

## The Integrated Research Data Lifecycle of the HELIPORT Project

HMC Conference 2022 // October 5, 2022

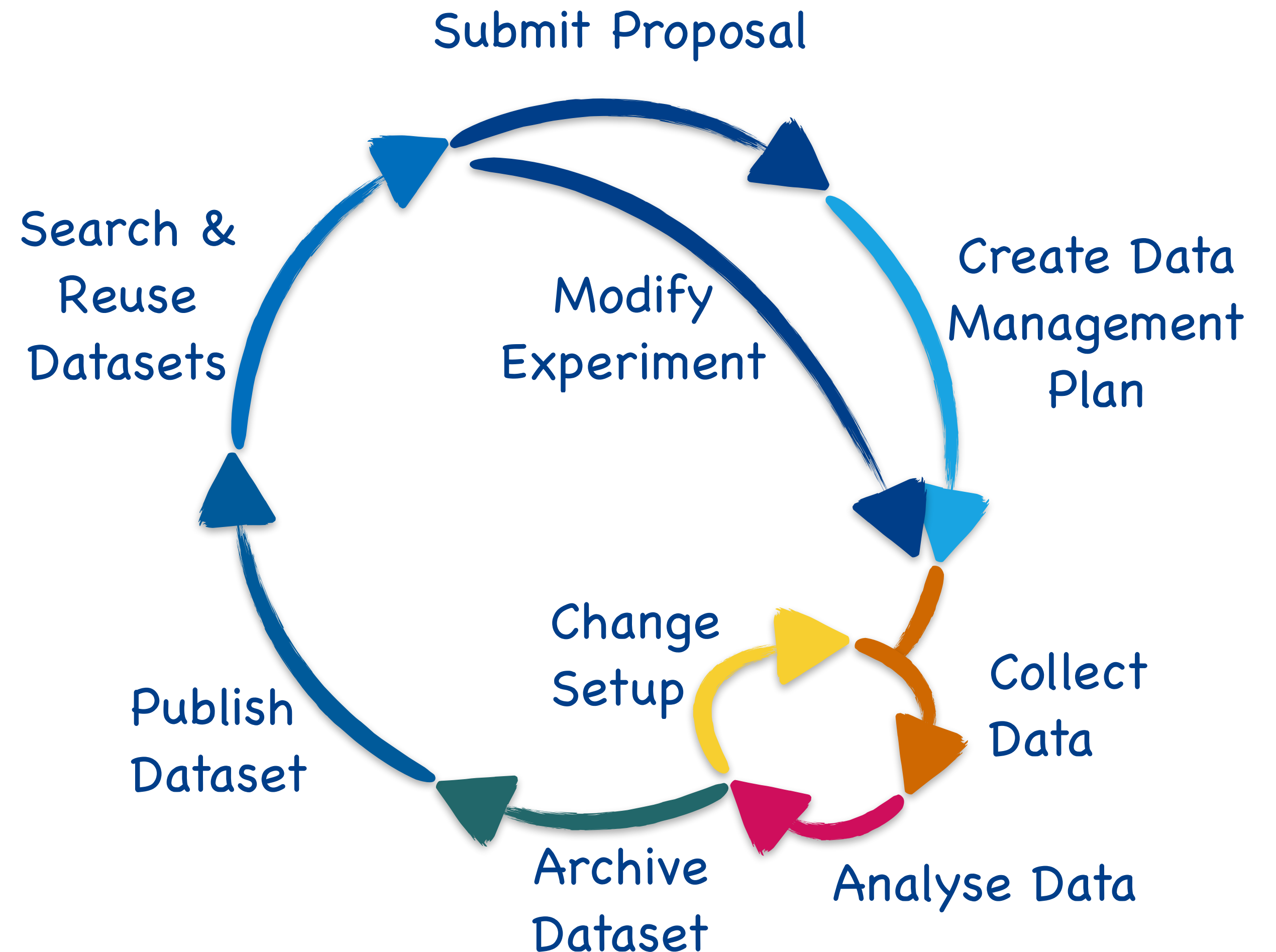
**Oliver Knodel**, Martin Voigt, Robert Ufer, David Pape, Mani Lokamani, Jeffrey Kelling, Stefan E. Müller, Thomas Gruber, Guido Juckeland, Alexander Kessler, Chien-Li Lee, Joachim Hein, Bernd Schuller // contact: o.knodel@hzdr.de





# Our 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.



# The Motivation to Develop HELIPORT

- HELIPORT was originally intended to provide only the **proposal's metadata**, to allow the assignment of resources.
- Over time, we realised that HELIPORT can also answer our scientists' most important questions, such as:

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 associated services/tools?



