

A FAIRly Integrated Scientific Project Lifecycle

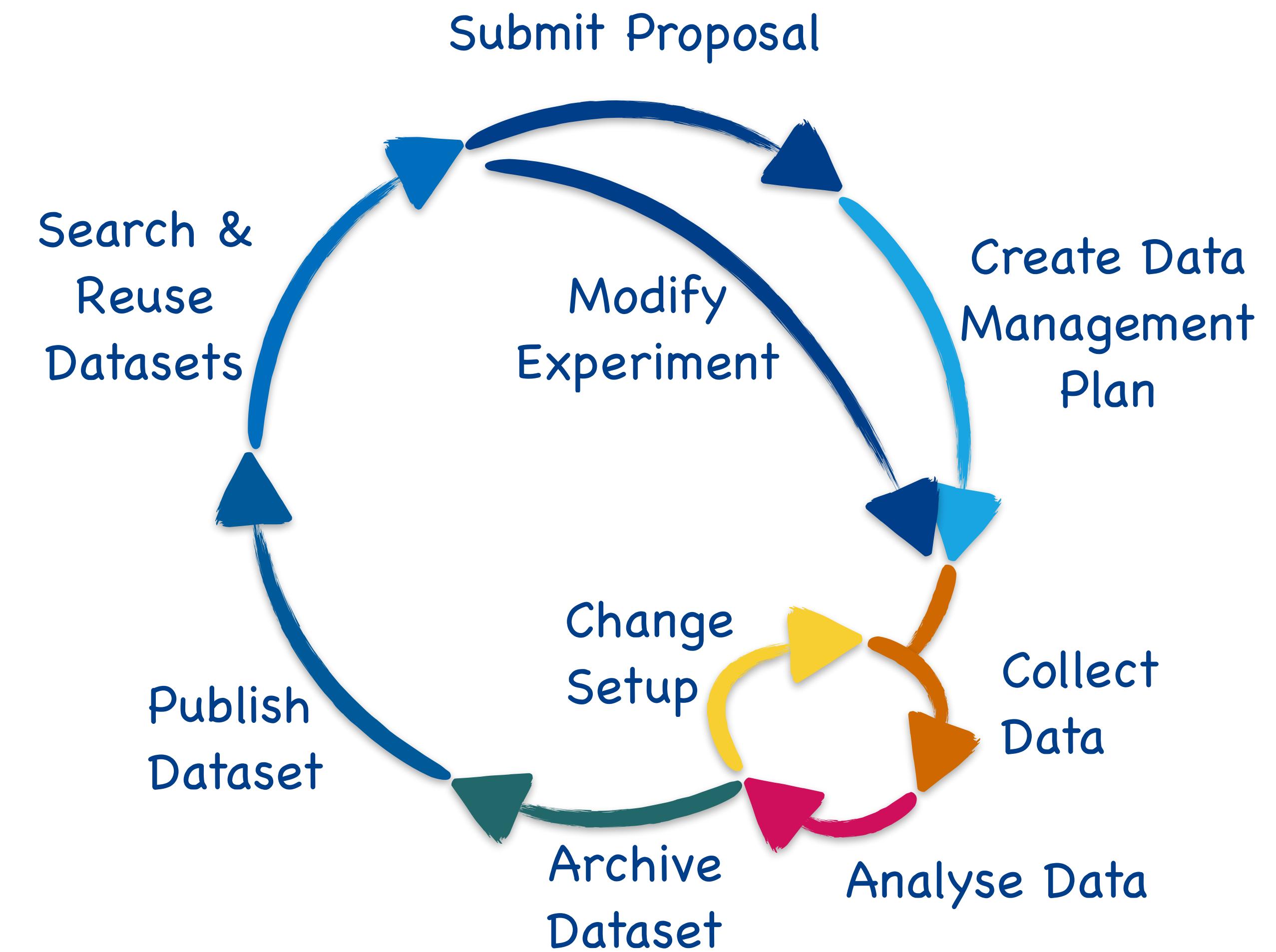


Oliver Knodel, Martin Voigt, Robert Ufer, David Pape, Mani Lokamani, Jeffrey Kelling, Stefan E. Müller, Thomas Gruber, Guido Juckeland, Alexander Kessler, Joachim Hein, Bernd Schuller // contact: o.knodel@hzdr.de



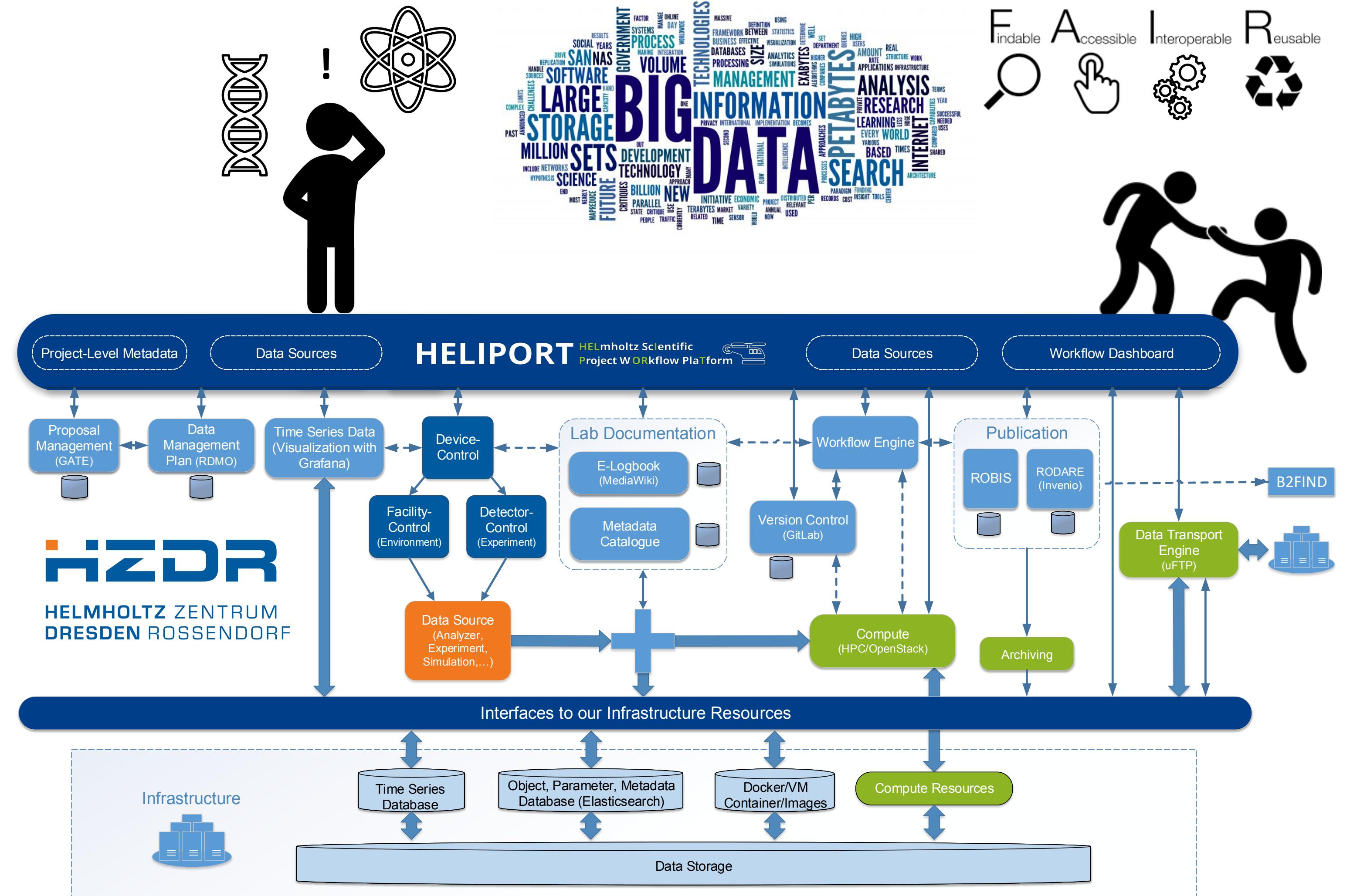
Our Challenge: An End-to-End Digital Data Lifecycle

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

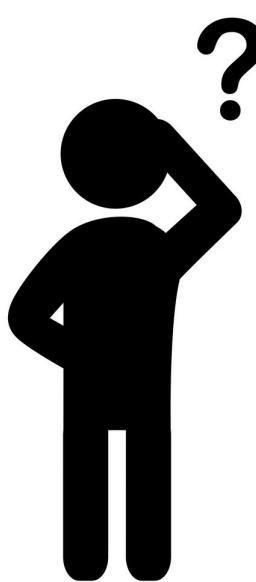


The Requirements and Conditions

- 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:

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

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

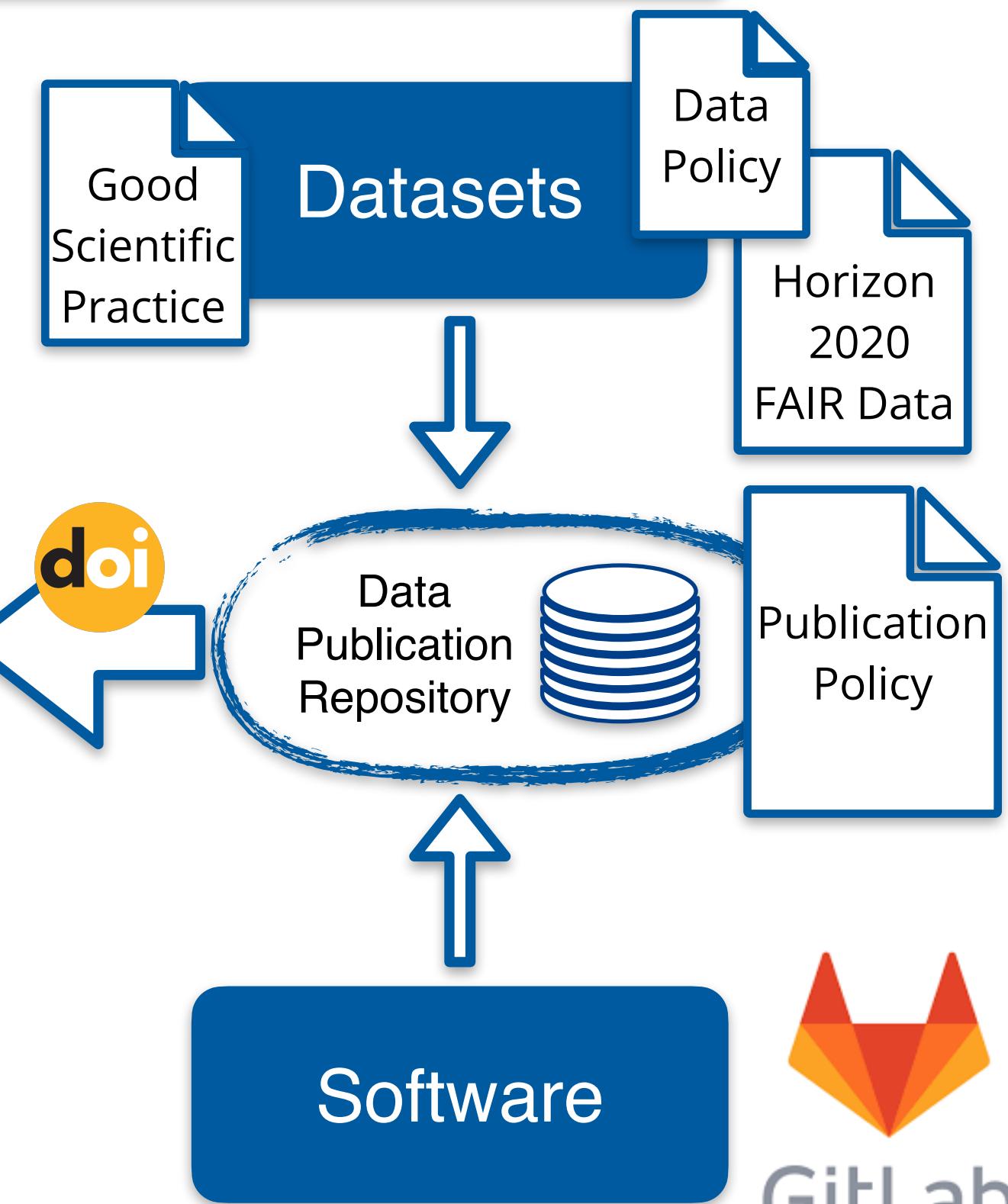


Which datasets or software can be **published** (and how)?

