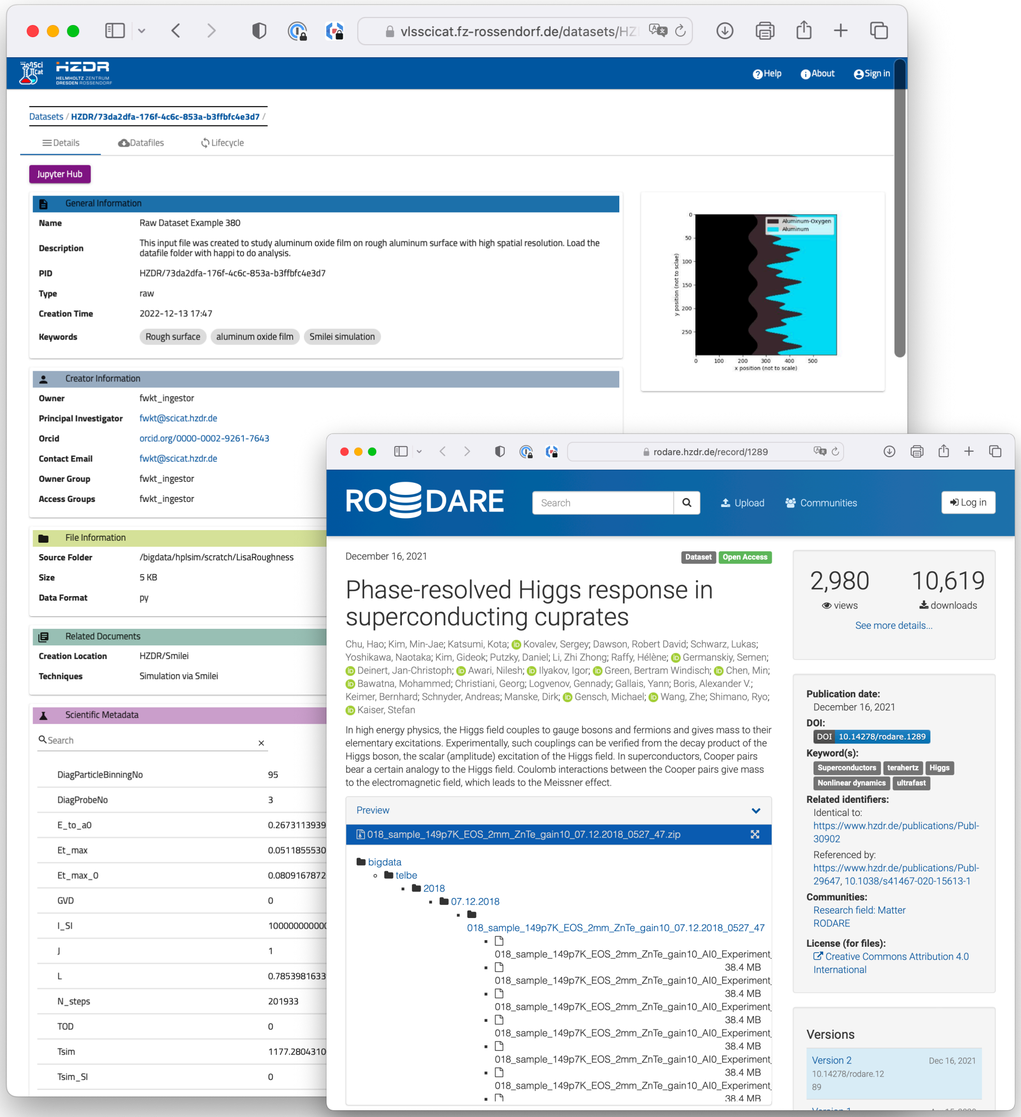
- With our data publication repository RODARE (based on InvenioRDM) we provide a platform for publishing datasets.
- In RODARE, bibliographic metadata is based on DataCite.
- For additional scientific metadata we use SciCat and reference the specific datasets in RODARE.



