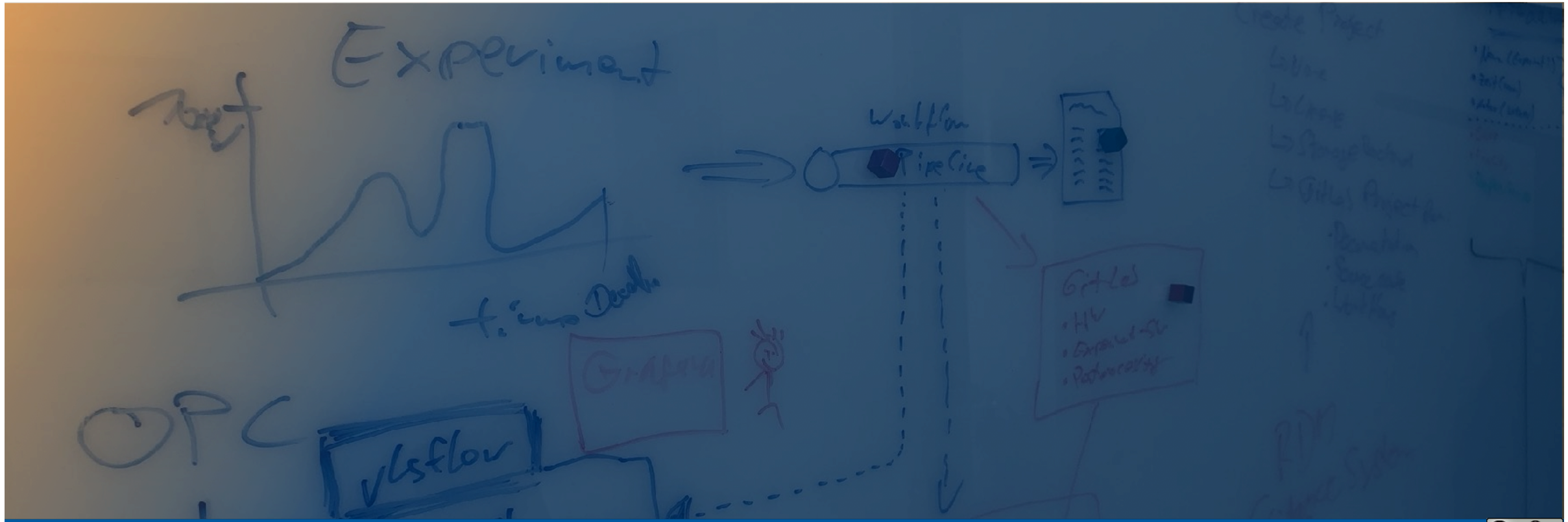
HELIPORT

HZDR
HELMHOLTZ ZENTRUM
DRESDEN ROSSENDORF



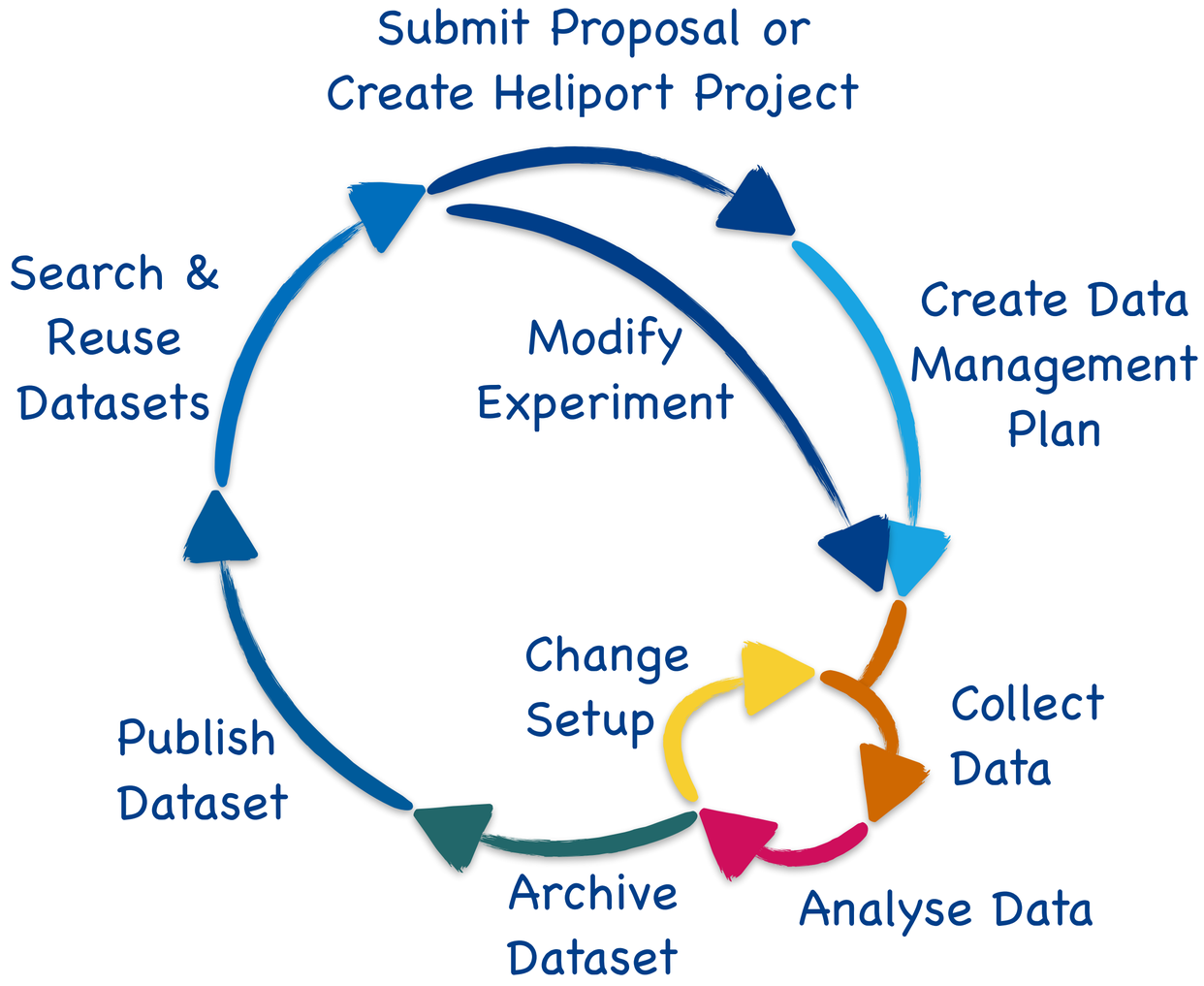
Full Integrated Research Data Lifecycle – The Project HELIPORT

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

DRESDEN concept

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.