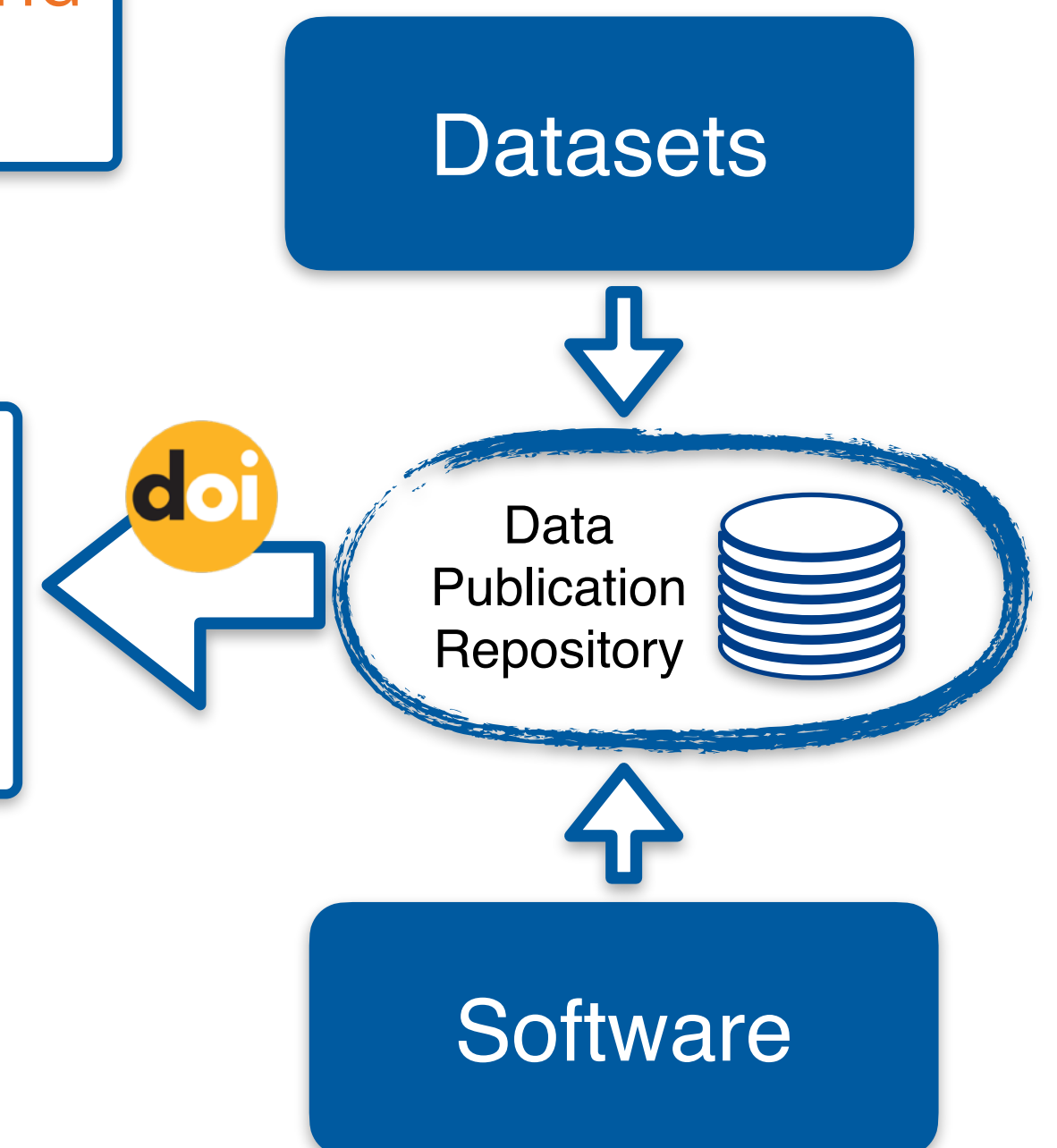
What are the necessary steps towards a full comprehensible and FAIR research experiment ensuring data provenance?

How we can **reduce computations and save energy?**



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

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



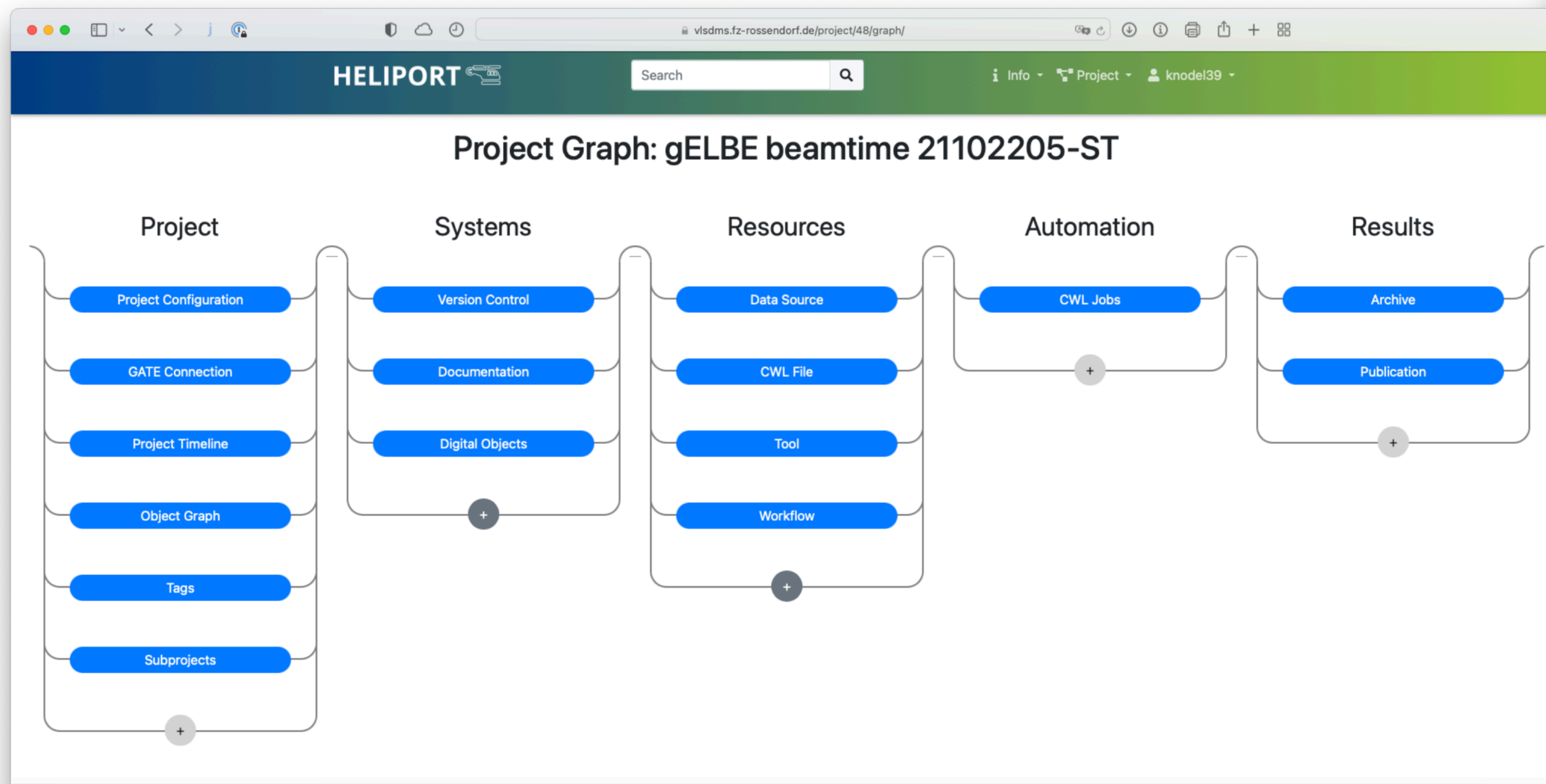


“ The HELIPORT project aims at developing a platform which accommodates the **complete life cycle** of a scientific project and links all corresponding programs, systems and workflows to create a more **FAIR** and comprehensible project description.