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

And how we can support them?!

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

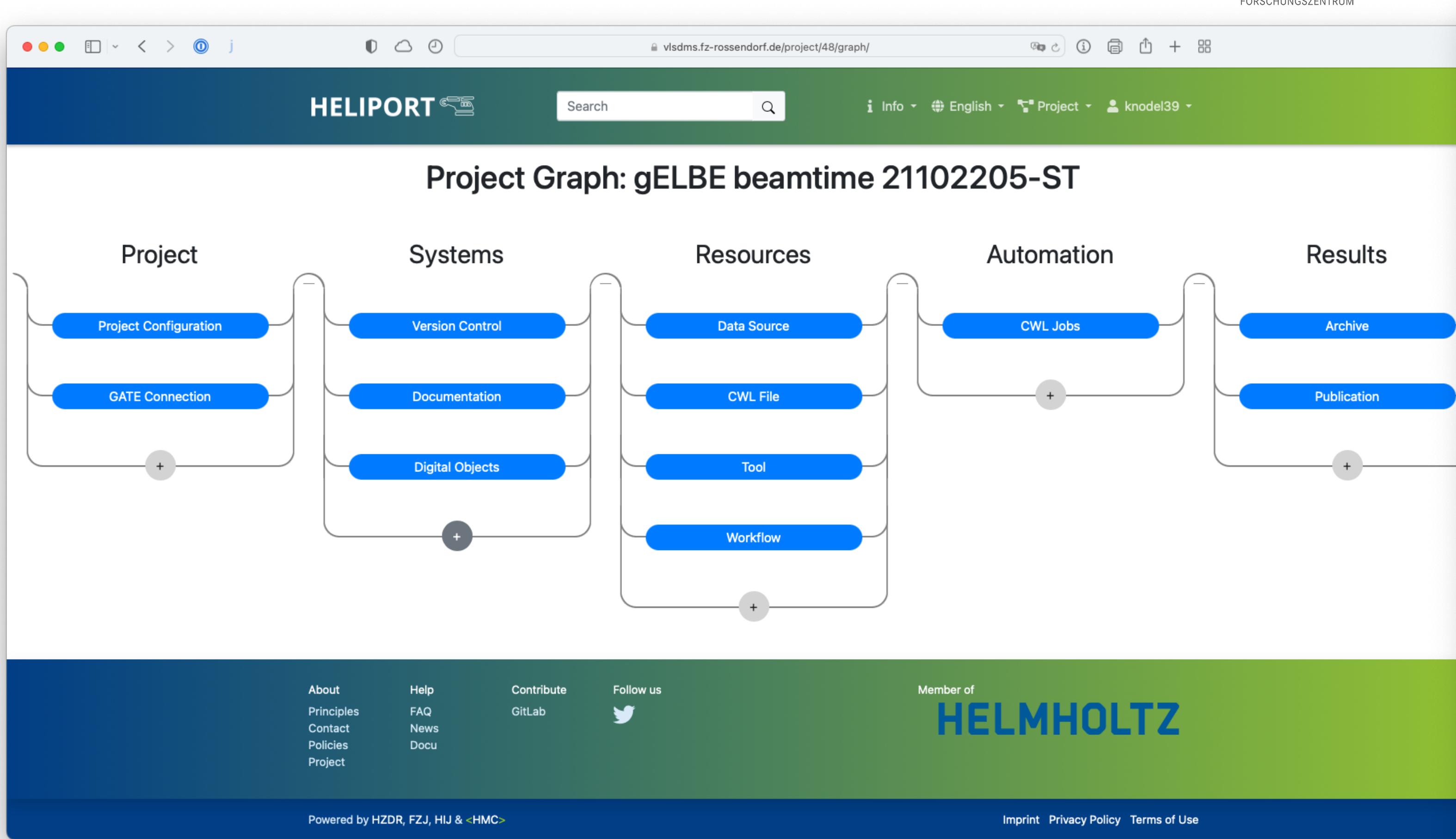


“ The HELIPORT project aims at developing a platform which accommodates the **complete life cycle** of a scientific project and links all corresponding programs, systems and workflows to create a more **FAIR** and comprehensible project description.

Project Members:

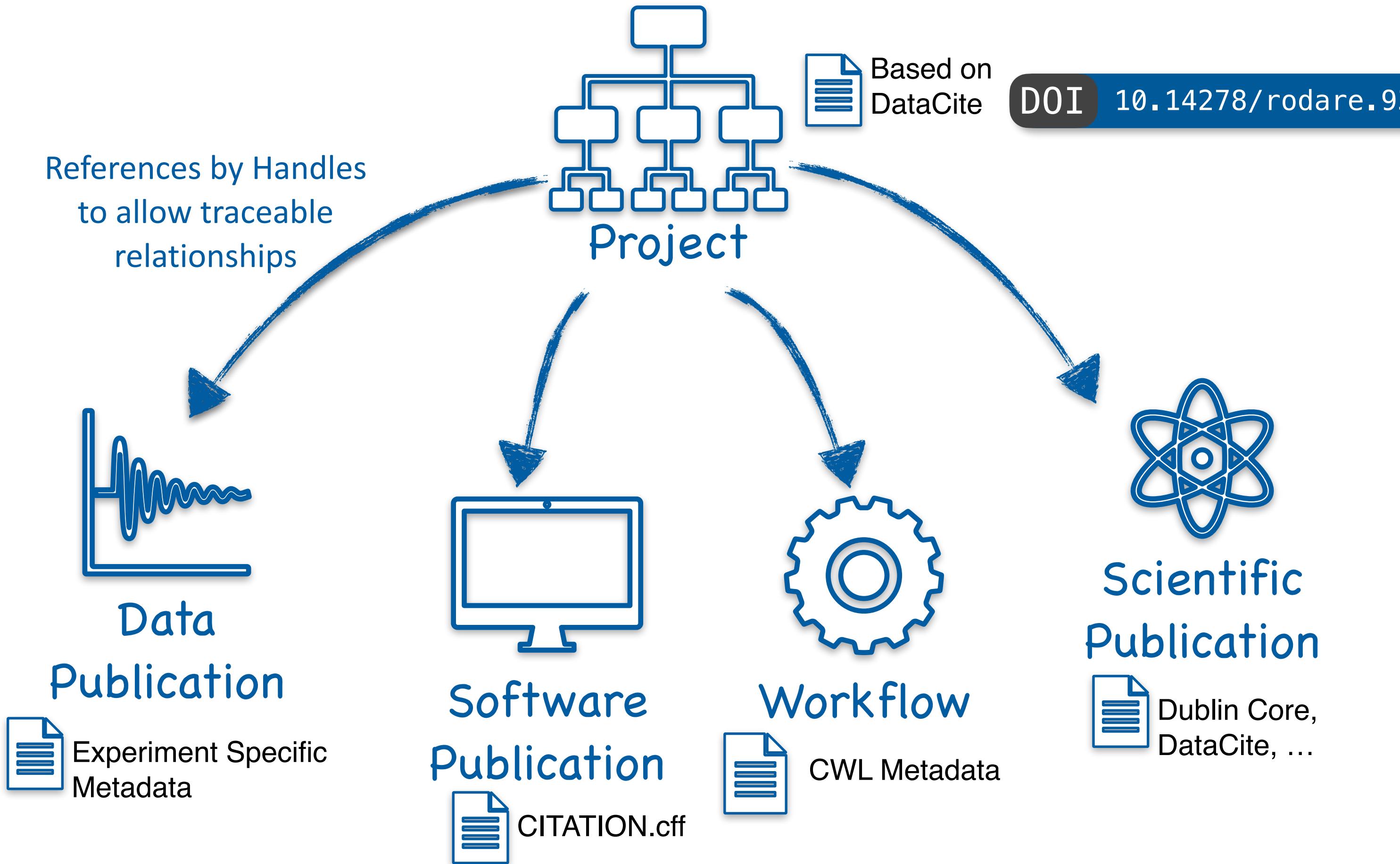


Founded by:



```
{
  "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://dd",
      "rdfs:label": "Test"
    }
  ],
  "heliport:has_DataManagementPlan": [
    {
      "heliport:data_management_plan_id": 6,
      "datacite:uri": "https://ddd",
      "datacite:hasDescription": "dddd"
    }
  ],
  "heliport:has_Documentation": [
    {
      "heliport:documentation_id": 7,
      "datacite:uri": "https://ddd",
      "heliport:documentation_system": "MediaWiki",
      "datacite:hasDescription": "dddd"
    }
  ],
  "heliport:has_DataSource": [
    {
      "heliport:data_source_id": 11,
      "datacite:uri": "http://ddd",
      "heliport:use_computer": null,
      "rdfs:label": "ddd",
      "datacite:hasDescription": ""
    }
  ]
}
```