…and is Your Research Data Usable?

Make your data

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

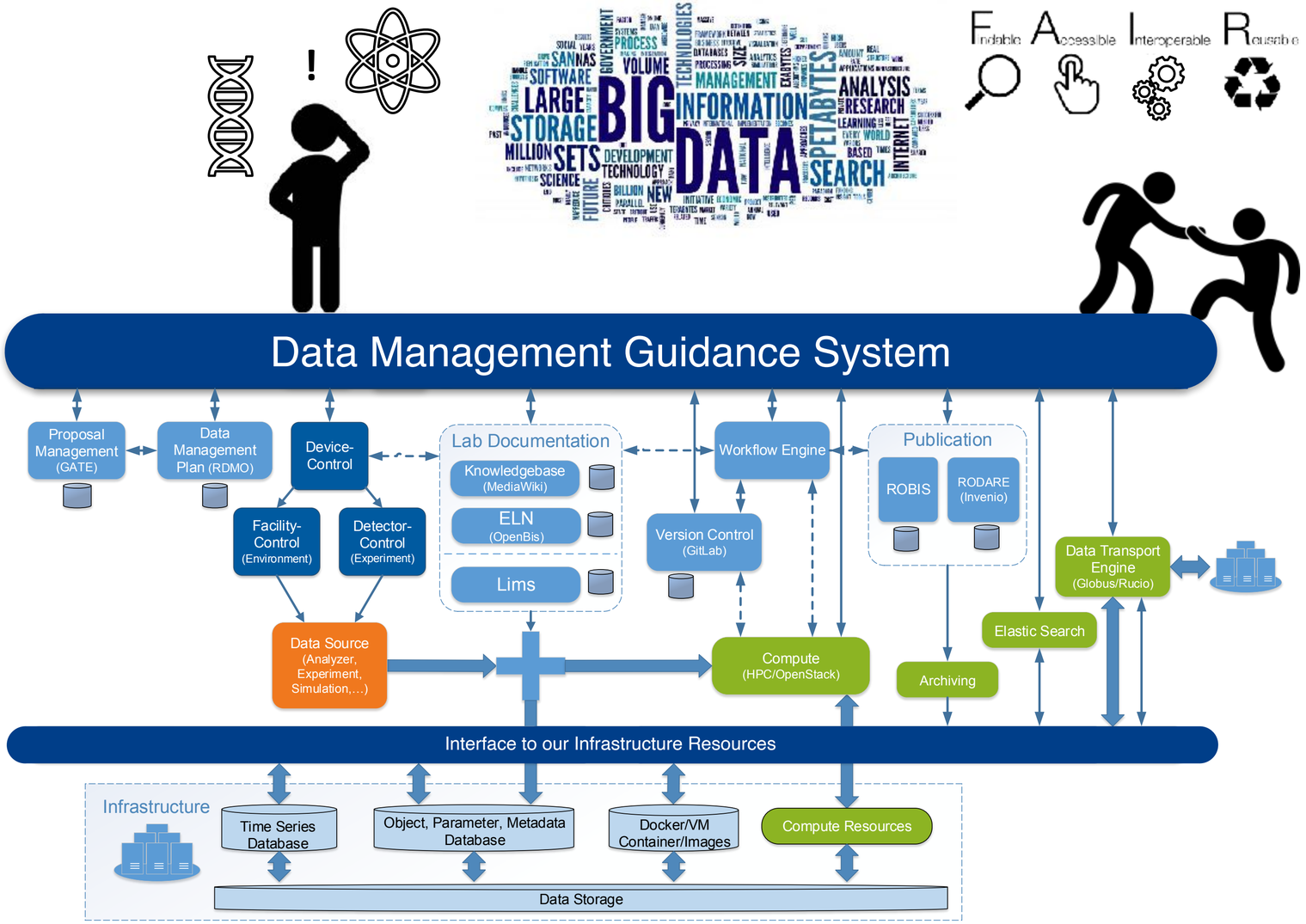
FAIR data principles

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).

RODARE DRESDEN concept HZDR

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.



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?

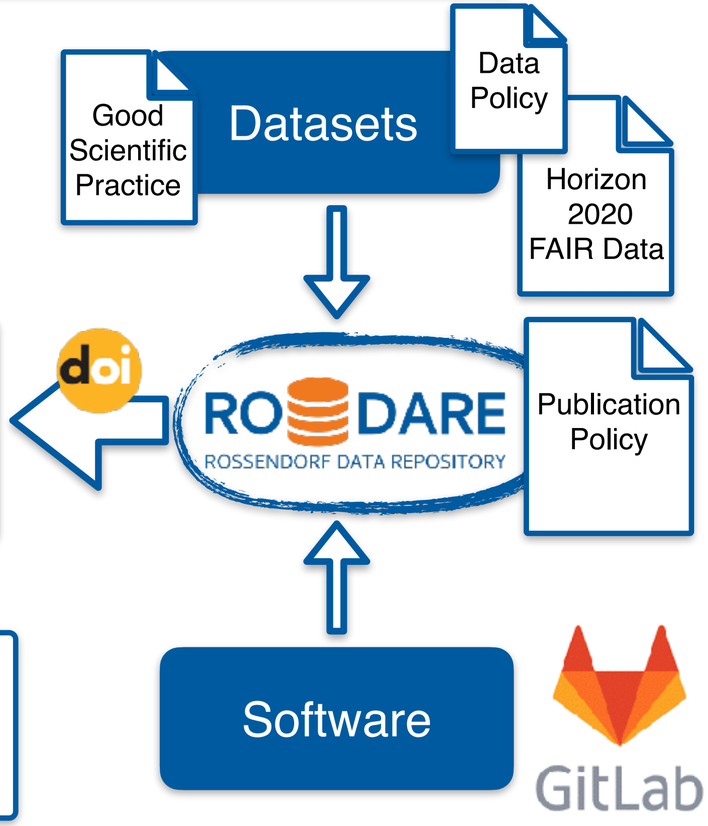
And how we can support them?!