Project Members:



Funded by:



```

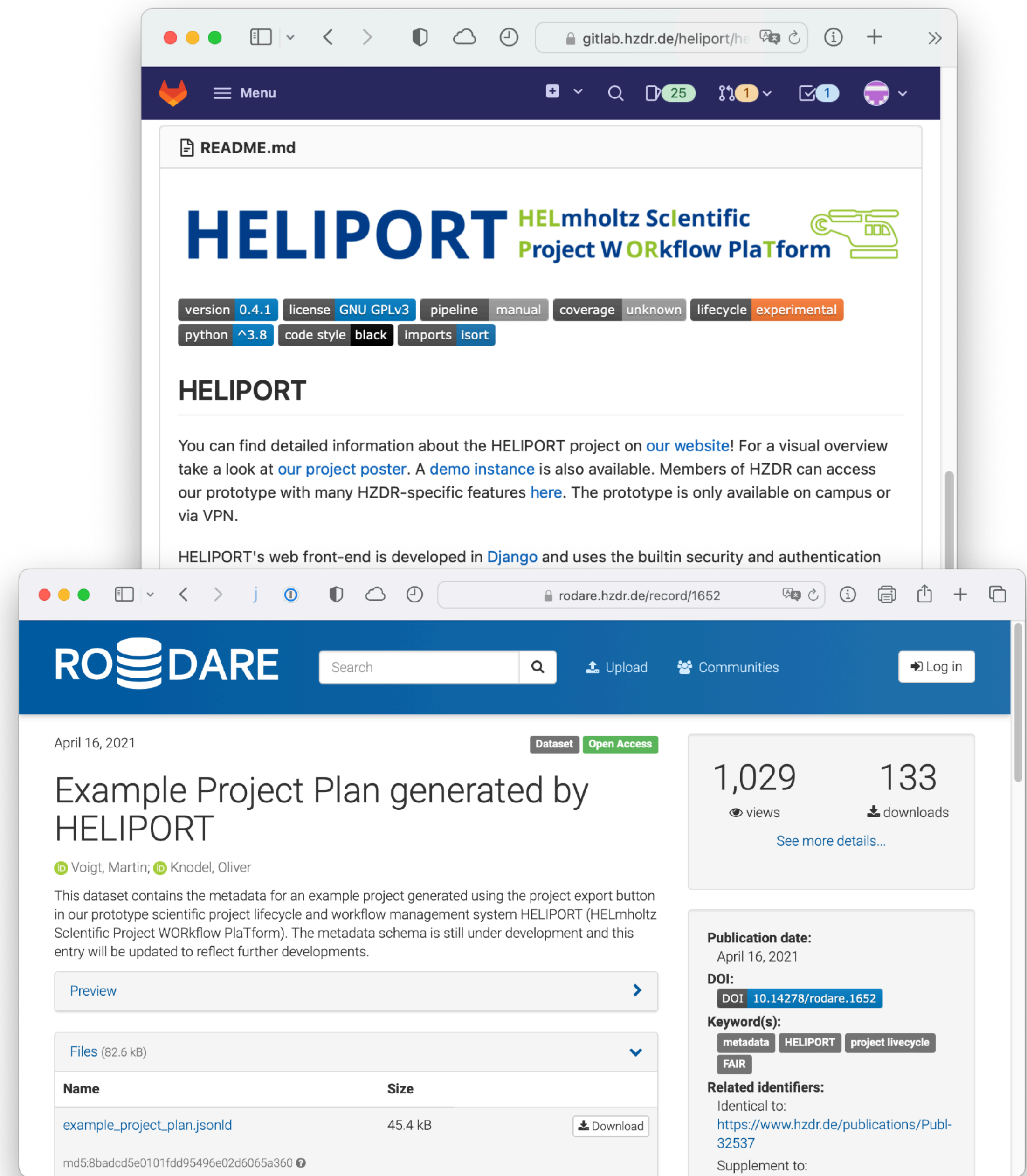
"namespaces": {
  "datacite": "http://purl.org/spar/datacite/",
  "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
  "heliport": "https://heliport/schema/",
  "time": "http://www.w3.org/2006/time#",
  "dc": "http://purl.org/dc/terms/"
},
"heliprot:project_id": 28,
"datacite:hasIdentifier": "HZDR.FWCC.2021.84769",
"heliprot:uuid": "09779261-200c-48c4-be9c-f298369d6a1c",
"datacite:handle": "https://hdl.handle.net/None",
"heliprot:project_name": "PaN Research Project",
"time:hasBeginning": "2021-04-01 09:14:34.296524+00:00",
"datacite:hasDescription": "",
"heliprot:group": "FWCC",
"heliprot:owner": {
  "datacite:hasIdentifier": "132739",
  "datacite:orcid": null,
  "rdfs:label": "Knodel, Dr. Oliver (FWCC) - 132739"
},
"heliprot:has_VersionControl": [
  {
    "heliprot:version_control_id": 15,
    "datacite:uri": "https://ddd",
    "rdfs:label": "Test"
  }
],
"heliprot:has_DataManagementPlan": [
  {
    "heliprot:data_management_plan_id": 6,
    "datacite:uri": "https://dddd",
    "datacite:hasDescription": "dddd"
  }
],
"heliprot:has_Documentation": [
  {
    "heliprot:documentation_id": 7,
    "datacite:uri": "https://dddd",
    "heliprot:documentation_system": "MediaWiki",
    "datacite:hasDescription": "dddd"
  }
],
"heliprot:has_DataSource": [
  {
    "heliprot:data_source_id": 11,
    "datacite:uri": "http://ddd",
    "heliprot:use_computer": null,
    "rdfs:label": "ddd",
    "datacite:hasDescription": ""
  }
]

```



# The First Year...

- Discussions on the modular HELIPOINT architecture and first prototype (D3) also available on [gitlab.hzdr.de](https://gitlab.hzdr.de) and: DOI 10.14278/rodare.947 ✓
- Project-level metadata schema as a first draft based on DataCite (D1): DOI 10.14278/rodare.1652 ✓
- The experiment-specific metadata is not directly part of the HELIPOINT metadata:
  - For the laser experiments we started discussions within the community (Mattermost Team, Laserlab Europe talks, ...).
  - We started the integrate/linkage of external metadata catalogues in HELIPOINT (e.g. ICAT, SciCat, ...).
- A first concept for mapping CWL to the UNICORE language has been completed (D2). ✓





# The Project Meeting at HI Jena in 2022

HELIPORT project meeting **IN PERSON**

Jun 22, 2022, 12:00 PM → Jun 23, 2022, 1:00 PM Europe/Berlin  
 Helmholtz Institut Jena  
 Alexander Kessler (Helmholtz Institut Jena), Oliver Knodel (Helmholtz-Zentrum Dresden-Rossendorf (HZDR))

**Description**

## HELIPORT HELmholtz Scientific Project WORKflow PlatForm

HELIPORT (Helmholtz Scientific Project WORKflow PlatForm) is a project funded by the Helmholtz Metadata Collaboration, and runs from July 2021 until June 2023. HELIPORT aims to make the entire life cycle of a scientific project findable, accessible, interoperable and reusable according to the FAIR principles, mentioned below. In particular, our data management solution deals with the areas from the generation of the data to the publication of primary research data, the workflows carried out and the actual research results. For this purpose, a concept was developed which shows the various essential components and their connections. Descriptions of the individual components can be found in our HZDR Data Management Strategy.