Heliport Metadata Ecosystem



Our Concept

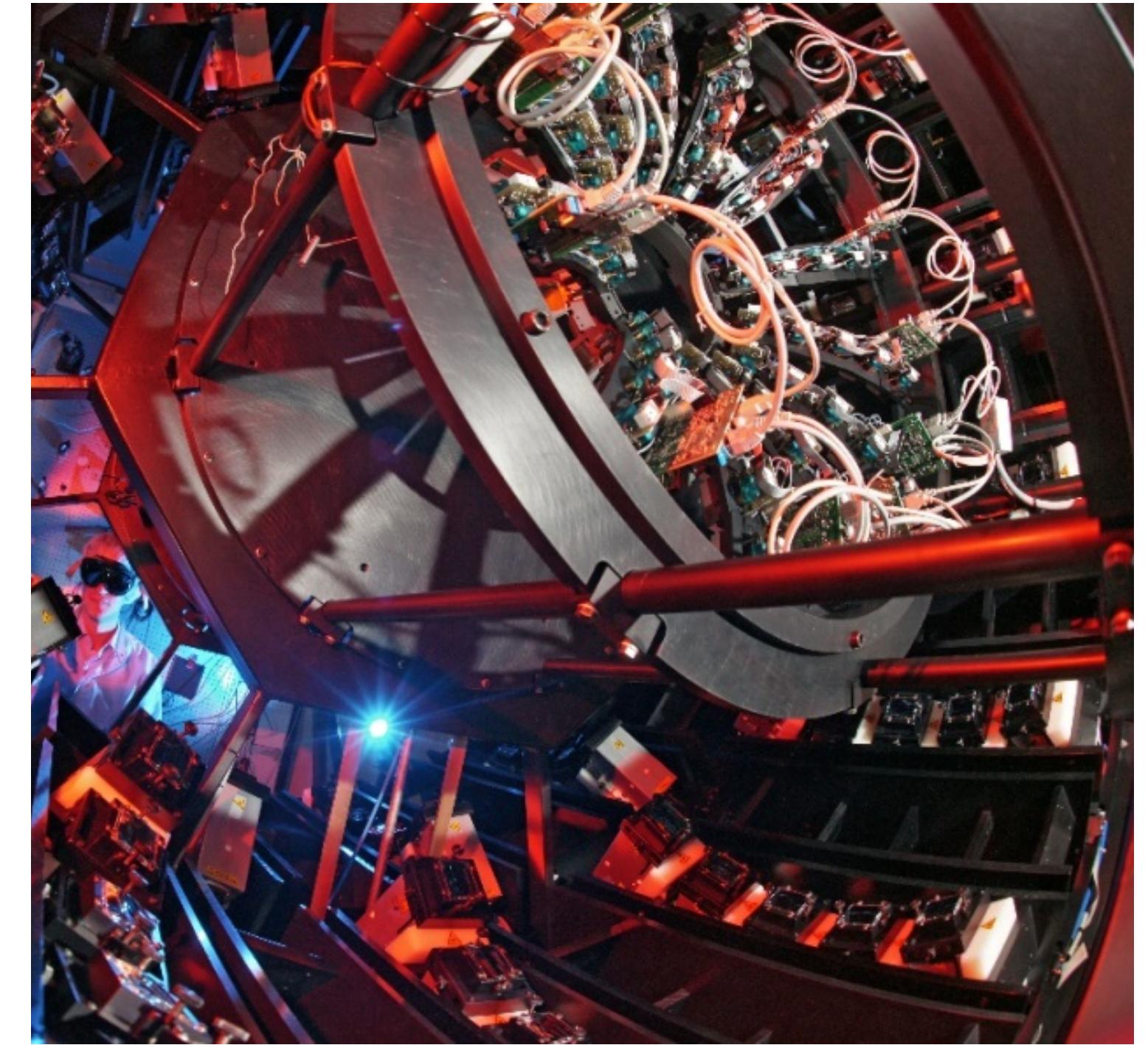
- In all stages of an experiment Heliport combines information about involved services with PIDs.
- Metadata (stored *near* the PID) is used to transfer information between different systems and a documentation of the project-level workflow is possible.
- The project-level metadata is distributed over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

Example I: POLARIS



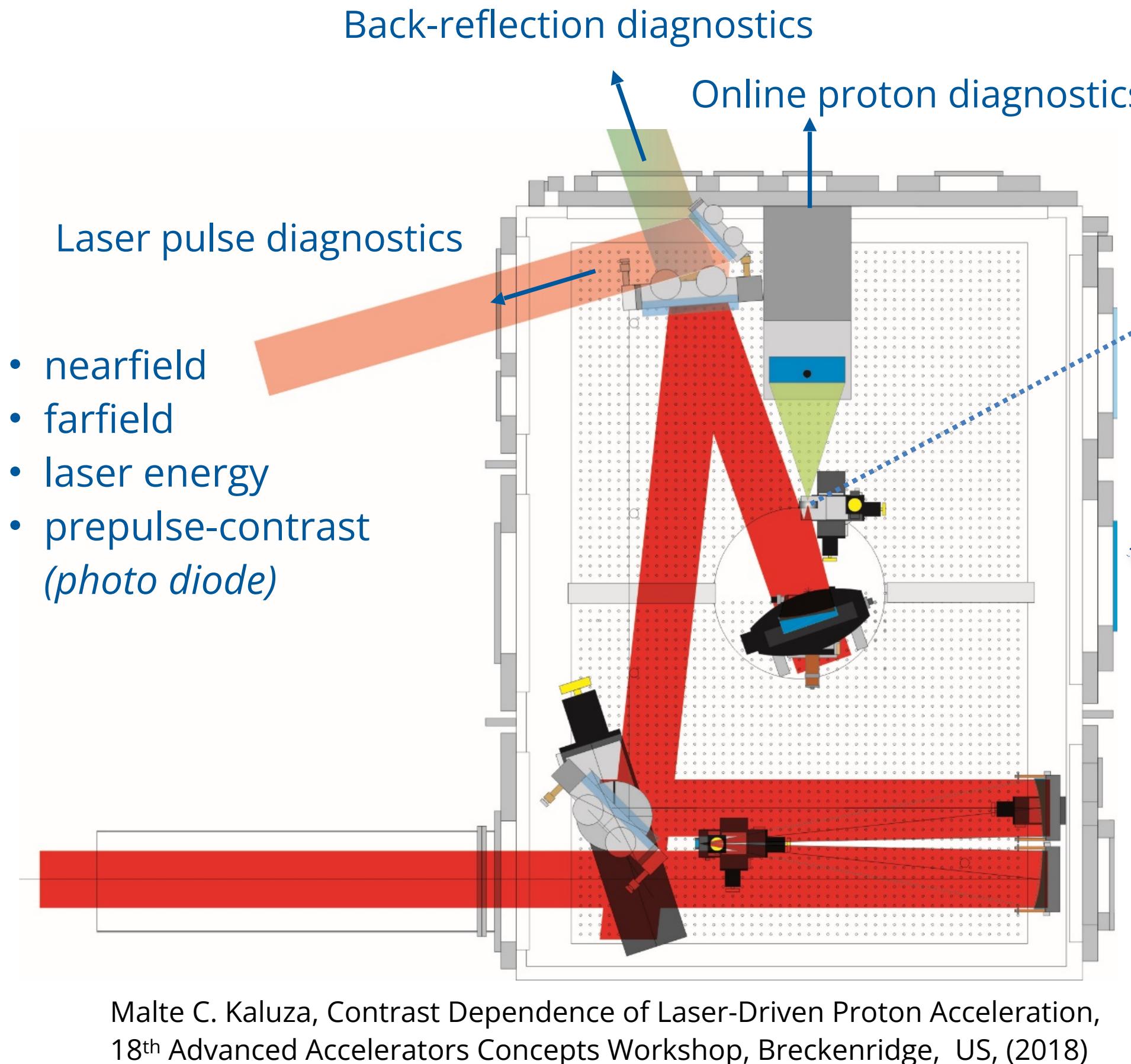
Example: The POLARIS Experiment @ HI Jena

- The **P**etawatt **O**ptical **L**aser **A**mplifier for **RI**NTENSIVE **E**XPERIMENT**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



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

Project Graph: PaN Research Project

```
GNU nano File: ...14.data
[Results]
q-Factor = 0,150082
Focus Area[0m^2] = 10,055498

[Parameters]
Calibration = 0,130000
Energy[J] = 10,000000
Size of region[px] = 2
#of Areas = 1
Target Point X = 0,000000
Target Point Y = 0,000000
Width Method = "FWHM"

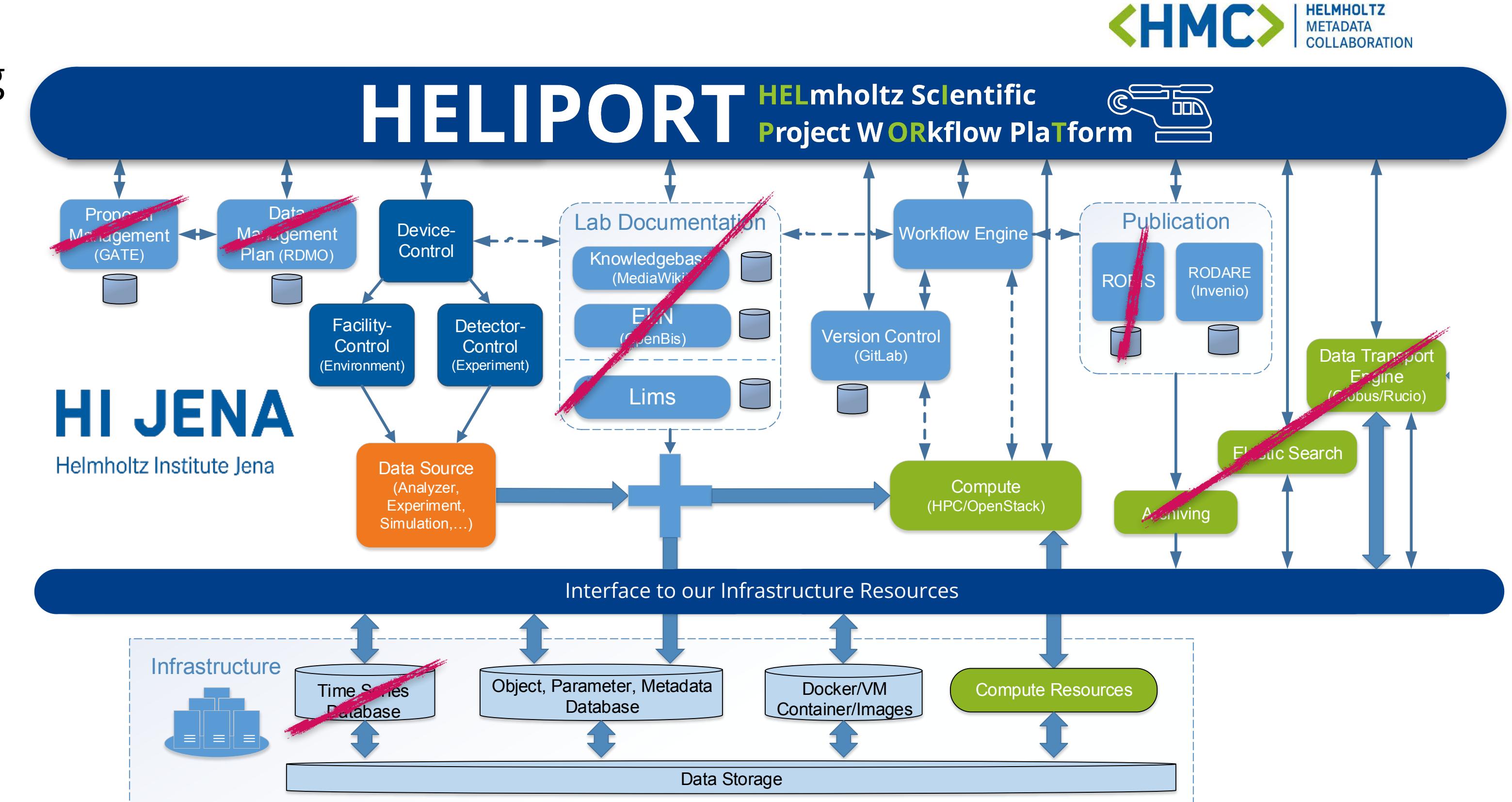
[Parameters.ROI]
GlobalRectangle.0 = 186
GlobalRectangle.1 = 346
GlobalRectangle.2 = 942
GlobalRectangle.3 = 1002

[Parameters.ROIContour.0]
ID = 0
Type = 4
Coordinate.0 = 186
Coordinate.1 = 346
Coordinate.2 = 942
Coordinate.3 = 1002

[G Get^O Wri^R Rea^Y Pre^K Cut^C Cur P
^X Exi^J Jus^W Whe^V Nex^U UnC^T To Sp]
```