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?

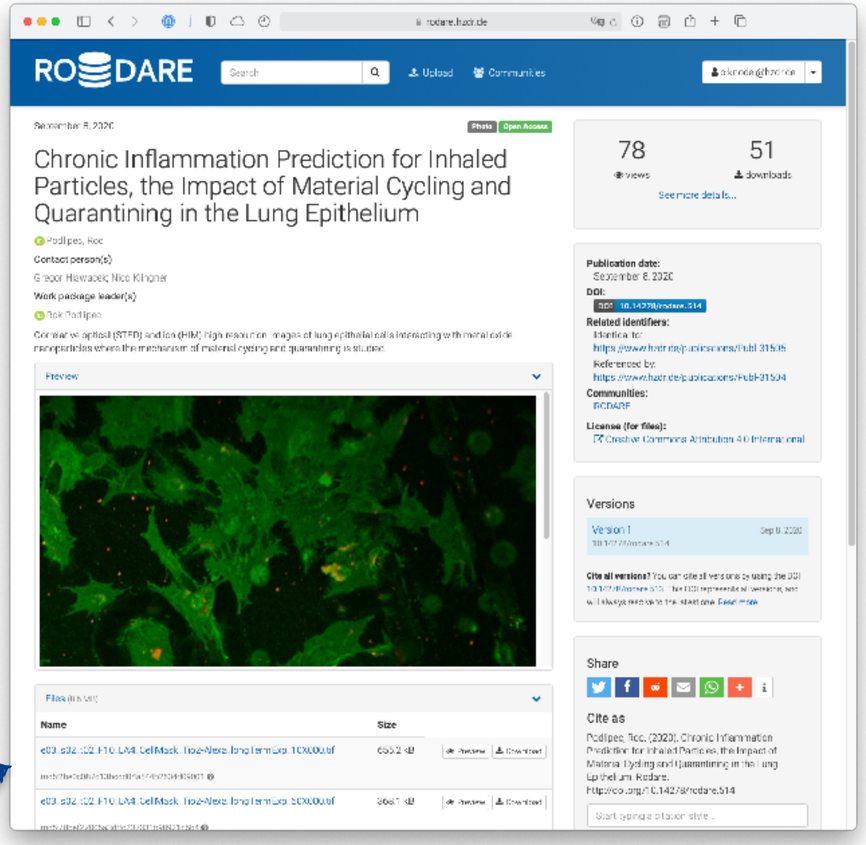
Where are data, software and how can I gain **access** to both of them?



DRESDEN concept 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:

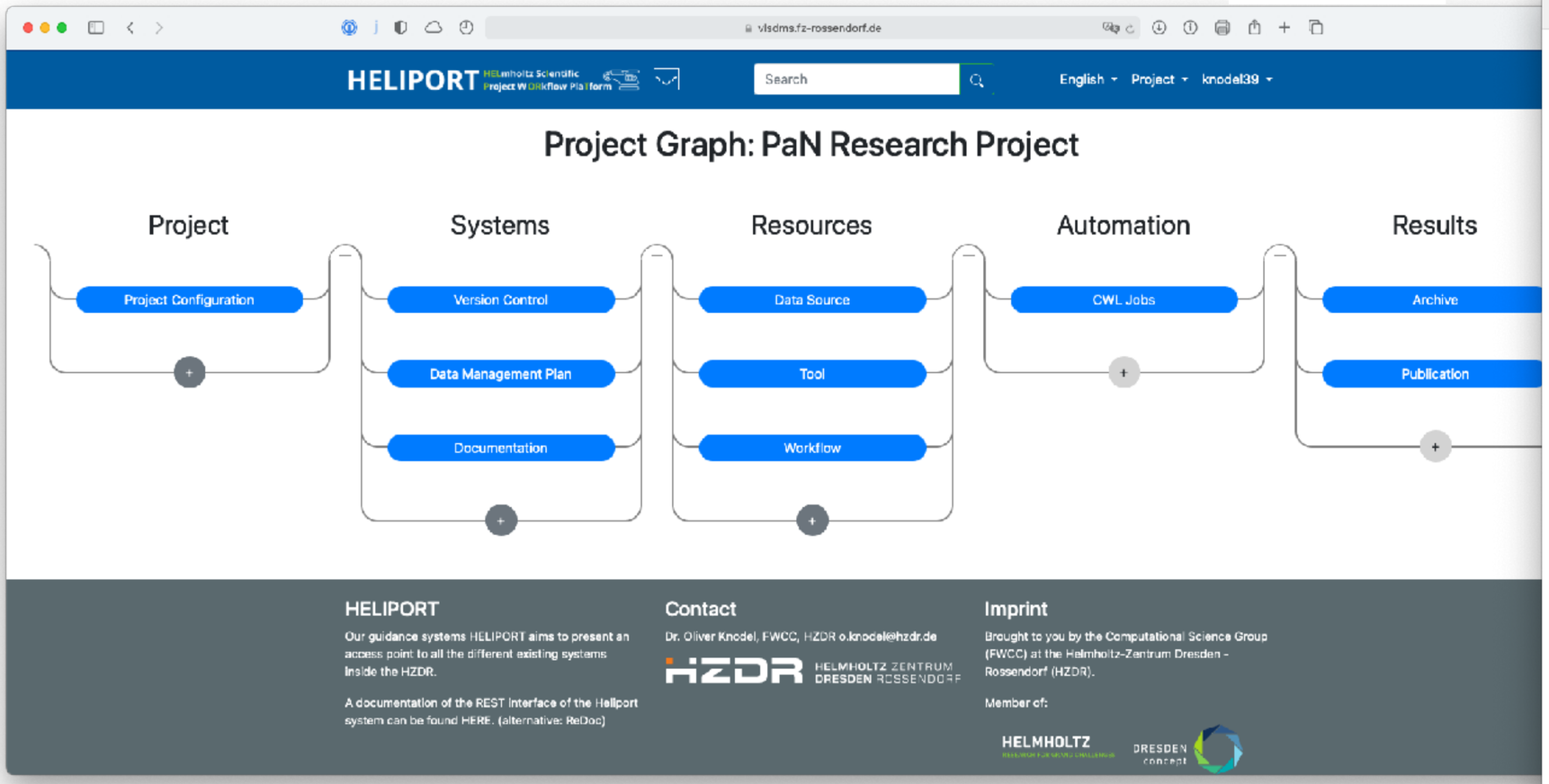
HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