This event is intended as an internal project meeting located at Helmholtz Institute Jena and open for all members of our HELIPORT project from HZDR, HIJ and FZJ.

**Registration** You are registered for this event. 10 [Check details](#)

**Participants**

- Alexander Kessler
- Bernd Schuller
- David Pape
- Jeffrey Kelling
- Joachim Hein
- Malte Christoph Kaluza
- Mani Lokamani
- Martin Voigt
- Oliver Knodel
- Stefan Mueller

Contact [heliport@hzdr.de](mailto:heliport@hzdr.de)

**WEDNESDAY, JUNE 22**

- 12:00 PM → 12:15 PM Welcome, Introduction to HELIPORT and Roadmap** (15m)
  - Speaker: Oliver Knodel (Helmholtz-Zentrum Dresden-Rossendorf (HZDR))
- 12:15 PM → 1:00 PM Features and latest extensions** (45m)
  - 1. Infrastructure Report: Updates on code, CI, packaging, and other infrastructure-related topics. (David Pape)
  - 2. Latest Developments and Features: Showcase of major updates since the last HELIPORT release on Rodare. (Martin Voigt)
  - 3. User Discussion: Discussion on use cases, features, usability, and more.
  - Speakers: David Pape (HZDR), Martin Voigt (HZDR)
- 1:15 PM → 2:15 PM HELIPORT@Jena Status and Metadata DB Discussion** (1h)
  - Speaker: Alexander Kessler (Helmholtz Institut Jena)
- 2:30 PM → 4:30 PM Interactive Session I: Deployment** (2h)
  - Speaker: David Pape (HZDR)
- 5:00 PM → 6:00 PM POLARIS Tour** (1h)
  - Speaker: Alexander Kessler
- 6:00 PM → 6:20 PM Summary Day 1** (20m)
  - Speaker: Oliver Knodel (Helmholtz-Zentrum Dresden-Rossendorf (HZDR))





# Important (but unplanned) Milestone: heliport.hzdr.de

- We received an increasing number of questions regarding heliport.
- To address the demands we created the website **heliport.hzdr.de** with:
  - Overall information on HELIPORT,
  - Documentation,
  - Ressources (Poster, Presentation, Paper)
  - News section and
  - Our two HELIPORT systems:



Demo  
Version

HZDR  
Version  
(Internal)