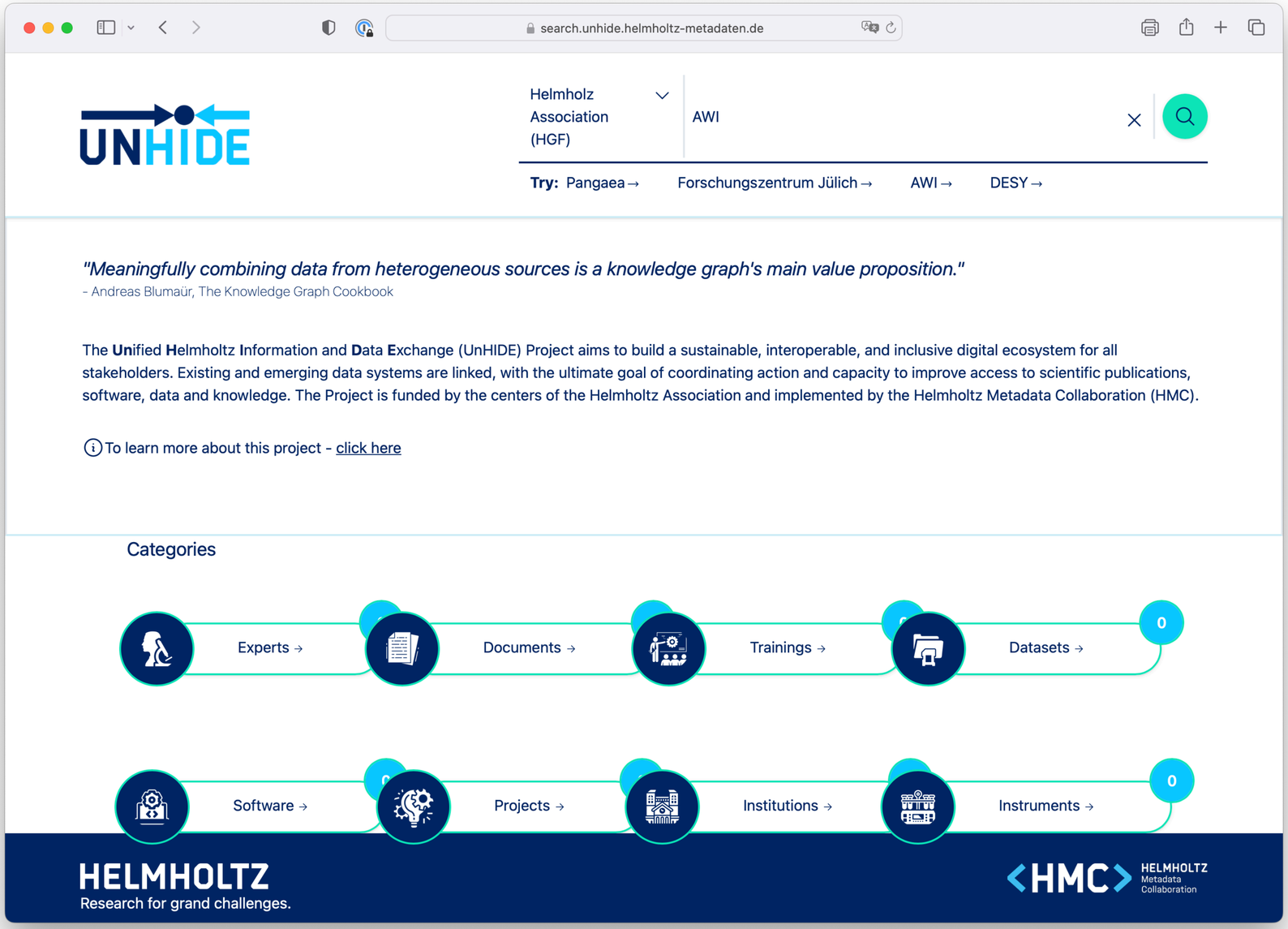
ROSSENDORF DATA REPOSITORY

Next Steps: Integration of an Additional Metadata Catalogue

- With our data publication repository RODARE (based on InvenioRDM) we provide a platform for publishing datasets.
- In RODARE, bibliographic metadata is based on DataCite.
- For additional scientific metadata we use SciCat and reference the specific datasets in RODARE.
- Within the DAPHNE4NFDI project, we want to establish such an approach for the photon and neutron community.