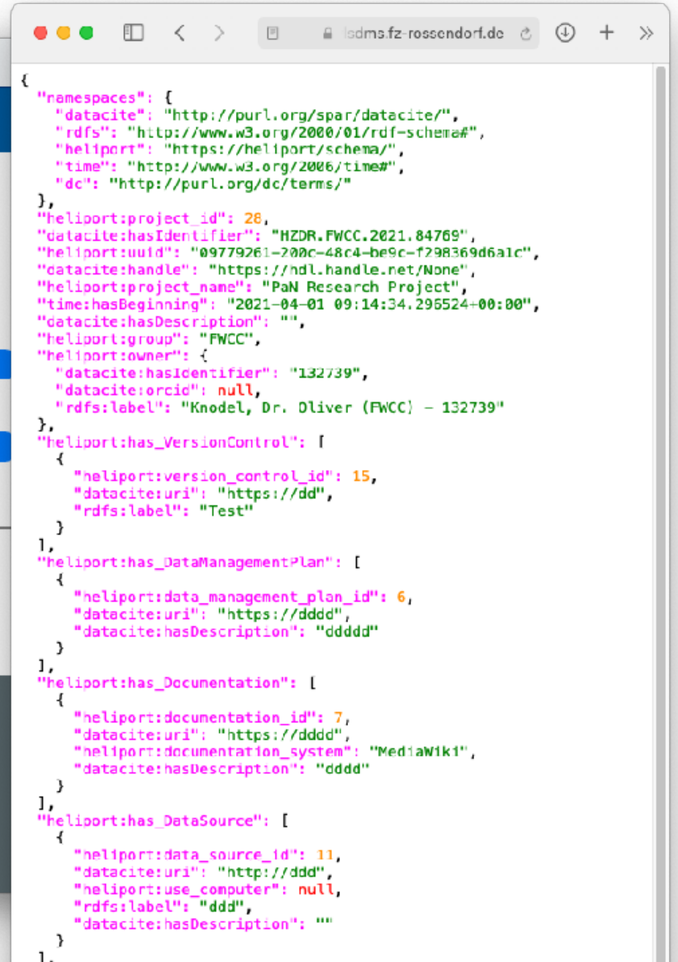
HI JENA Helmholtz Institute Jena

JÜLICH FORSCHUNGSZENTRUM

Founded by:

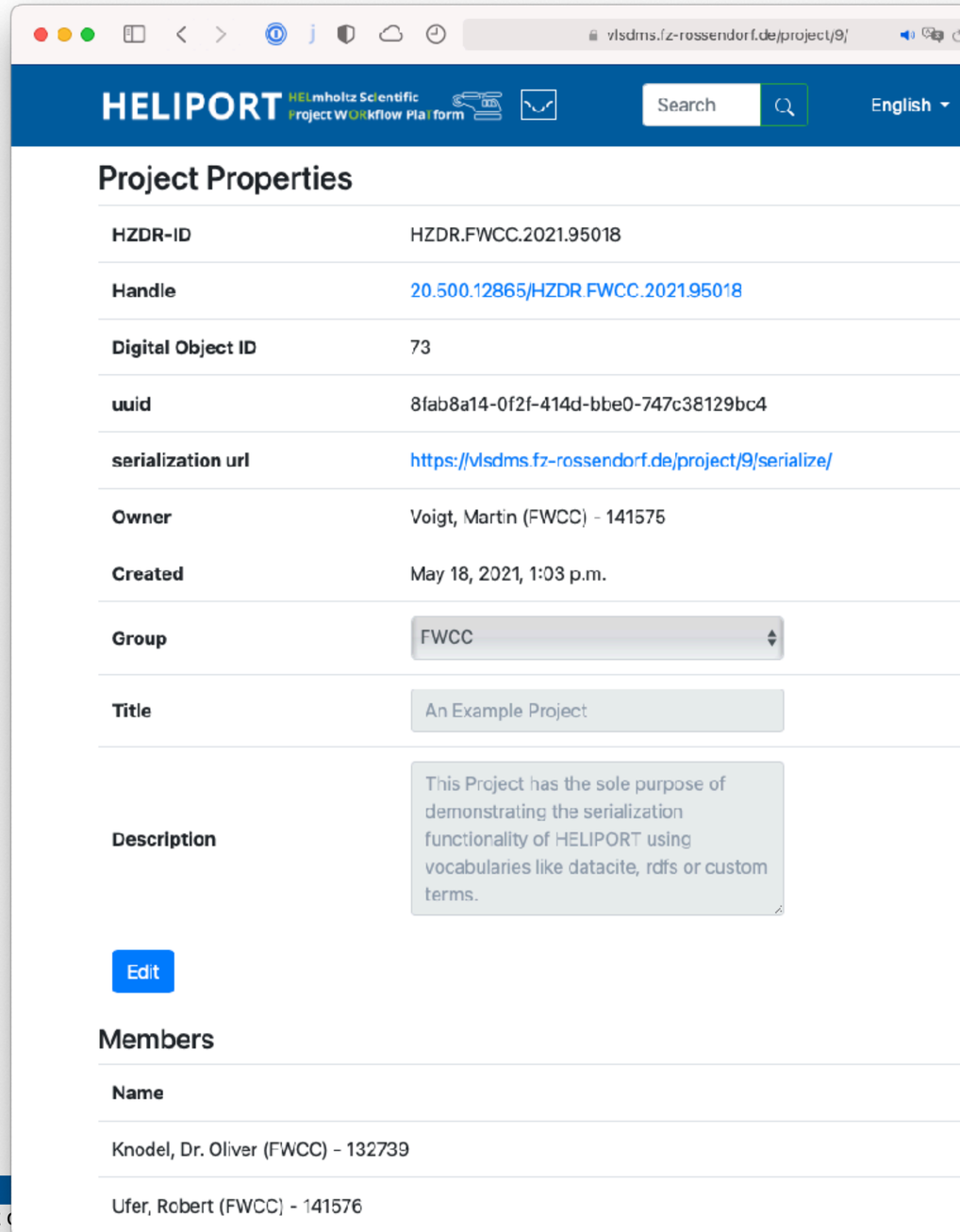
HMC HELMHOLTZ METADATA COLLABORATION



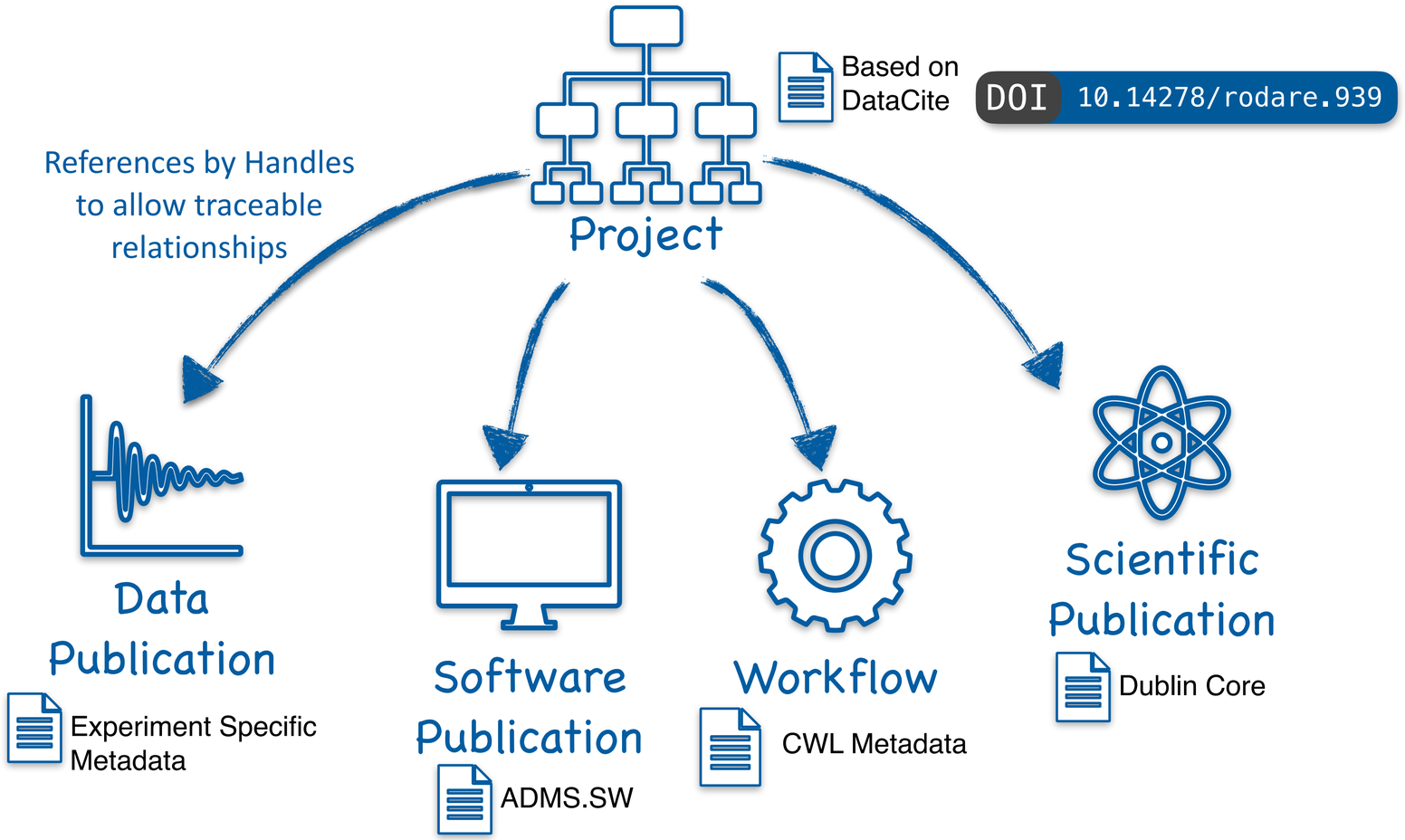


The Heliport Metadata Scheme


```
{
  "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": {
        . . .
      },
. . .
```