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

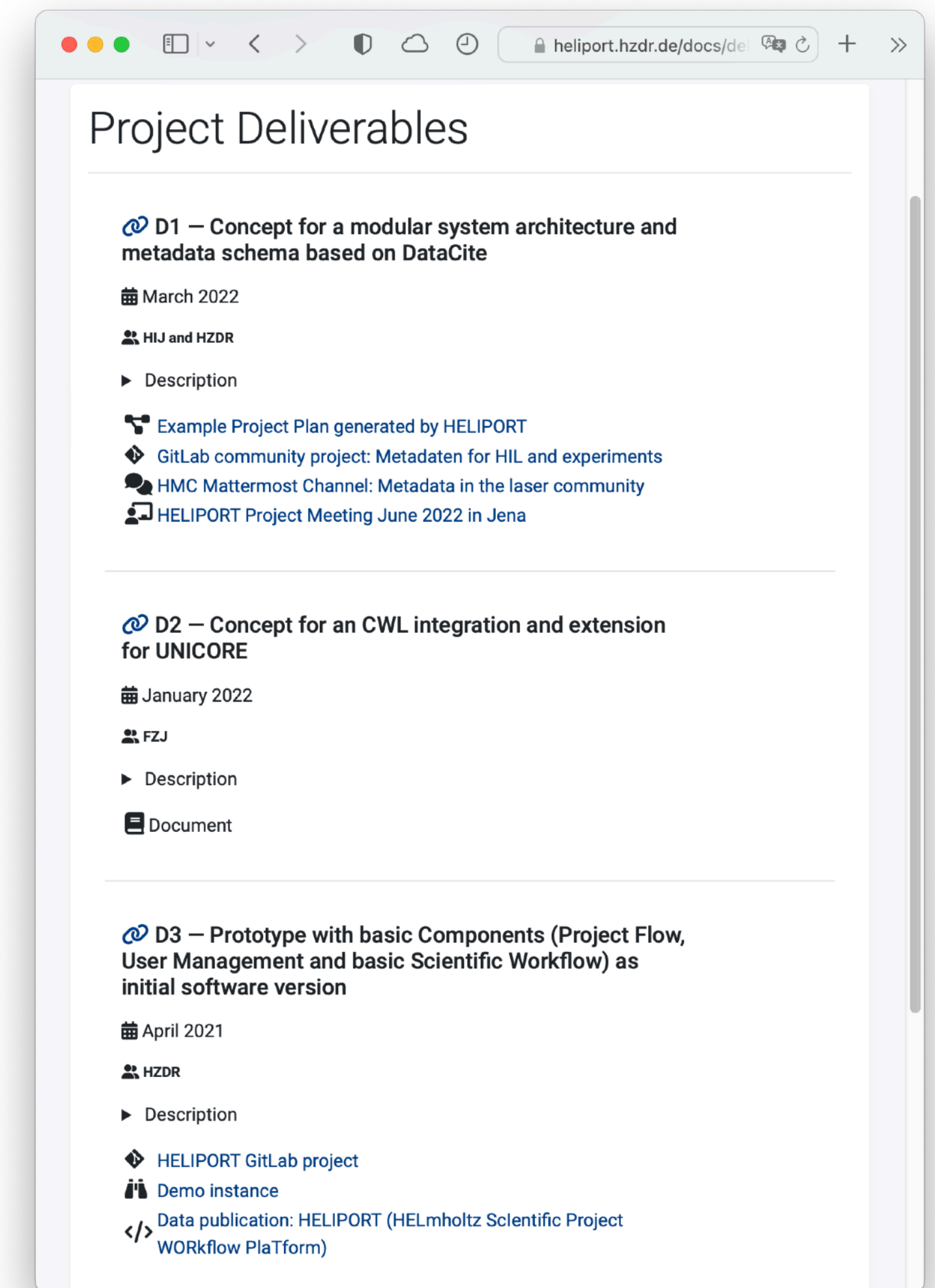
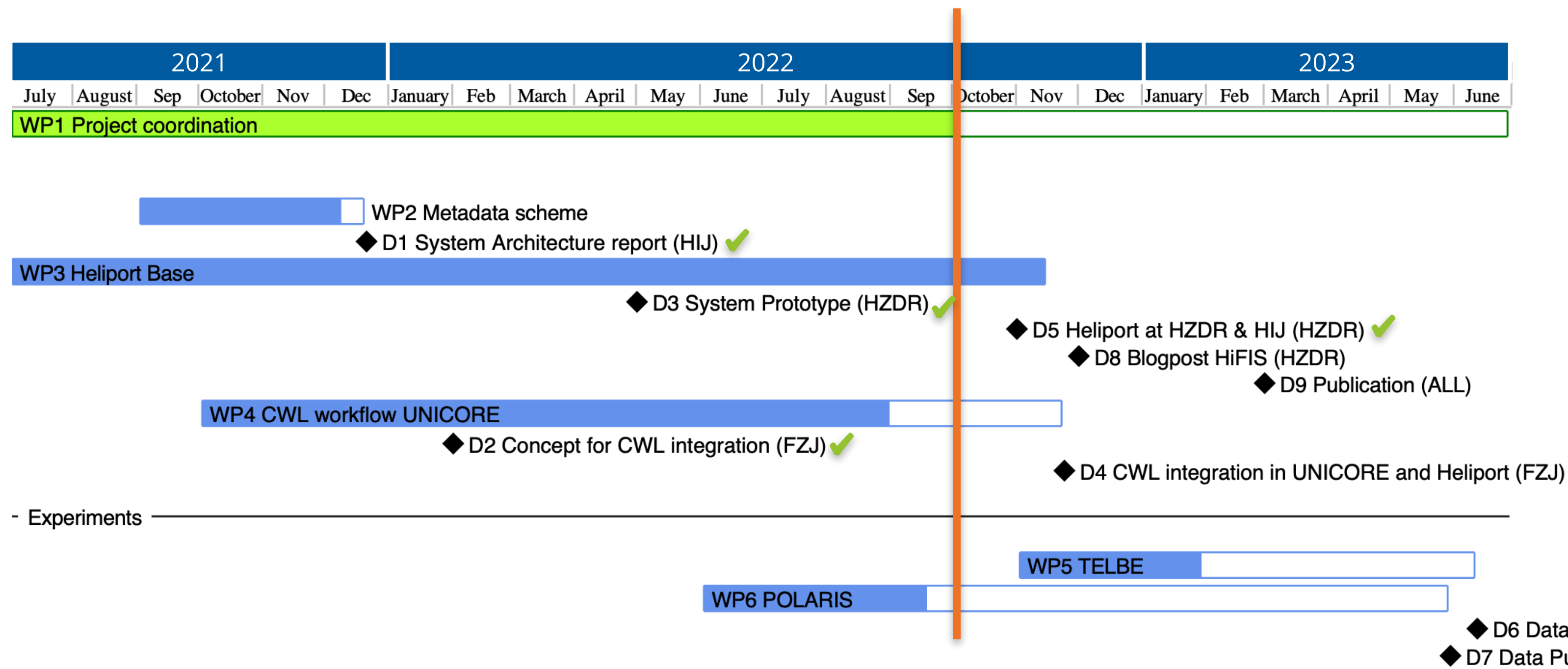
Below the text, it says "Intuitive and structured user interface". The text continues: "The clearly structured user interface of HELIPORT let's you easily create, manage and create scientific projects. In the future an Elasticsearch powered search backend will help you find information from other (similar) projects to compare and improve your methods and workflows with other projects."

The interface screenshot shows a "Project Graph: gELBE beamtime 21102205-ST" with sections for Project, Systems, Resources, Automation, and Results. The Project section includes Project Configuration, GATE Connection, and Digital Objects. The Systems section includes Version Control, Documentation, and Digital Objects. The Resources section includes Data Source, Data File, Tool, and Workflow. The Automation section includes Data Jobs. The Results section includes Archive and Publication. There is also a "Jobs" table with columns for Name, Cluster Name, Directory on Cluster, and Status.