Future: Providing Metadata for UnHIDE (Unified Helmholtz Information and Data Exchange)



HMC HELMHOLTZ Metadata Collaboration

HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

HI JENA Helmholtz Institute Jena

JÜLICH FORSCHUNGSZENTRUM

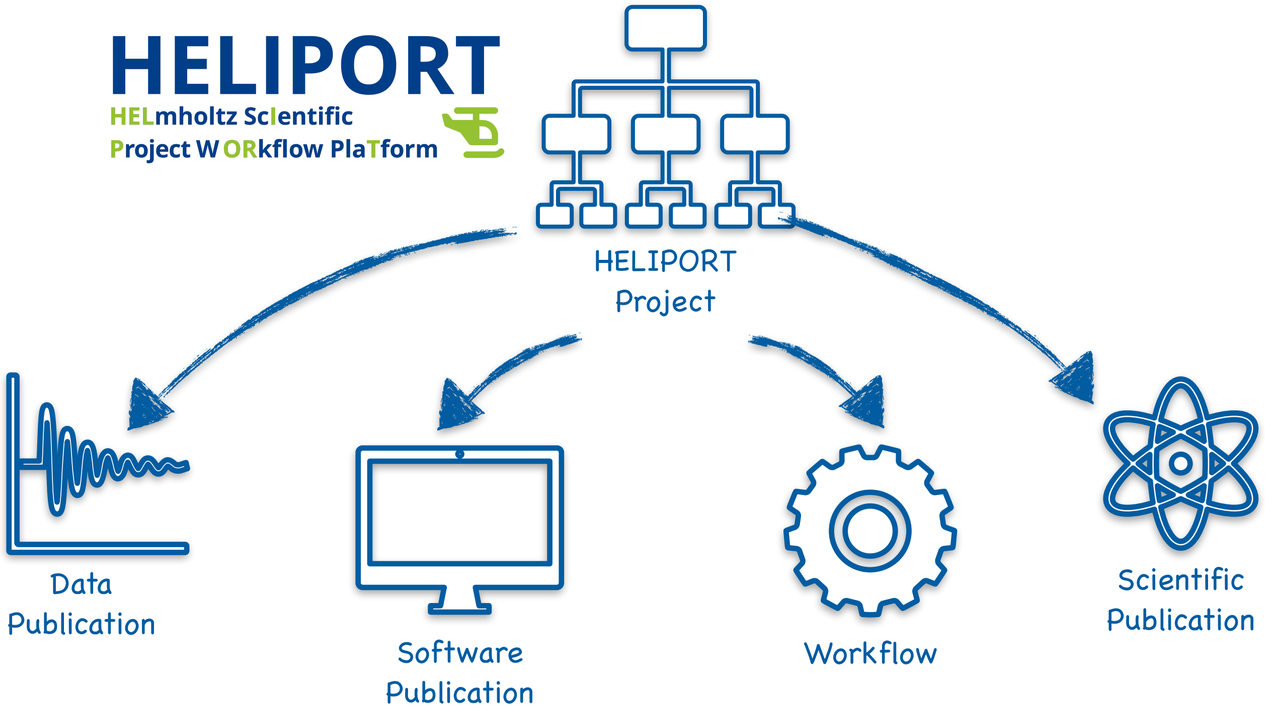
GFZ Helmholtz-Zentrum POTSDAM

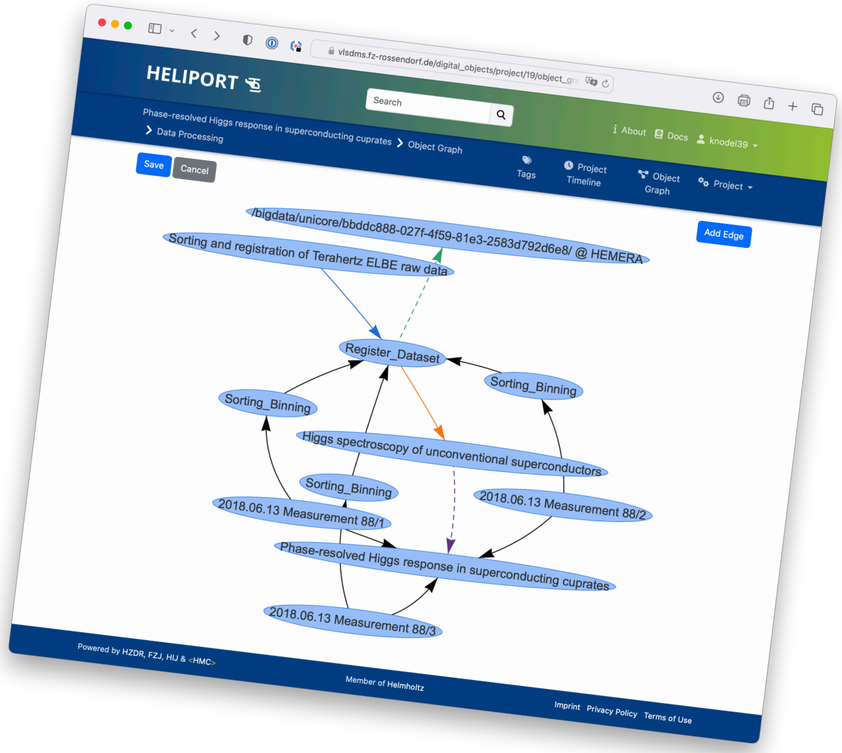
GSI

HZDR

Conclusions

— HELIPORT describes and collects all metadata from all services and systems involved in an scientific experiment.