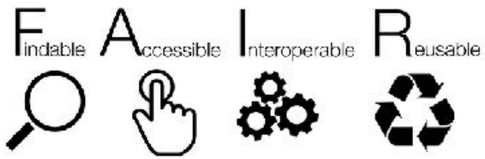



Heliport Metadata Ecosystem



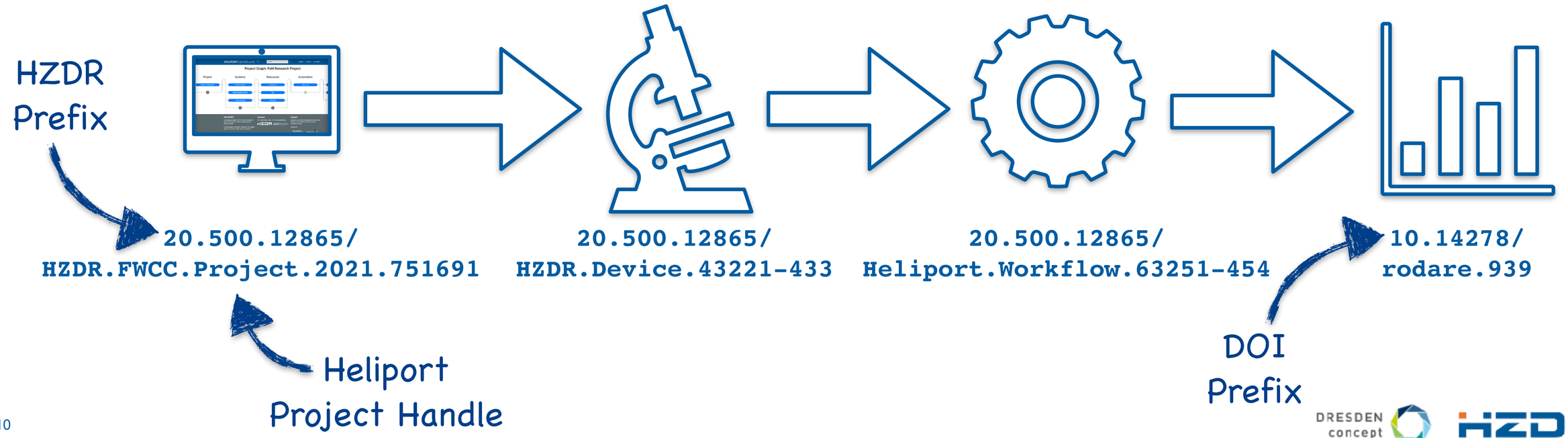
Our Objective

- 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.
- In the end every digital object should have an uniform PID, describing metadata in an open and widely used format to be



Handle Management Support 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.