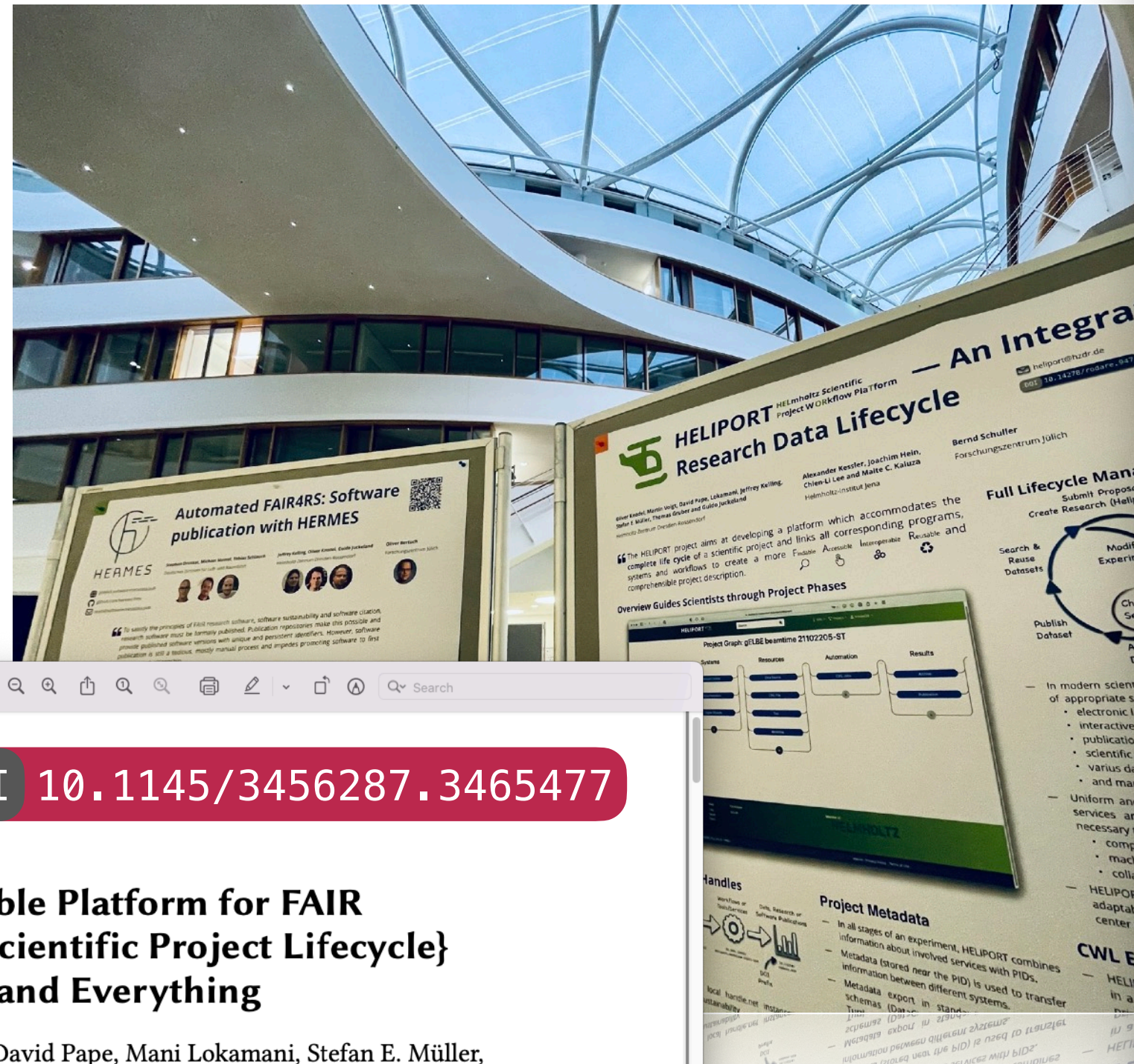
# Overview of our Deliverables and Results

- Our first three deliverables are completed and the results are available on our website.
- We speak with different Helmholtz members, universities and European partners about HELIPOINT and build a HELIPOINT community.
- Overview of work packages and milestones:





# Our HELIPORT Outreach



Workshop Presentation DOI 10.1145/3456287.3465477

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

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

**ABSTRACT**  
Modern scientific collaborations and projects (MSCPs) employ various processing stages, starting with the proposal submission, continuing with data acquisition and concluding with final publications. The realization of such MSCPs poses a huge challenge due to (1) the complexity and diversity of the tools, (2) the heterogeneity of various involved computing and experimental platforms, (3) flexibility of analysis targets towards data acquisition and (4) data throughput. Another challenge for MSCPs is to provide additional metadata according to the FAIR principles for all processing stages for internal and external use. Consequently, the demand for a system, that assists the scientist in all project stages and archives all processes on the basis of metadata standards like DataCite to make really everything transparent, understandable and citable, has risen

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

**1 INTRODUCTION**  
An essential objective of modern cutting-edge research should be to enable accessibility of the acquired research data and its re-usability across different research fields and their respective communities. The current generation of scientists is therefore faced with the challenging task of transferring experimental investigations into a data oriented research flow with strong focus on documenting every step closely following the FAIR [41] principles. The FAIR principles are well-established as standards in the field of research data management. The three pillars F (Findable), A (Accessible) and

HELIPORT | Blog

[About](#) | [News](#) | [System](#) | [Docs](#) | [Contribute](#)

### HELIPORT and HERMES at the Virtual HMC Conference

VIRTUAL EVENT | 05.-06.10.2022

by David Pape on October 5, 2022 Photo by HMC

This week, on the 5th and 6th of October, we have the chance to present our HELIPORT and HERMES posters at the virtual conference of the Helmholtz Metadata Collaboration (HMC) who are kindly funding these projects.

### HELIPORT and HERMES at MT-DMA in Hamburg

by Oliver Knodel on September 26, 2022 Photo by Oliver Knodel

Our third presentation of the new HELIPORT poster takes place in Hamburg at the 8th Annual "Matter and Technologies" Meeting at DESY. Today we have the other HMC project we are involved in next to our poster: Automated FAIR4RS software publication with HERMES.

### HELIPORT Poster at SaxFDM in Leipzig

22. September 2022

by Oliver Knodel on September 22, 2022 Photo by David Pape

Our second presentation of the new HELIPORT poster takes place in Leipzig at the "3. Sächsische FDM-Tagung" with the topic "Forschungsdatenmanagement im Spannungsfeld zwischen Idealen, Anforderungen und Praxis" - a perfect topic and event for our project.

### Poster Presentation at SNI in Berlin

by Oliver Knodel on September 6, 2022 Photo by David Pape

The "German Conference for Research with Synchrotron Radiation, Neutrons and Ion Beams at Large Facilities" was a wonderful opportunity to present our new HELIPORT poster to a potential user group and to discuss future requirements.

### HELIPORT project meeting June 2022 in Jena

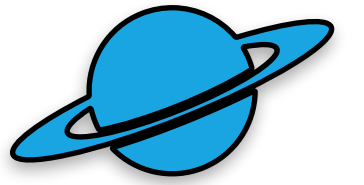
by Oliver Knodel on June 24, 2022 Photo by Oliver Knodel

Our first in-person and real-life HELIPORT project meeting took place at Helmholtz Institute Jena!