HELIPORT and the Infrastructure at Helmholtz Institute Jena

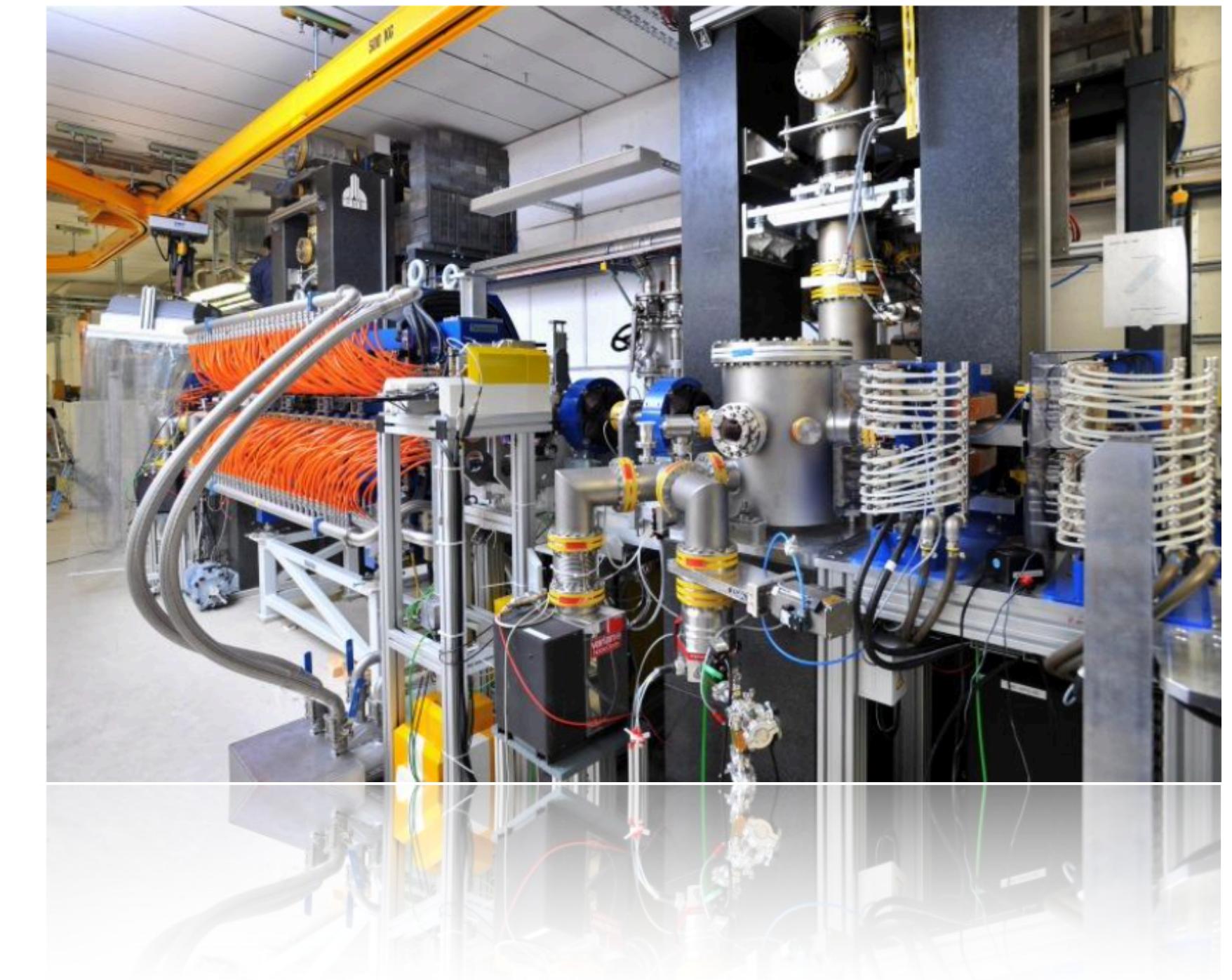
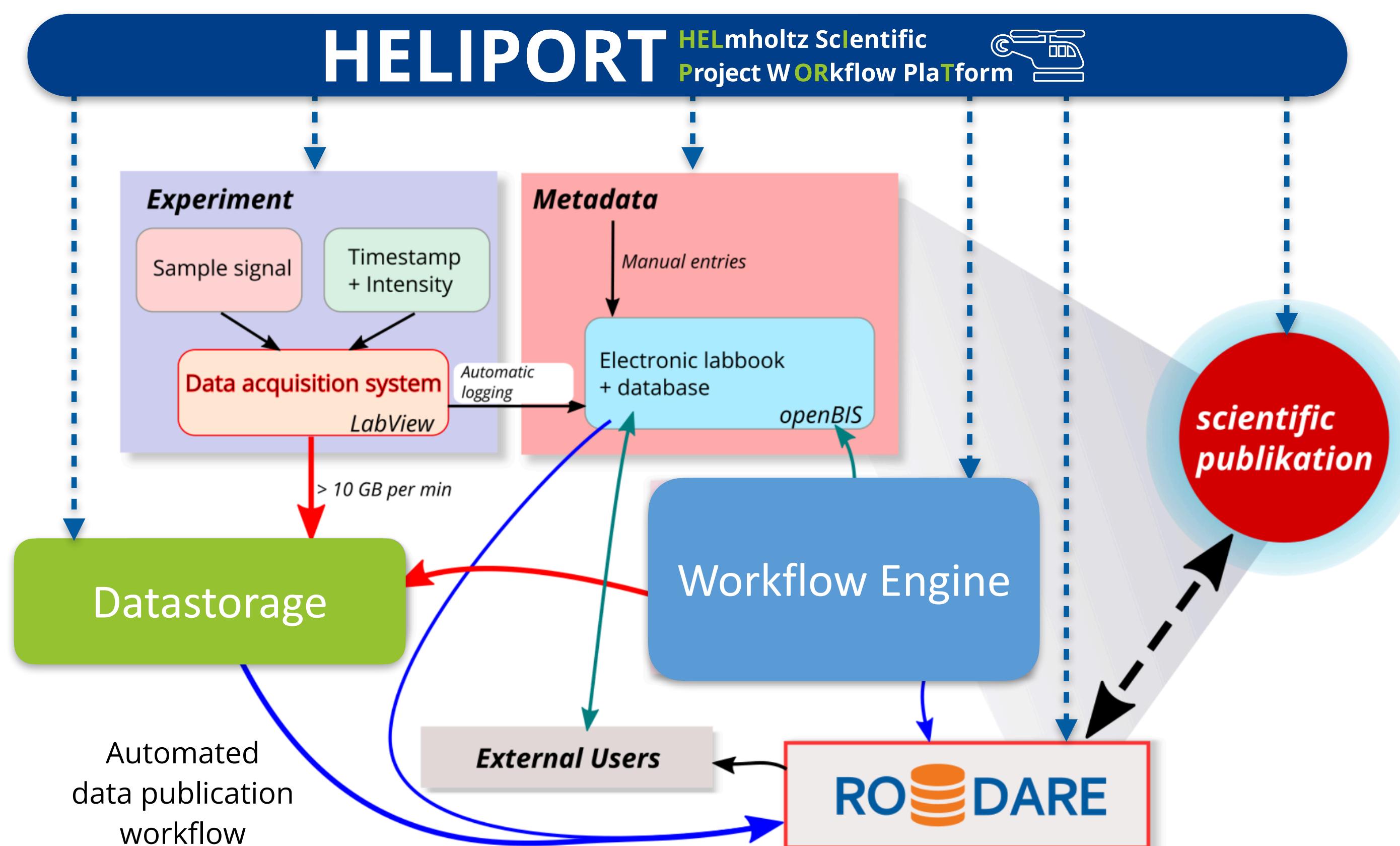
- HELIPORT was initially developed to provide access to the computing Infrastructure at the HZDR.
- The infrastructure at Jena is different, but the important systems are available:
 - HPC Infrastructure and Data Management*,
 - Data Acquisition, and
 - Git Repositories
- And others are in development: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



Example II: TELBE

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.



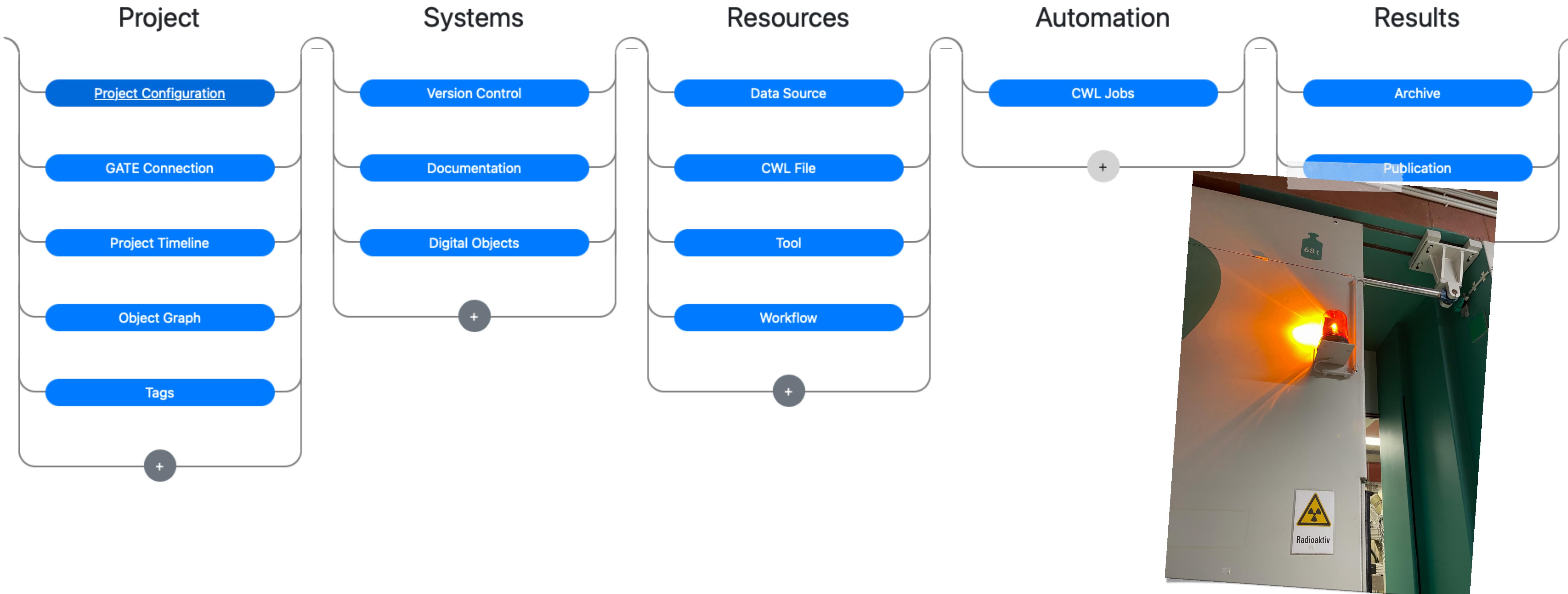
Appearance and Components

...explained with Example III



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:



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, ...