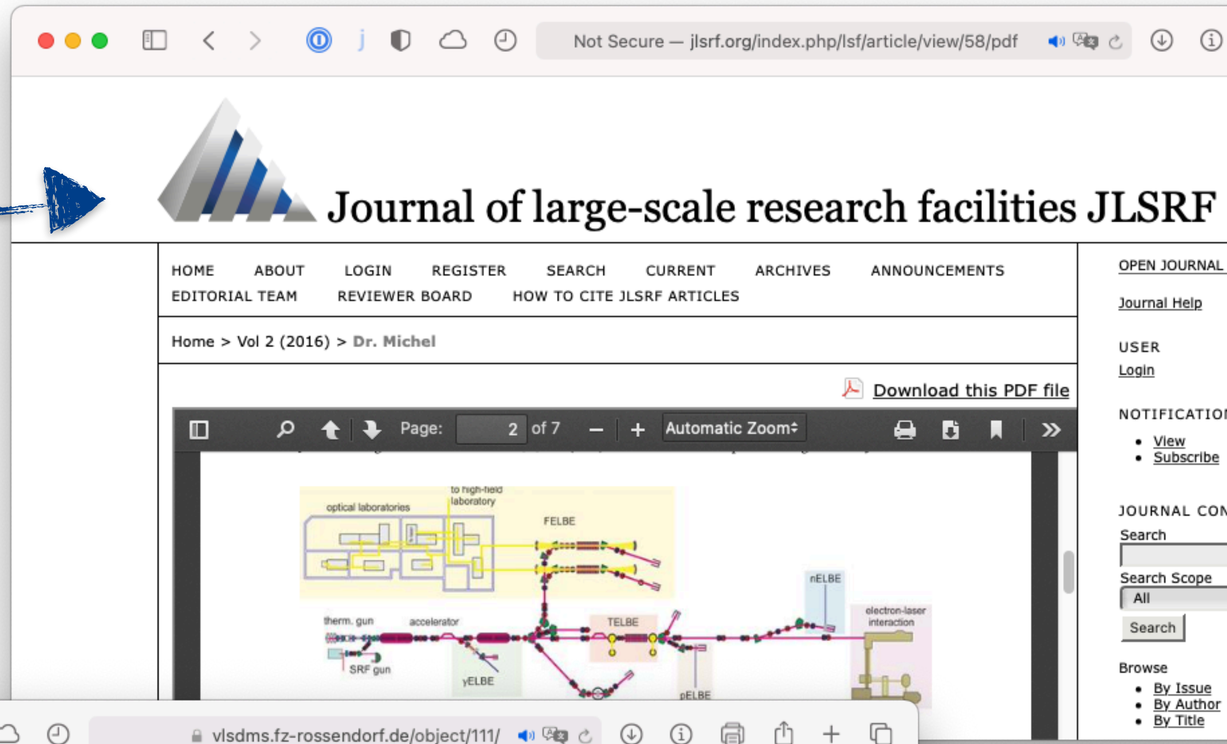


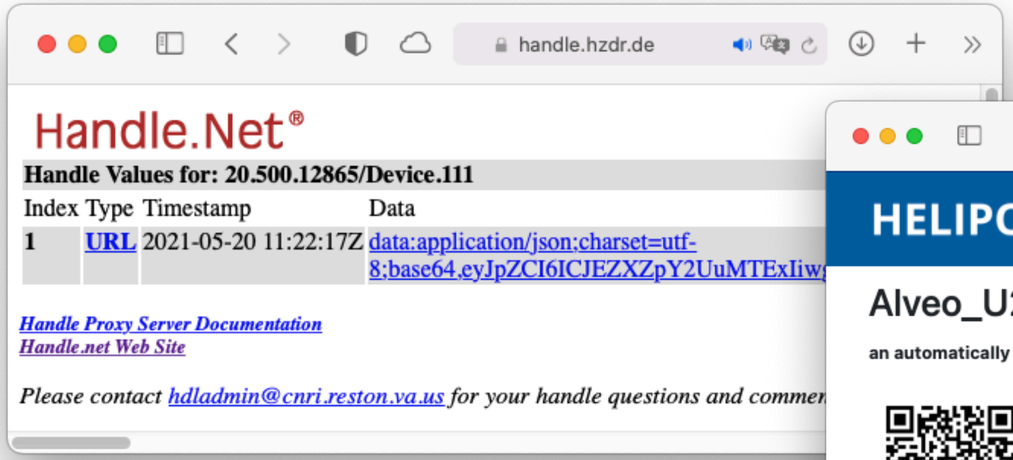
The Role of Handles

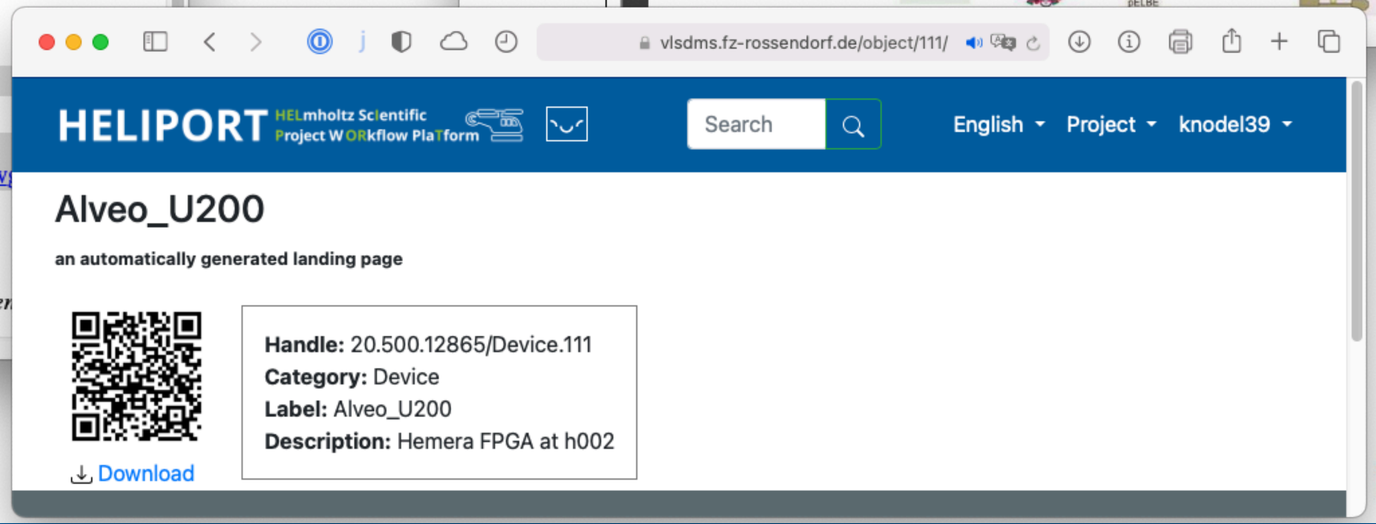
— We provide doi s only for data publications (in our data repository) and our large scale facilities.

— Our *internal* Handles are generated in Heliport (automated or by hand) and have **external** and **internal** (Heliport) landing pages.

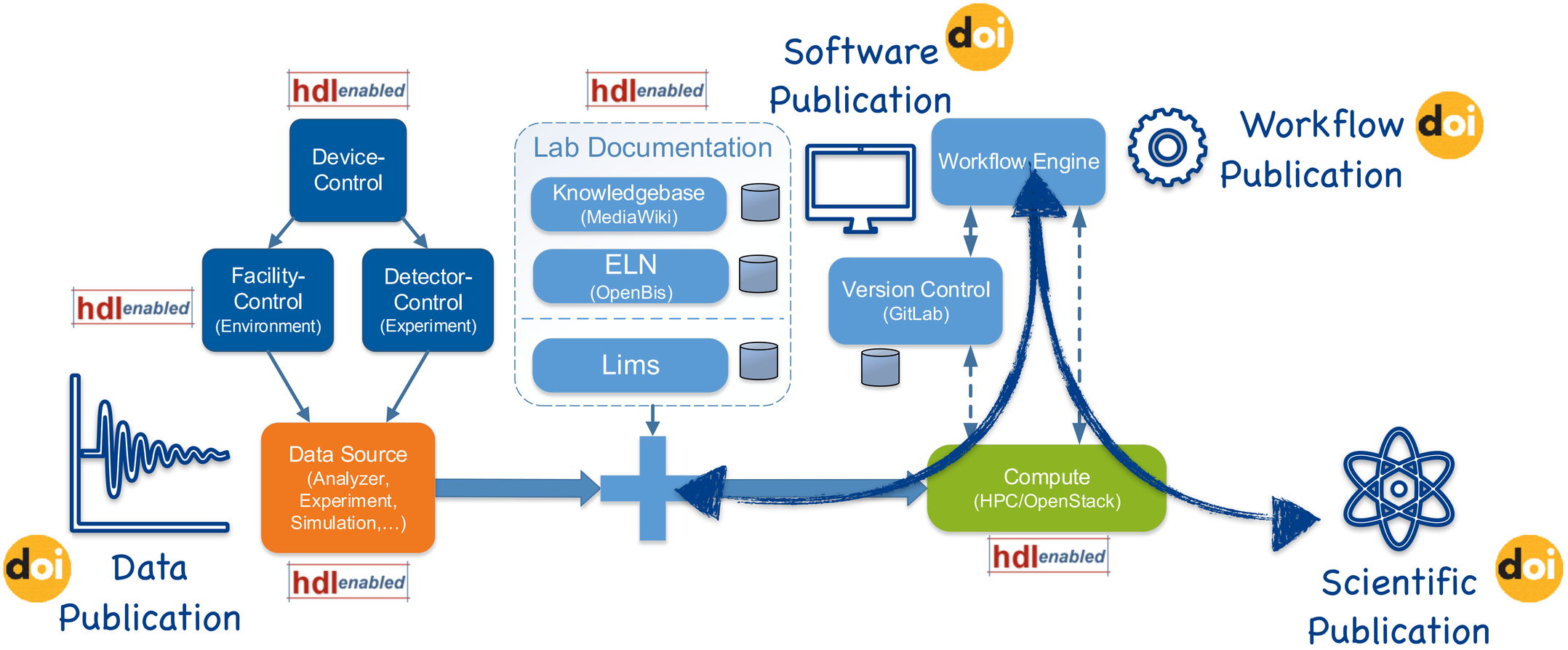
ELBE Center for High-Power Radiation Sources DOI 10.17815/jlsrf-2-58