Previous 1 Next



# Conclusions and Outlook

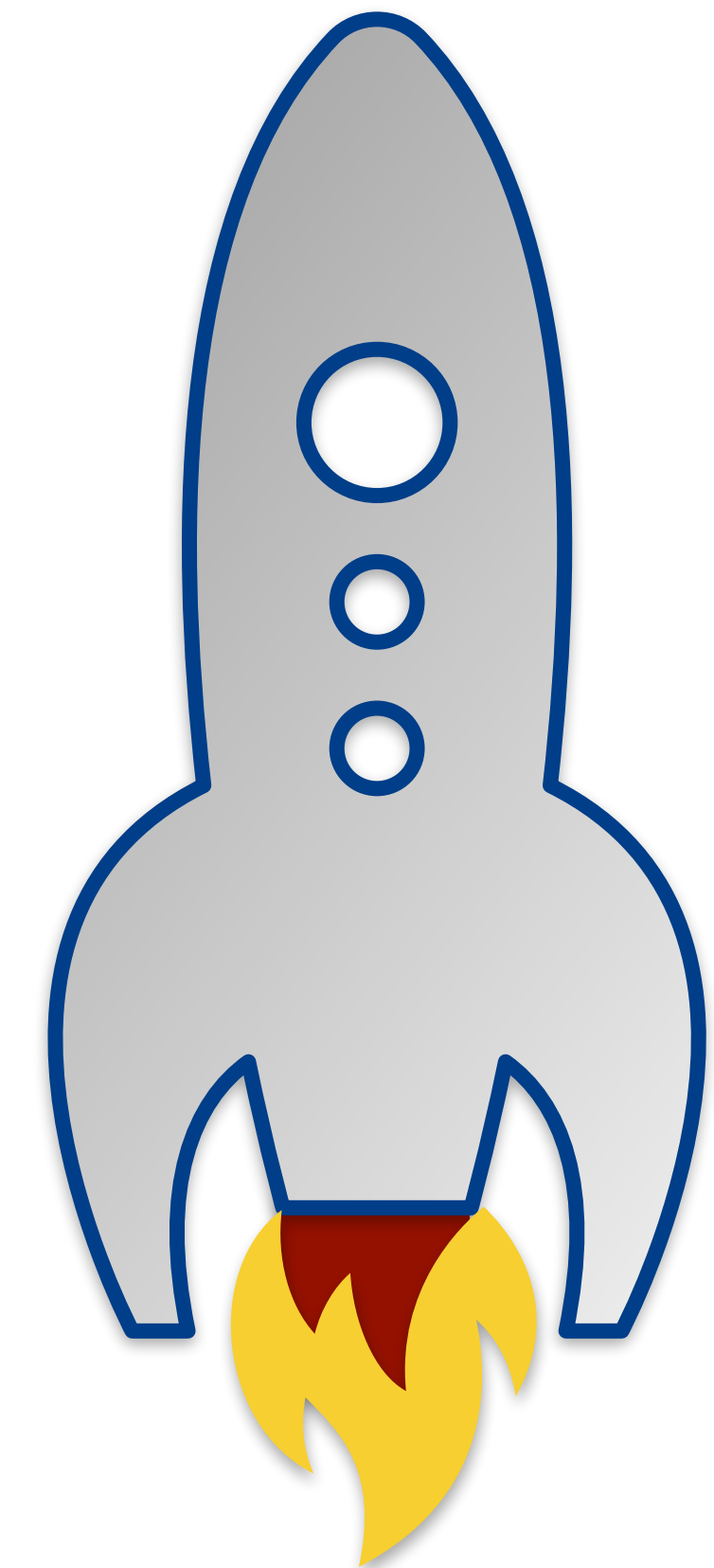


## The first year:

- In the first year we created a useable HELIPORT prototype at HZDR and HIJ.
- We developed first plugins and established our HELIPORT website with various resources and a HELIPORT demo system.
- The metadata discussion on laser metadata in the laser community successfully started.
- The HZDR cluster is accessible using UNICORE.
- ➔ We found different interested stakeholders, gave presentations (posters, talks) and walked the first steps towards a HELIPORT community.

## The next year:

- We move forward and bring further interested stakeholders into our project.
- The CWL integration for UNICORE gives us an additional workflow language for our projects TELBE and POLARIS.
- Conversion to or implementation of our experiments with HELIPORT:
  - TELBE at HZDR
  - POLARIS at HIJ





# HELIPORT and HELIPORT@TEIBE Poster in Postersession I & II

Poster Postersession Postersession I

**HELIPORT HELmholtz Scientific Project WORKflow PlatForm**  
**An Integrated Research Data Lifecycle**

Oliver Knodel, Martin Voigt, David Pape, Lokamani, Jeffrey Kelling, Stefan E. Müller, Thomas Gruber and Guido Juckeland  
 Alexander Kessler, Joachim Hein, Chien-Li Lee and Malte C. Kaluza  
 Bernd Schuller

Helmholtz-Zentrum Dresden-Rossendorf  
 Helmholtz-Institut Jena  
 Forschungszentrum Jülich

DOI 10.14278/rodare.947  
 DOI 10.5281/zenodo.7104942  
 License CC BY-4.0

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 **F... A... L... R...** and comprehensible project description.

**Full Lifecycle Management**

Submit Proposal or Create Research (Heliport) Project

Search & Reuse Datasets, Modify Experiment, Create Data Management Plan, Collect Data, Analyse Data, Archive Dataset, Publish Dataset, Change Setup

**Overview Guides Scientists through a Project**

**Project Metadata**

- 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.
- Metadata export in standardized formats and schemas (DataCite JSON/XML, RDF, JSON-LD, Turtle, ...).
- The project metadata is distributed over all relevant linked systems.

**CWL Execution and Monitoring**

- HELIPORT's encapsulation of individual steps in a computational workflow follows the FAIR principles and enables reusability.
- Analysis and pre-/post-processing steps can be documented and reproduced.

**Digital Object and Handles**

- HELIPORT interfaces with local handle.net instances (e.g. handle.hzdr.de) to enable sustainability.
- Automated generation of uniform, globally unique PIDs for digital objects of all systems, jobs, services, ...
- With digital objects, object relations and landing pages, HELIPORT improves **Provenance** and **Comprehensibility**.

**Project Resources**

- Public available documents at [heliport.hzdr.de](http://heliport.hzdr.de) including tutorials and presentations.
- Deployment guide of the HELIPORT prototype with UNICORE integration and first system plugins.
- Software and data publications of **prototype** and **project metadata** for an example project.

Funded by: HZDR, JÜLICH, HELMHOLTZ METADATA COLLABORATION, HELMHOLTZ

Poster Postersession Postersession II

**Integrated Data Workflow using HELIPORT at TELBE**

**Project Achievements**

- HELIPORT Prototype with Unicore Integration and first system plugins deployed on HZDR infrastructure for TELBE Experiment
- Presentations of the HELIPORT idea at workshops: LaserLab Europe, P-RECS, LEAPS IP, ExPaNDS, ...
- Data publications of prototype and project metadata.