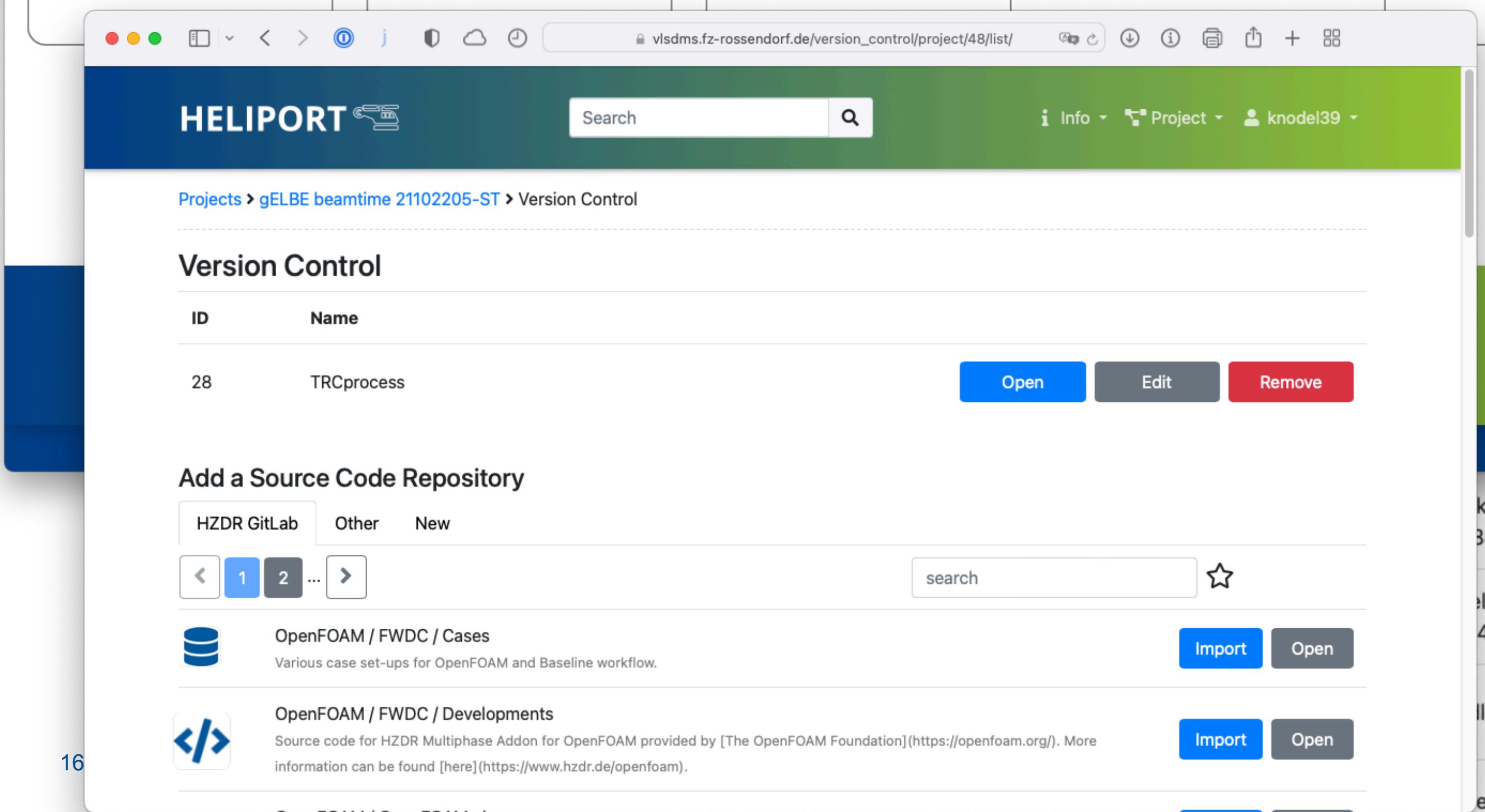
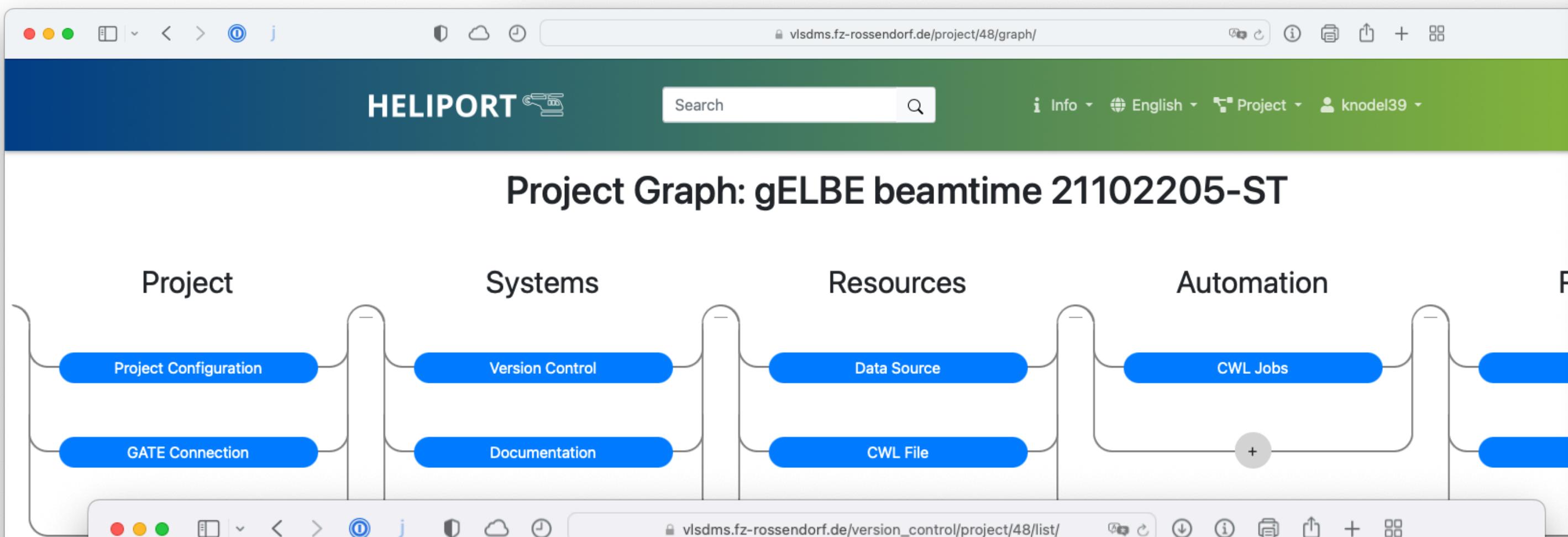
Next Metadata Fields

- Instrument or beam-line,
- Funding,
- Scientific method,
-

The screenshot shows two main sections of the HELIPORT application. The top section displays 'Gate Project' details for a project titled 'gELBE beamtime 21102205-ST'. It includes fields for GATE-ID (2205), Title (Tests of the detector system for the Stopping Target Monitor of the MU2E experiment in a high flux pulsed gamma), Proposer (Mueller, Dr. Stefan (FWCC) - 7394), Abstract (description of the gELBE pulsed gamma beam), and Proposal (21102205-ST). The bottom section shows 'Co-Proposers' (Ferrari, Dr. Anna (FWKH) - 5161) and 'Experimentalists' (Ferrari, Dr. Anna (FWKH) - 5161, Knodel, Dr. Oliver (FWCC) - 132739). The right side of the interface shows 'Settings' for a user profile, including fields for User-ID (knodel39), Surename (Knodel), Givenname (Oliver), E-Mail (o.knodel@hzdr.de), Group (FWCC), Institute (FZR), and ORCID (https://orcid.org/0000-0000-0000-0000). An 'Edit' button is located at the bottom right.

Gate Project	
GATE-ID	2205
Title	Tests of the detector system for the Stopping Target Monitor of the MU2E experiment in a high flux pulsed gamma (Resubmission of 20101909-ST due to COVID pandemic)
Proposer	Mueller, Dr. Stefan (FWCC) - 7394 (Owner of Project "gELBE beamtime 21102205-ST")
Abstract	The gELBE pulsed gamma beam, with narrow pulses set to about 600 kHz repetition rate - the choice of the ELBE micropulses at 406 kHz or 812.5 kHz is ideal in our case- is the unique facility in the world suited to study the performance of the Stopping Target Monitor detector of the Mu2e Experiment. The STM monitor has the crucial role to normalize the conversion rate of the flavor muon conversion rate in the Mu2e detector. The STM monitor has the crucial role to have at ELBE the unique possibility to study the conversion rate of the flavor muon conversion rate in the Mu2e detector. The STM monitor has the crucial role to have at ELBE the unique possibility to study the conversion rate of the flavor muon conversion rate in the Mu2e detector.
Proposal	21102205-ST
Restricted	no
Responsible Experimentalist	Mueller, Dr. Stefan (FWCC) - 7394
Local Contact	Schwengner, Dr. Ronald (FWCC) - 5161
Co-Proposers	
Person	Ferrari, Dr. Anna (FWKH) - 5161
Experimentalists	
Person	Ferrari, Dr. Anna (FWKH) - 5161
Person	Knodel, Dr. Oliver (FWCC) - 132739
Settings	
User-ID	knodel39
Surename	Knodel
Givenname	Oliver
E-Mail	o.knodel@hzdr.de
Group	FWCC
Institute	FZR
ORCID	https://orcid.org/0000-0000-0000-0000

Documentation and Repositories



This screenshot shows a web-based documentation management system interface. The top navigation bar includes standard browser controls, a URL bar (vlsdms.fz-rossendorf.de/documentation/project/48/list/), and user account information (Info, Project, knodel39). The main header features the word "HELIPORT" with a small graphic, a search bar, and navigation links for "Projects", "gELBE beamtime 21102205-ST", and "Documentation".

Documentation

ID	Description	System	Actions
11	Experimental Setup (Room 540/109)	MediaWiki	Open Edit Remove
12	HedgeDoc - Mu2e @ELBE Labbook	HedgeDoc	Open Edit Remove
13	HedgeDoc - Mu2e @ELBE Requirements and Procedure	HedgeDoc	Open Edit Remove
16	Cloud storage containing Pictures, Software, Presentations related to the beamtime (Password: ELBE2021)	Lims	Open Edit Remove