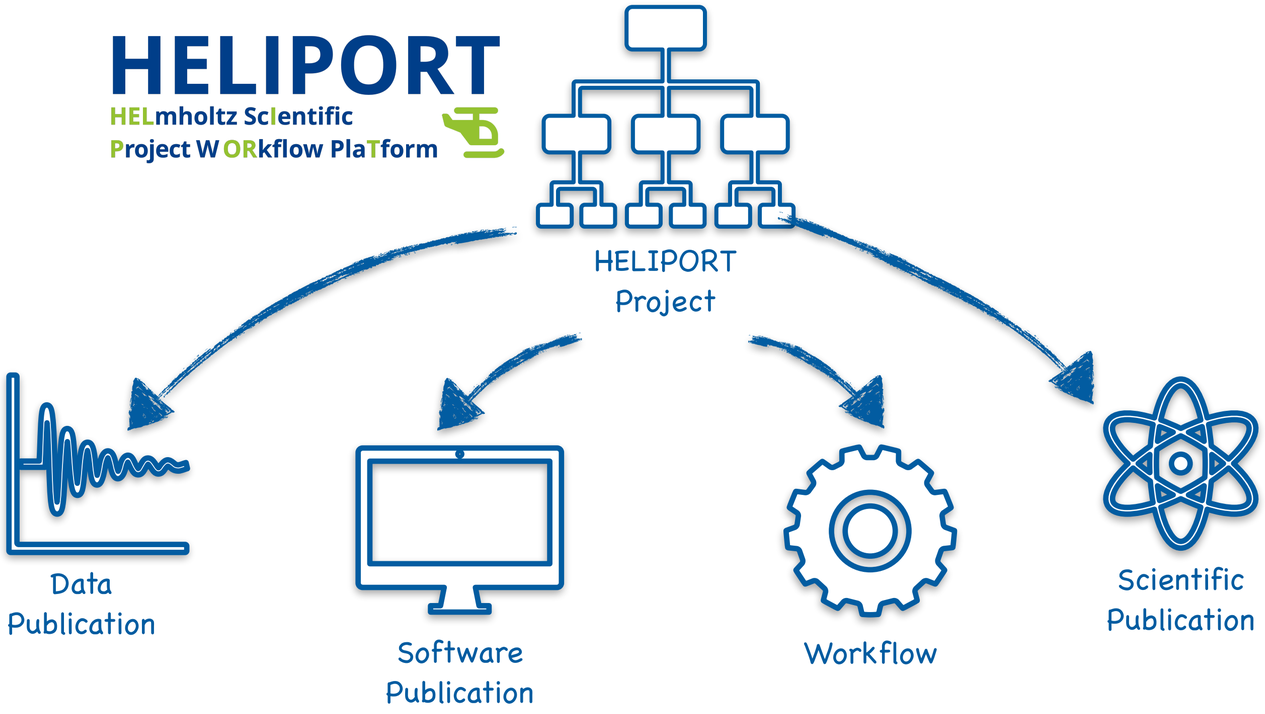


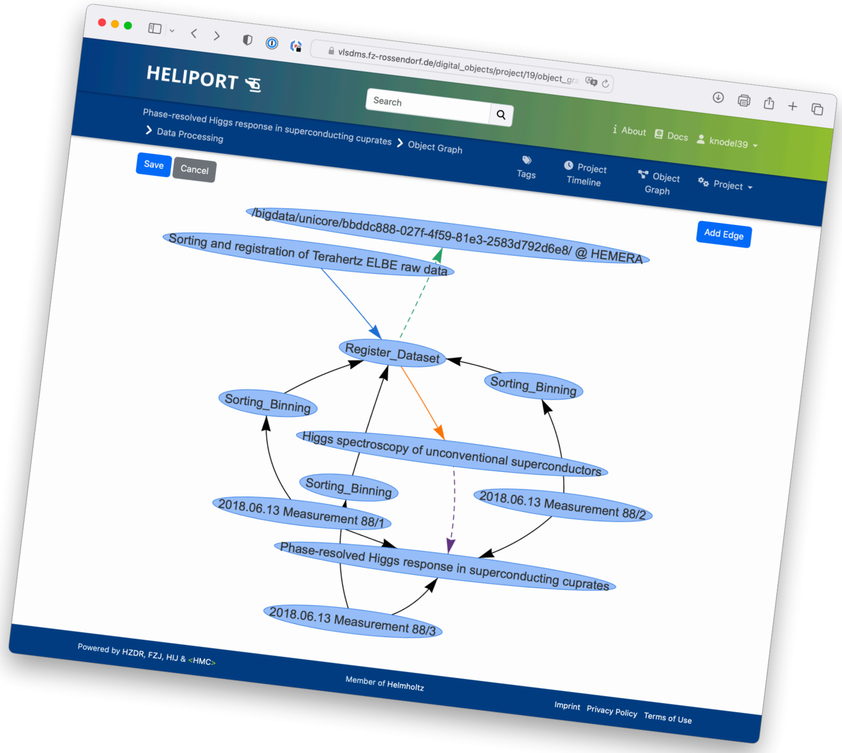


Conclusions

— 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.