**Work In Progress**

- Discussions on specific metadata support (TELBE)
- User Authentication using LDAP/HelmholtzAAI
- Full integration in to HELIPORT

**Stage 1: Data Acquisition at TELBE | UNICORE Workflow**

- Data acquisition using LABVIEW on local computer at TELBE
- Metadata on sensors acquired during the experiment
- UNICORE workflow assembly from templates
- Job dispatch

**Stage 2: UNICORE Job Execution on HPC-HEMERA**

UNICORE Gateway UNICORE X Server  
 UNICORE TSI HPC-HEMERA

- UNICORE job processed by UNICORE Gateway and X Server
- User Authentication via LDAP or OIDC
- UNICORE job execution by UNICORE TSI as Slurm job

**Stage 3: Metadata Management by HELIPORT**

- UNICORE Job triggers metadata acquisition by HELIPORT
- Overview of processed workflows
- Status and Restart options

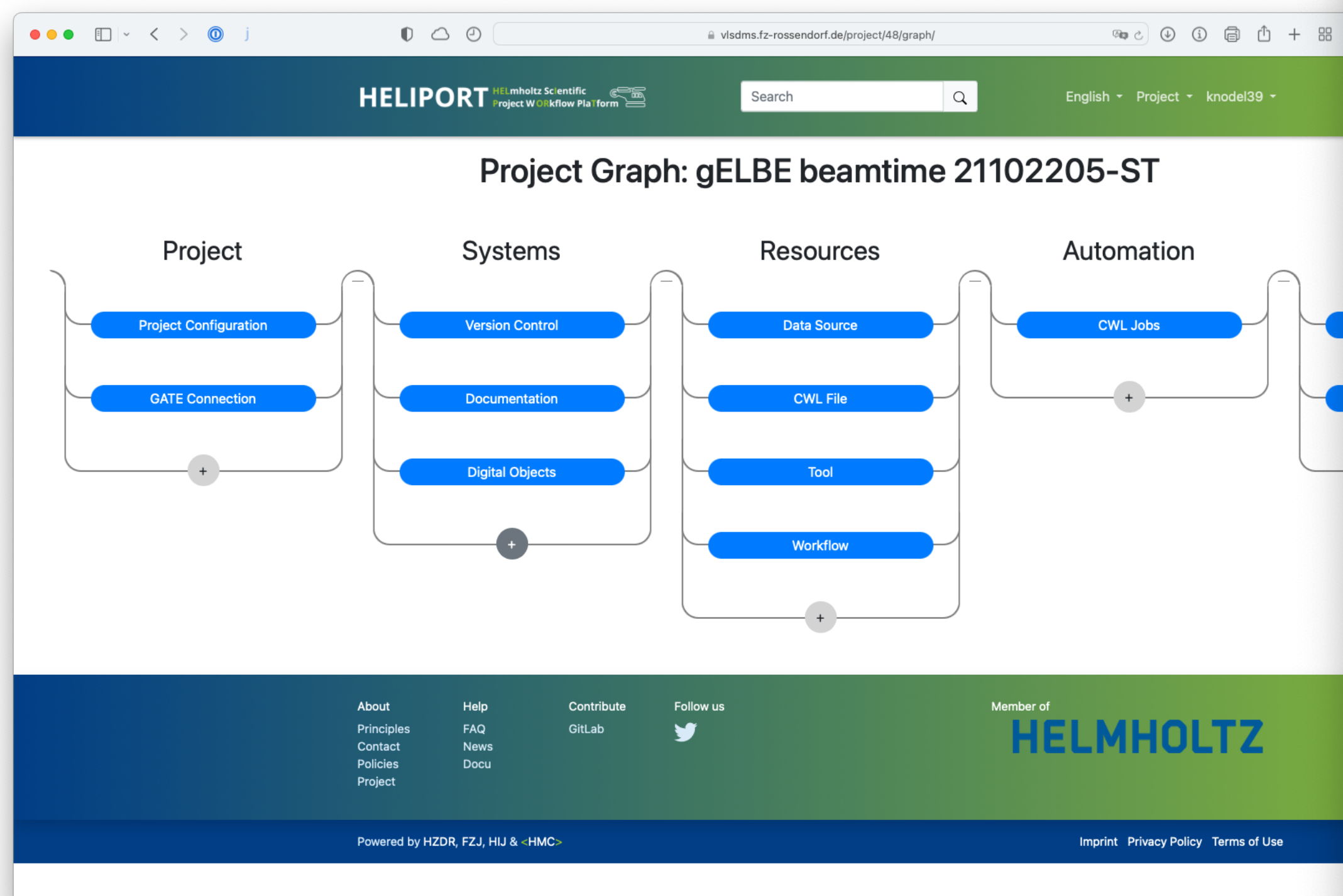
DOI 10.5281/zenodo.7054583

HZDR: Lokamani, Martin Voigt, David Pape, Thomas Gruber, Jeffrey Kelling, Jan-Christoph Deinert, Stefan Mueller, Oliver Knodel and Guido Juckeland



# Resources

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



Demo: [heliport.hzdr.de/app](https://heliport.hzdr.de/app)

**HELIPORT** HELmholtz Scientific Project WOrkflow PlaTform

The guidance system HELIPORT aims to make the entire life cycle of a project at the HZDR findable, accessible, interoperable and reusable according to the FAIR principles. mentioned below. In particular, our data management solution deals with the areas from the general... the actual research results. For this... their connections. Descriptions...

Workshop Presentation **DOI 10.1145/3456287.3465477**

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

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

**ABSTRACT**  
Modern scientific collaborations and projects (MSCPs) employ various processing stages, starting with the proposal submission, continuing with data acquisition and concluding with final publications. The realization of such MSCPs poses a huge challenge due to (1) the complexity and diversity of the tools, (2) the heterogeneity of various involved computing and experimental platforms, (3) flexibility of analysis targets towards data acquisition and (4) data throughput. Another challenge for MSCPs is to provide additional metadata according to the FAIR principles for all processing stages for internal and external use. Consequently, the demand for a system, that assists the scientist in all project stages and archives all processes on the basis of metadata standards like DataCite to make really everything transparent, understandable and citable, has risen

**1 INTRODUCTION**  
An essential objective of modern cutting-edge research should be to enable accessibility of the acquired research data and its re-usability across different research fields and their respective communities. The current generation of scientists is therefore faced with the challenging task of transferring experimental investigations into a data oriented research flow with strong focus on documenting every step closely following the FAIR [41] principles. The FAIR principles are well-established as standards in the field of research data management. The three pillars F (Findable), A (Accessible) and

Intuitive and structured user interface