Add a Documentation

Description

Link

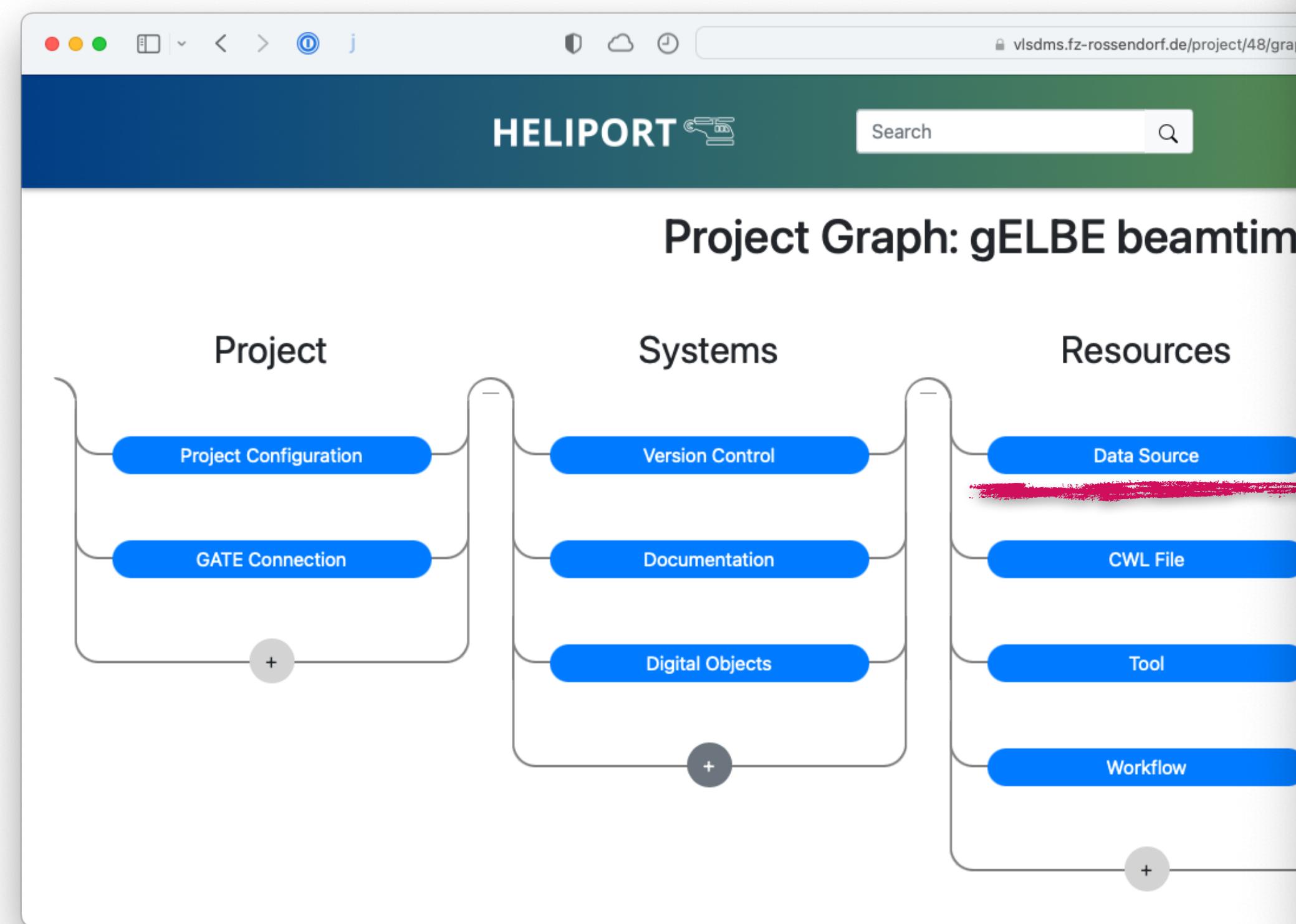
System

or create one with [OpenBis](#), [MediaWiki](#), [MediaWiki \(FWK\)](#), [Lims](#) or [HedgeDoc](#)

The bottom right corner displays logos for CHI JENA and HZDR, indicating their involvement in the project.

Data Sources can be Registered...

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



The screenshot shows the HELIPORT Data Sources interface. The URL in the address bar is `vlsdms.fz-rossendorf.de/data_source/pro`. The page title is "HELIPORT".

The navigation path is: Projects > gELBE beamtime 21102205-ST > Data Sources > Mu2e on bigdata > Beamtimedata2021 > Beamtime_DSPEC_data

Filter options at the top include "Select:", "All", "Files", "Directories", a search bar, and "Select Pattern".

The main content area is titled "Beamtime_DSPEC_data" and lists several items:

Item Type	Content	Action Buttons
Folder	15092021	Publish Zipped x Add Tag Sept. 21, 2021, 2:16 p.m.
Folder	16092021	Add Tag Sept. 21, 2021, 3:05 p.m.
Folder	20092021	Add Tag Sept. 21, 2021, 2:16 p.m.
File	ELBE TI Current-data-2021-09-17 16_52_59.xlsx	Publish x Add Tag 53 KiB Sept. 21, 2021, 2:16 p.m.
File	ELBE_2021-09-15.png	Publish x Add Tag 121 KiB Sept. 21, 2021, 2:16 p.m.
File	ELBE_2021-09-16.png	Add Tag 112 KiB Sept. 21, 2021, 2:16 p.m.

Integration in an Overall Data Publication

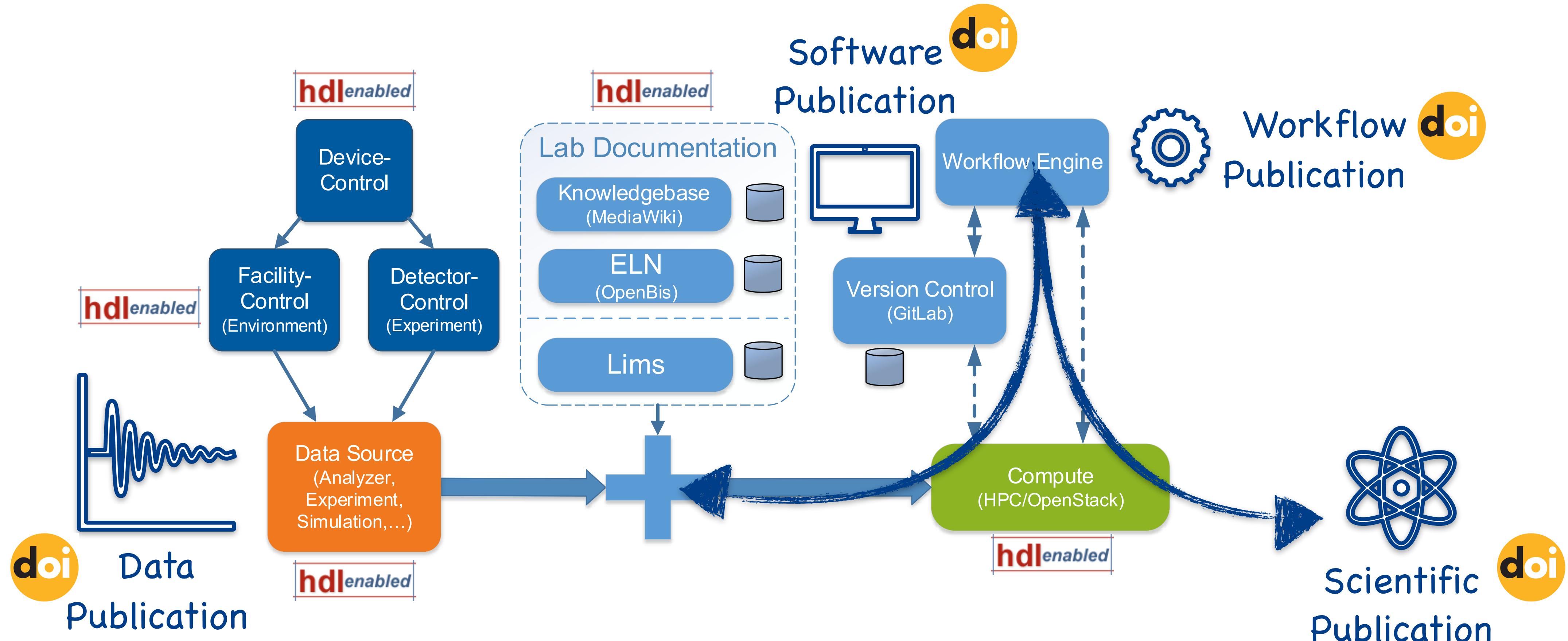
The figure displays three screenshots illustrating the integration of HELIPORT and RO2DARE for data publication:

- Left Screenshot (HELIPORT):** Shows the 'Gate Project' details for 'gELBE beamtime 21102205-ST'. It includes fields like GATE-ID (2205), Title (Tests of the detector system for the Stopping Target Monitor of the MU2E experiment in a high flux pulsed gamma beam), and Abstract (description of the experiment). A blue arrow points from this screen to the central 'Project Graph'.
- Middle Screenshot (Project Graph):** Titled 'Project Graph: gELBE beamtime 21102205-ST', it shows a diagram with two columns: 'Systems' (Version control, Documentation, Digital Objects) and 'Resources' (Data Source, CWL File, Tool, Workflow). A blue arrow points from the HELIPORT screenshot to this graph.
- Right Screenshot (RO2DARE):** A dataset page for record 1343. The title is 'Tests of the detector system for the Stopping Target Monitor of the MU2E experiment in a high flux pulsed gamma beam'. The abstract describes the experiment. Below the abstract is a preview of a zip file containing data from 'Beamtime_DSPEC_data.zip', showing files like 'ELBE_Run2_Nal.Spe' and 'ELBE_Run6_LaBr.Spe'. The page also shows statistics (244 views, 66 downloads) and publication details (December 20, 2021).

Automated data publication with:

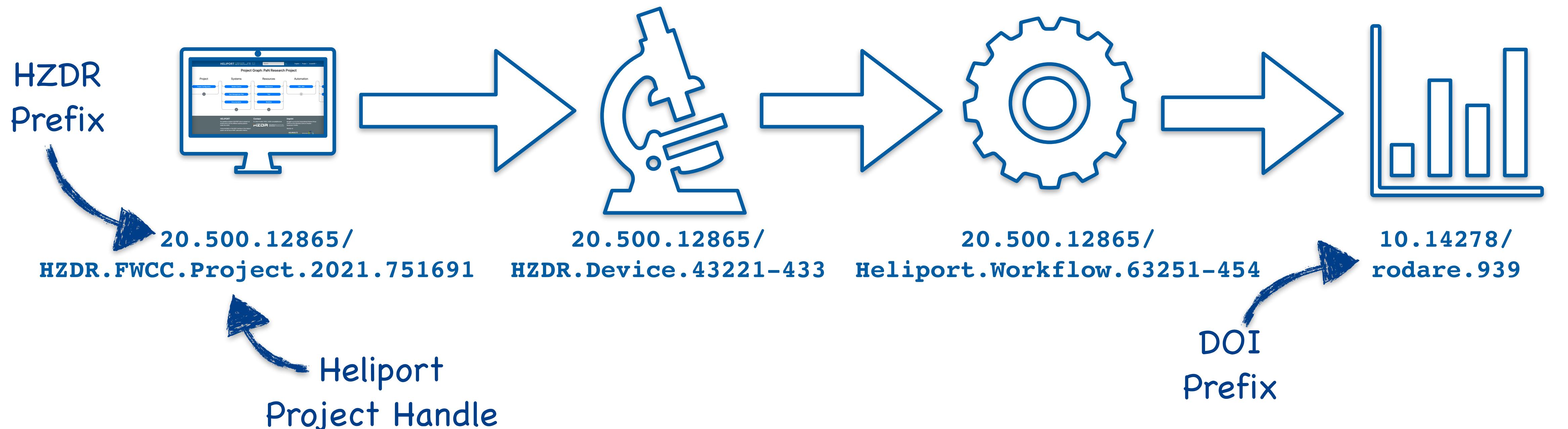
- Metadata from Proposal System
- Data sources registered and selected in HELIPORT

Different Types of (Data) Publications and Data Provenance



Persistent 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

A screenshot of the HELIPORT web interface. The page title is "HELIPORT HELmholtz Scientific Project Workflow Platform". The main content area shows a table of "Jobs" with columns: ID, Name, Cluster Login, Directory on Cluster, and Status. There are several rows of jobs listed, each with a set of blue and red icons for managing the job.

Workflow Engine

Version Control

Compute
(HPC, OpenStack)



A screenshot of the GitLab web interface. The page title is "GitLab Projects Groups Activity Milestones Snippets Help Search or jump to...". The main content area shows a repository named "DAQAlg2PPGA" with a file named "core_template.c". The code editor shows a C program with various functions and comments.

UNICORE

JÜLICH
FORSCHUNGSZENTRUM

slurm
workload manager

HELI

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



A screenshot of the HELIPORT web interface showing the "Edit a Scientific Workflow" page. The page title is "HELIPORT HELmholtz Scientific Project Workflow Platform". The main content area shows a form for editing a workflow with fields for "Name" (curl and cat stdout and stderr) and "Description". Below the form is a graphical workflow diagram with nodes labeled "link", "curl", "cat", and "cat_1". At the bottom are buttons for "Save", "Cancel", "Fit to Screen", and "Delete Selection".