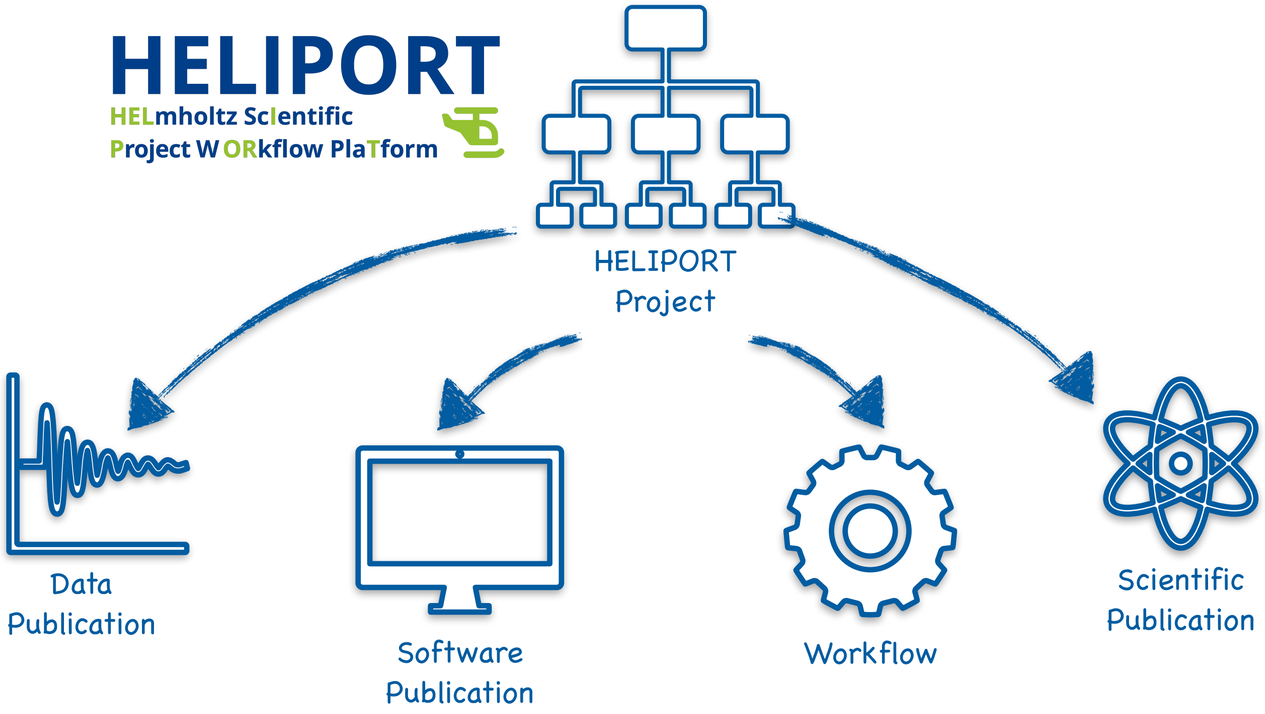


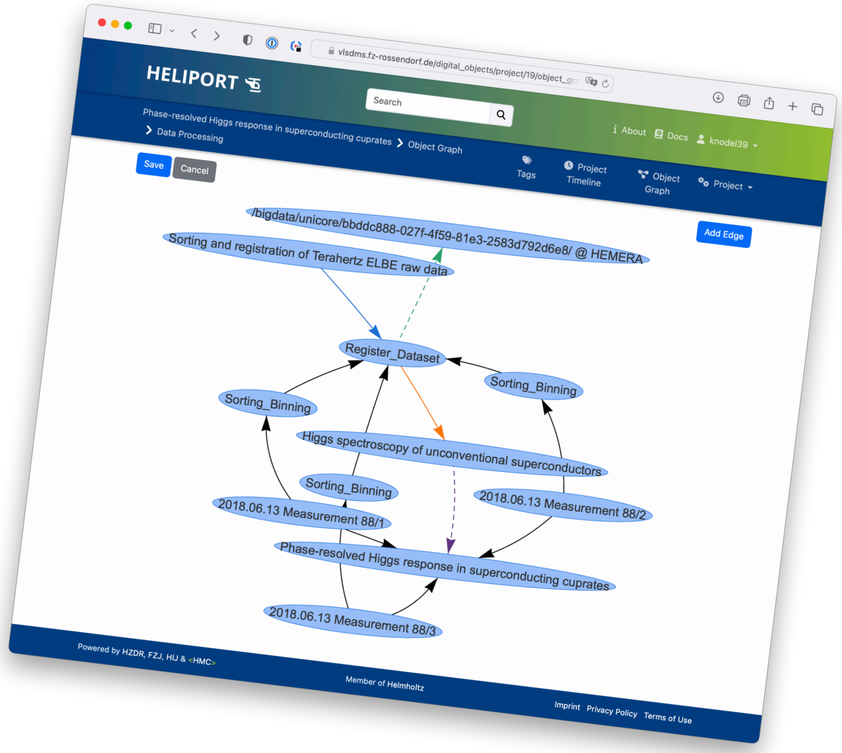
Conclusions

— 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.