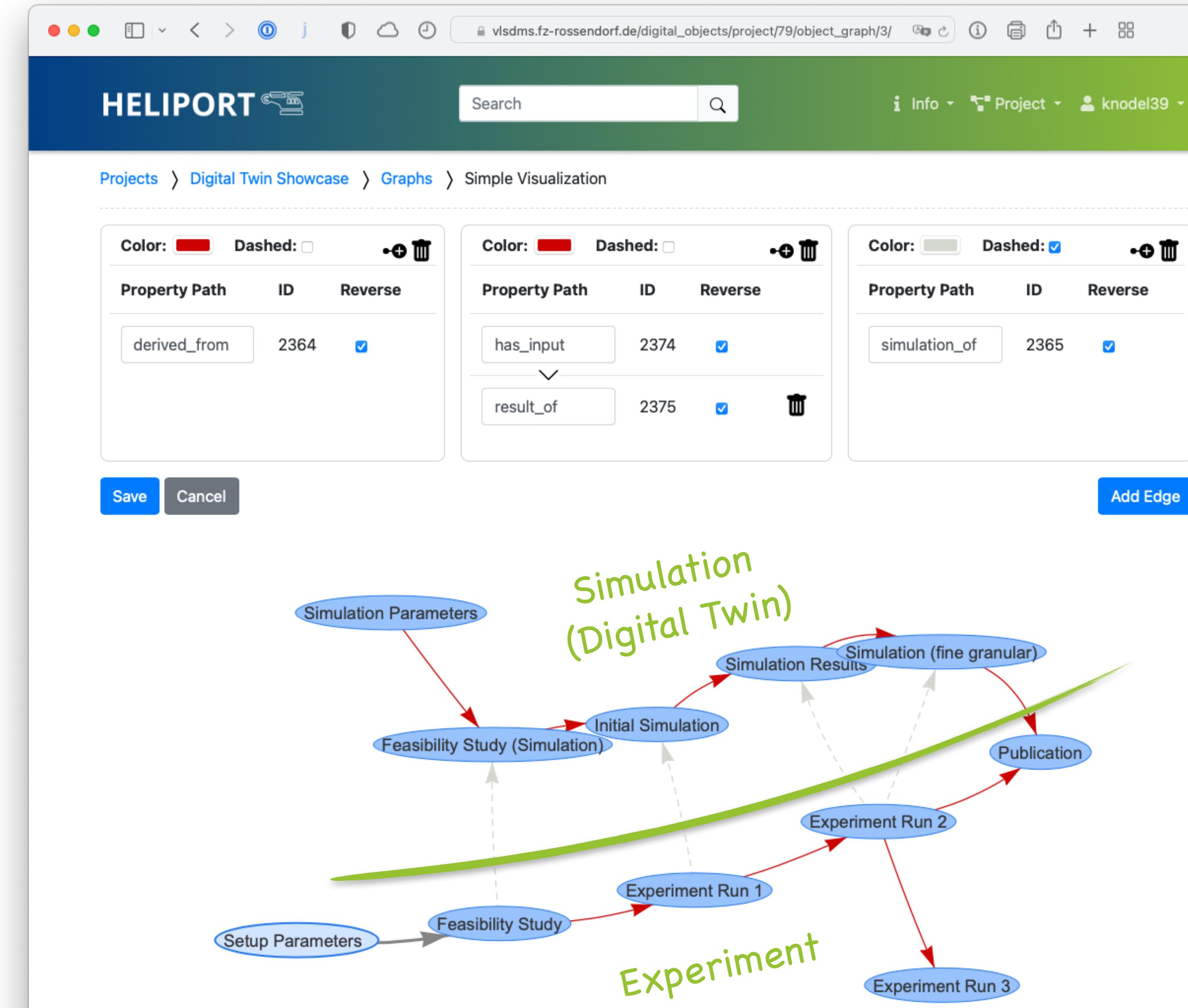
Scientific Software Development and Reproducible Workflows II

The figure displays three screenshots of the HELIPORT web application, which is a Helmholtz Scientific Project Workflow Platform.

- Left Screenshot:** Shows the "Jobs" page. A table lists several jobs with columns for ID, Name, Cluster Login, Directory on Cluster, and Status. Each job entry includes a set of icons for managing the task. The URL is vlstdms.fz-rossendorf.de/cwl_execution/project/17/job/list/.
- Middle Screenshot:** Shows the "Edit a Scientific Workflow" page. It features a workflow diagram with nodes labeled: full_name, mp_name, new_step, natoms, new_step_1, and supercell. Arrows indicate the flow between these nodes. Below the diagram is a detailed description of the workflow step: "Create supercell from element and structure". The URL is vlstdms.fz-rossendorf.de/cwl_execution/project/41/workflow/80/update.
- Right Screenshot:** Shows the "Test job Run" details for a specific job. The page displays a log of the job's execution, including download progress for various packages and tools. The URL is vlstdms.fz-rossendorf.de/cwl_execution/project/41/job/60/detail/.

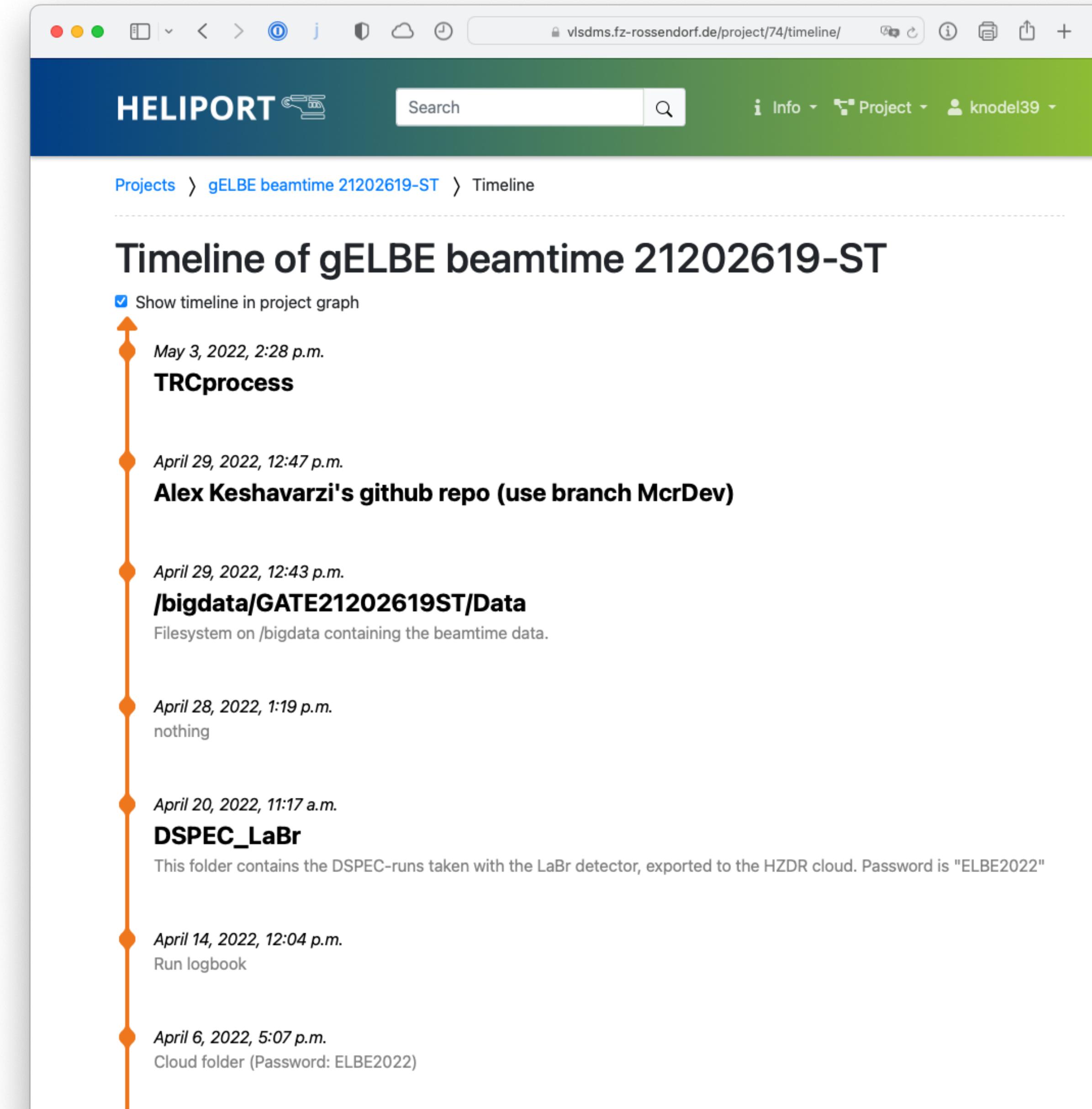
Relations Between Digital Objects

- 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, beamline 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.

The screenshot shows the Heliport REST API documentation in ReDOC. The left sidebar lists various API endpoints under categories like 'api', 'gate-connection', 'version-control', etc. The main content area is focused on the 'createDigitalObject' endpoint. It shows the 'REQUEST BODY SCHEMA: application/json' with fields: 'project' (integer, required), 'handle' (string, nullable, max length 100 characters), 'relation' (string, required), 'category' (string, required), and 'description' (string, required). Below this, there's a 'Responses' section showing a green button labeled '> 201'. To the right, there's a 'Request samples' section with a 'Payload' tab showing the JSON schema for the request body, and a 'Response samples' section with a '201' status code tab showing the response body.

POST /digital_objects/api

Request samples

Payload

Content type
application/json

Copy Expand all Collap

{
 "project": 0,
 "handle": "string",
 "relation": "string",
 "category": "string",
 "description": "string"
}

Responses

> 201

201

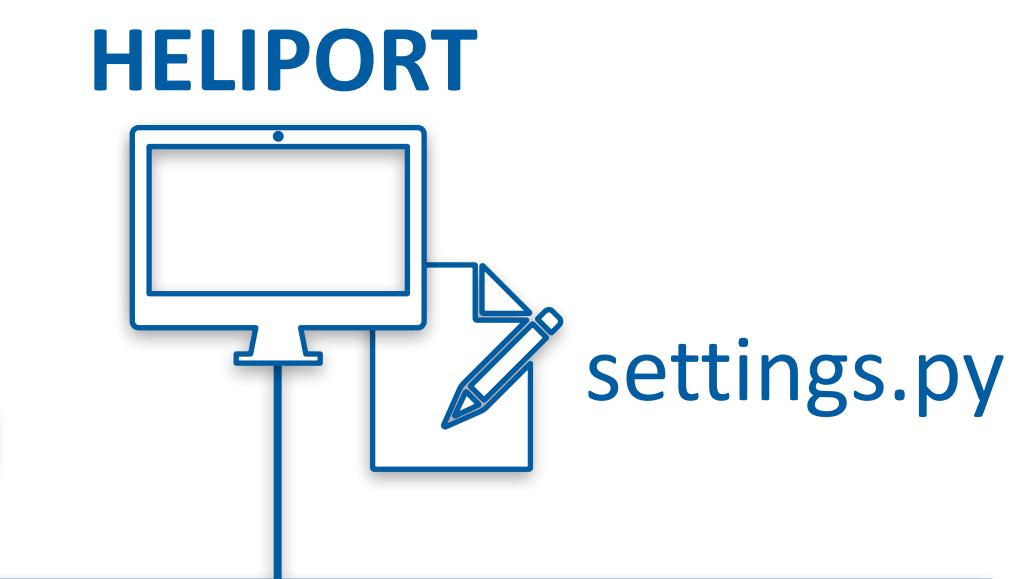
Content type
application/json

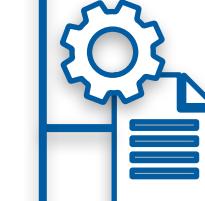
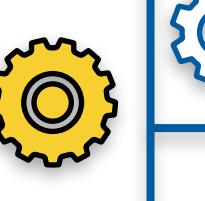
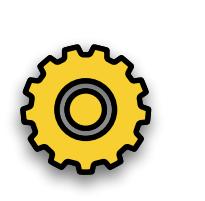
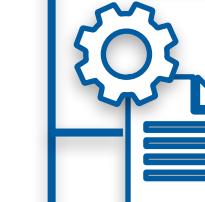
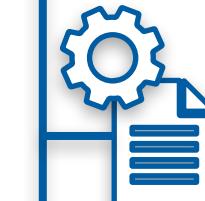
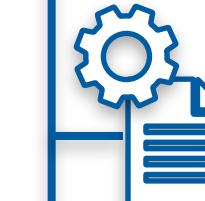
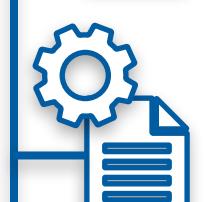
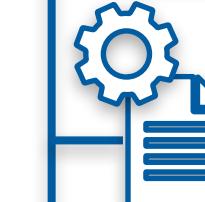
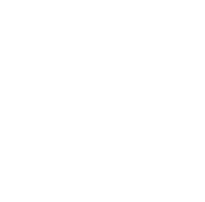
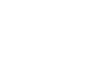
Copy Expand all Collap

{
 "digital_object_id": 0,
 "project": 0,
 "handle": "string",
 "relation": "string",
 "category": "string",
 "description": "string"
}

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](https://doi.org/10.14278/rodare.947)



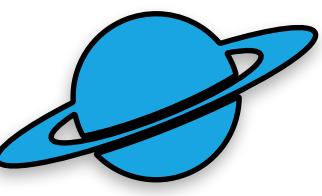
Project & Infrastructure	Systems	Resources	Automation	Results
 Core Ressources 	 Version Control 	 Data Sources 	 Slurm Job 	 Publication 
 HELIPORT Project 	 Data Management Plan (RDMO) 	 Tools 	 Unicore 	 Dataexport Invenio 
 Proposal Management (HZDR Gate) 	 Documentation 	 Arbitrary CWL 	 ...	 Internal Archive 
 About Page 	 Digital Objects 	 General Workflows 	 Aiida Workflows 	 ...
<p>...  Available  In development  Planned</p>				

 Available

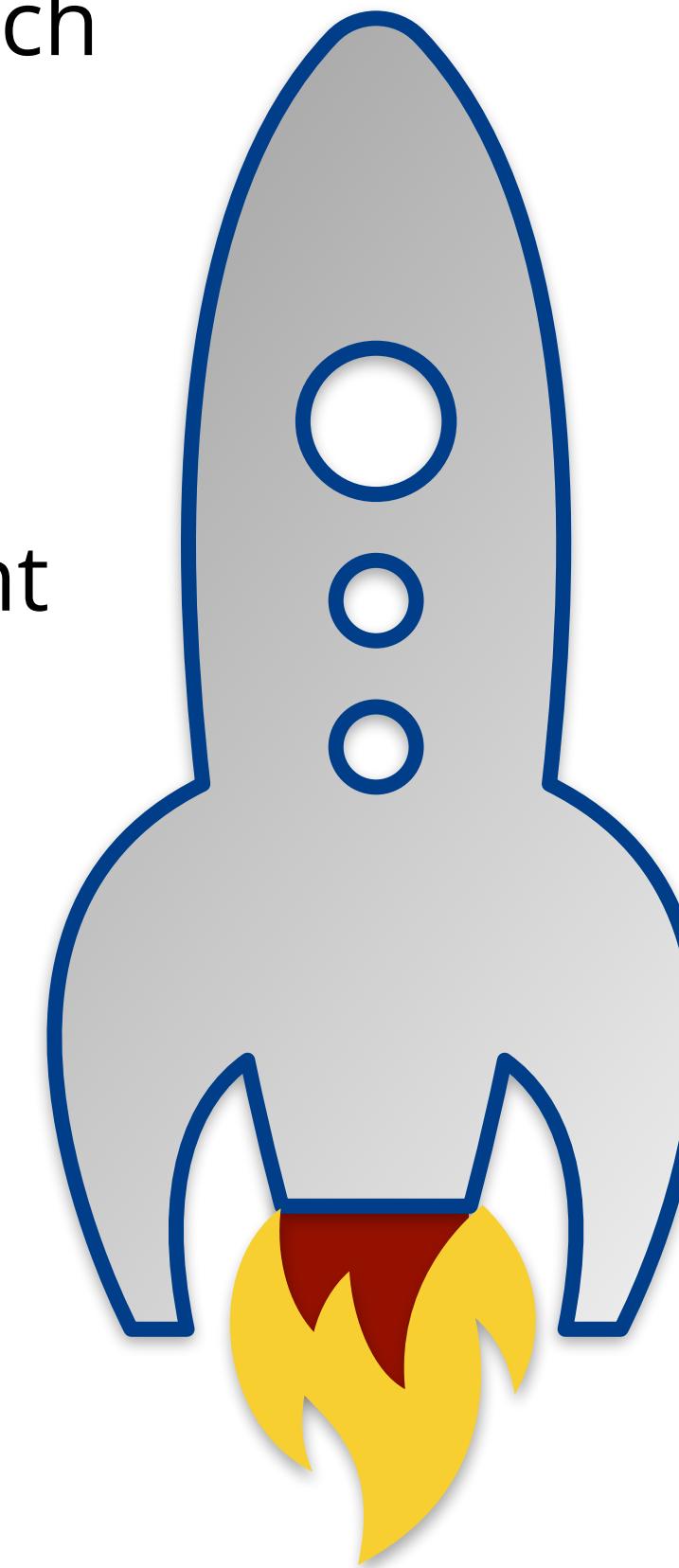
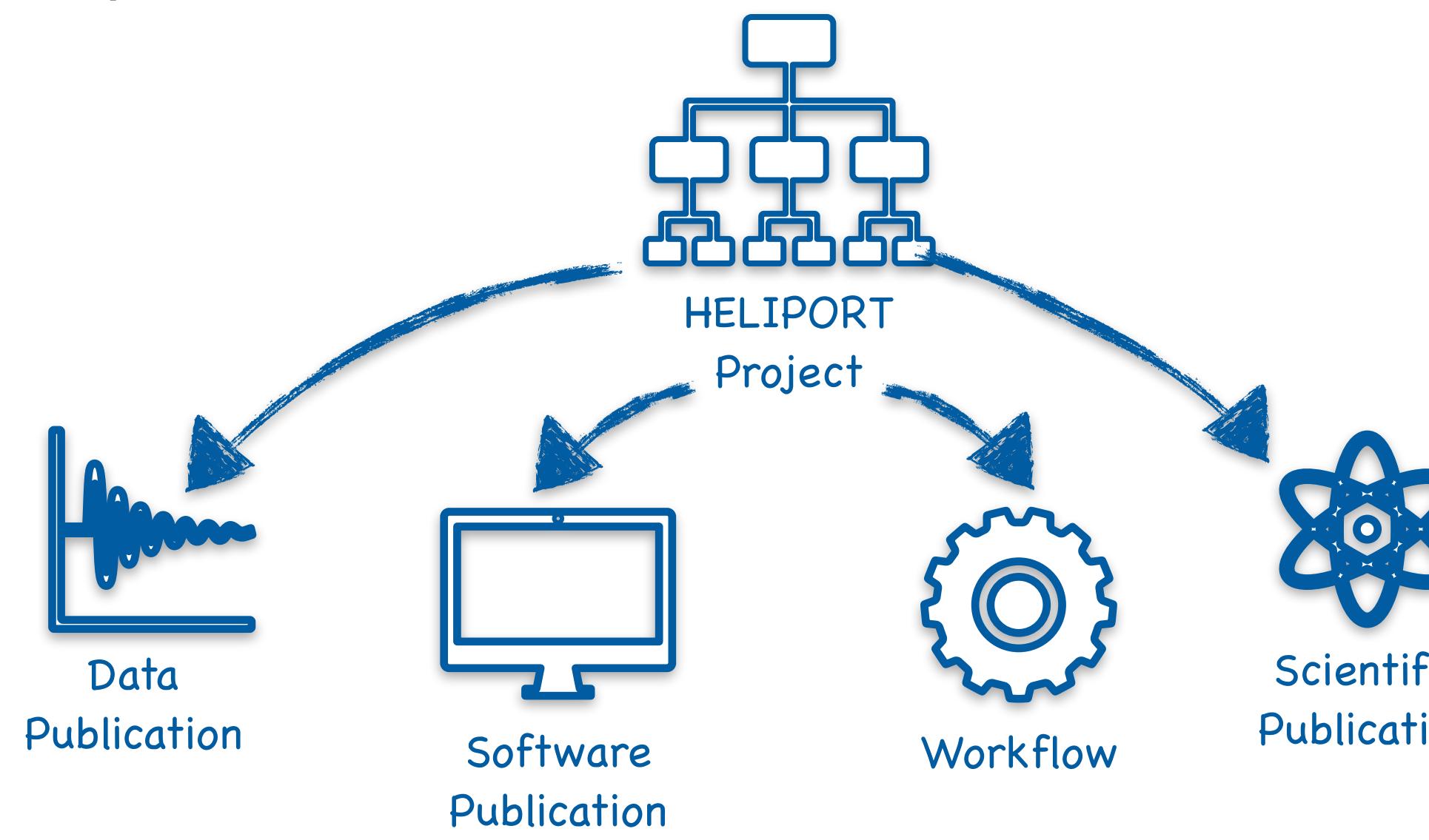
 In development

 Planned

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.



Resources

Website: heliport.hzdr.de

Project Graph: gELBE beamtime 21102205-ST

Project

- Project Configuration
- GATE Connection

Systems

- Version Control
- Documentation
- Digital Objects

Resources

- Data Source
- CWL File
- Tool
- Workflow

Automation

- CWL Jobs
- Archive

Results

Demo: heliport.hzdr.de/app

HELIPORT HELmholtz Scientific Project WORkflow PlaTform

The guidance system HELIPORT aims to make the entire workflow interoperable and reusable according to the FAIR principles, meaning it deals with the areas from the generation of the data to the publication of the actual research results. For this purpose, a concept was developed ...

HELIPORT (0.0.3)

Download OpenAPI specification: [Download](#)

api

listProjects

Showing the most general HELIPORT project properties. Request more detailed information by appending url/s to the url Go to a specific project by appending its id/s to the url For more information on how to authenticate look in HELIPORT (user > settings)

QUERY PARAMETERS

limit	integer Number of results to return per page.
offset	integer The initial index from which to return the results.
search	string A search term.
group	string group
owner	string owner

GET /api/projects/

Response sample

```

200
Content type application/json

Copy Expand all Collapse all

{
  "count": 123,
  "next": "http://api.example.org/account/next",
  "previous": "http://api.example.org/account/prev",
  "results": [
    {
      ...
    }
  ]
}

```

API Doc: heliport.hzdr.de/redoc/