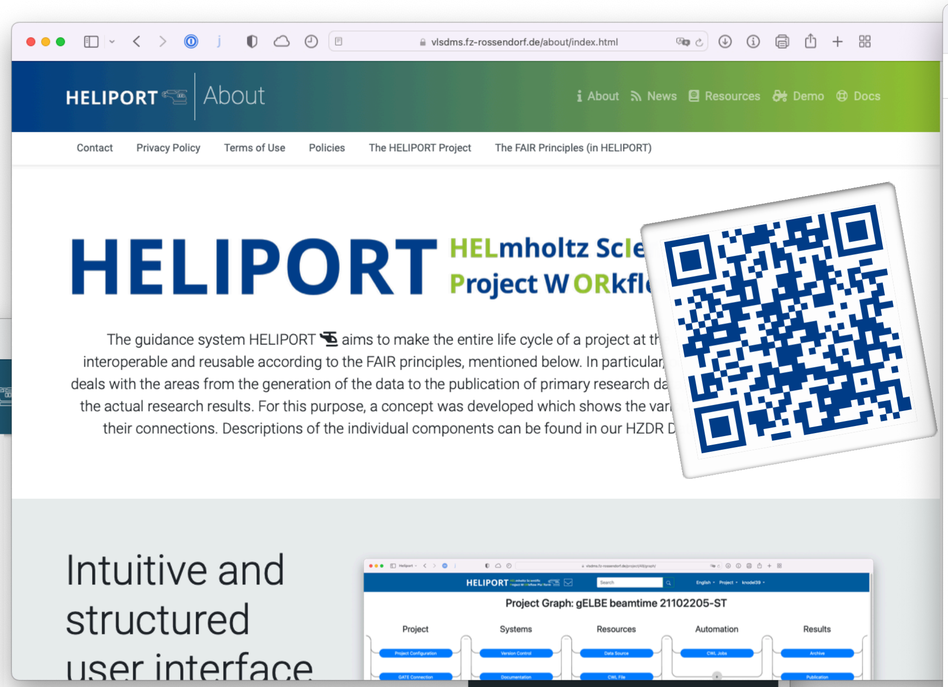


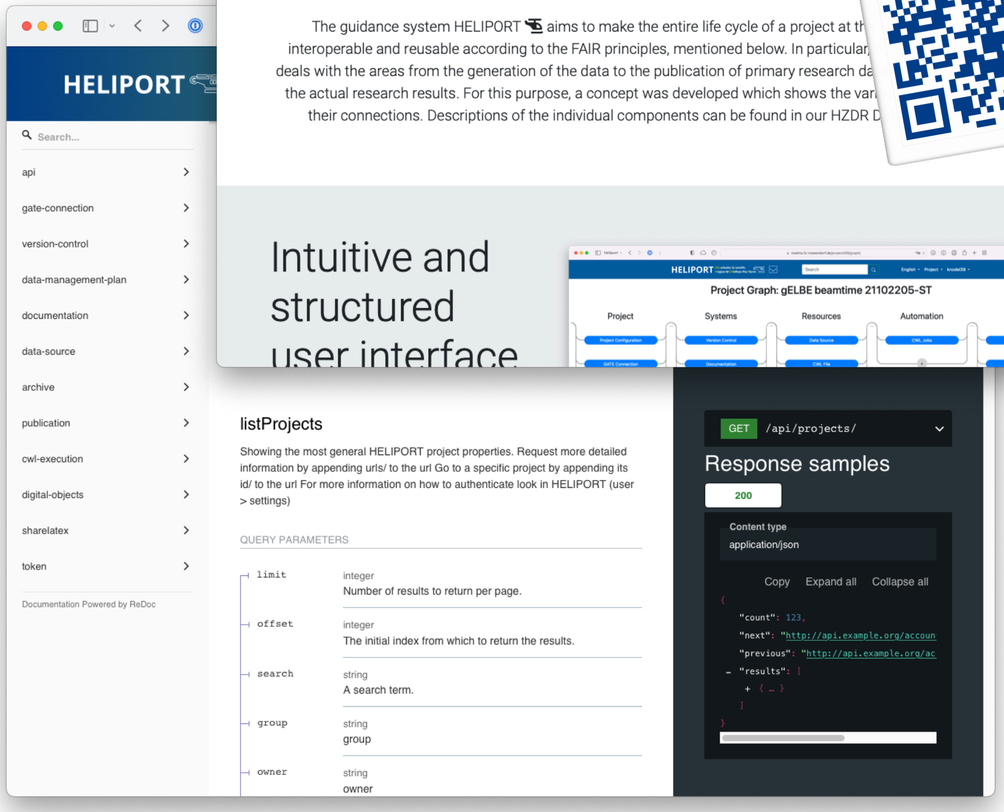


Resources

Website: heliport.hzdr.de

Repository: codebase.helmholtz.cloud/heliport





API Doc: heliport.hzdr.de/redoc/