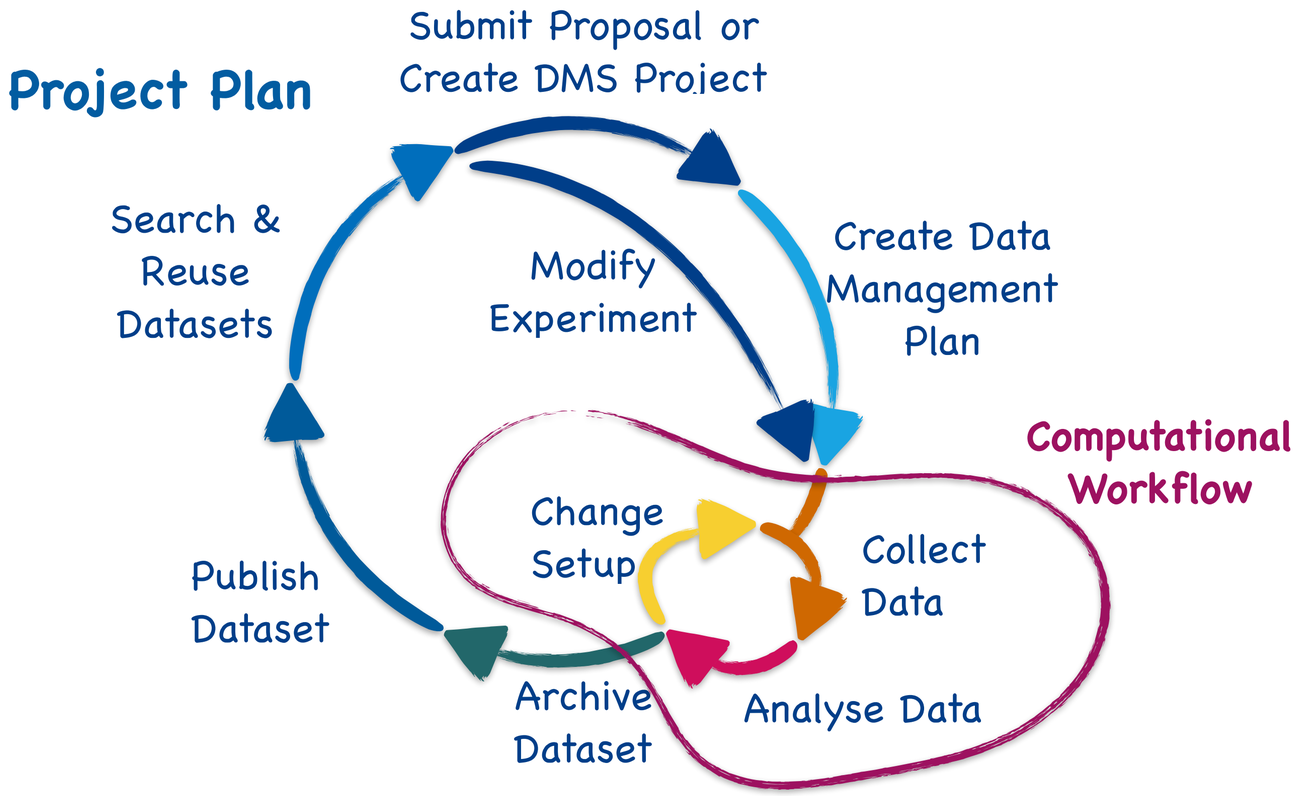


Different Types of Publications based on our Infrastructure

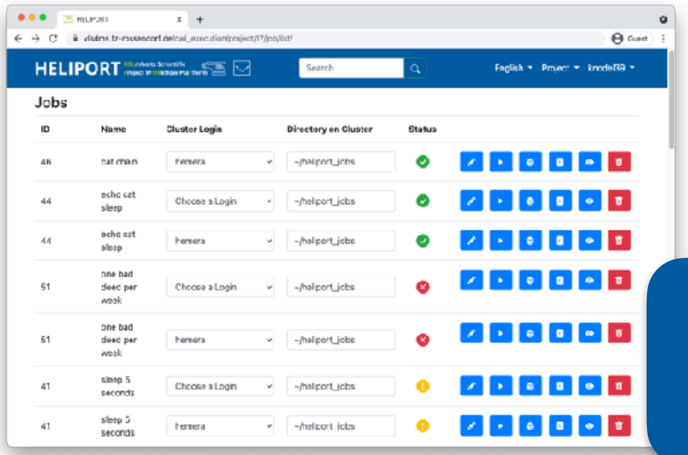


HELIPORT has a build-in Integration of Computational Workflows

- We choose the Common Workflow Language (CWL) to describe our workflows.
- Heliport is intended to fill the gap between:
  - The workflow itself and the surrounding project information and data locations,
  - Software versions and the generated particular data products.
- Computational Workflows can be:
  - User specific analysis jobs used during the experiment,
  - Recurring background jobs in the pre- and post-processing of the experiment.



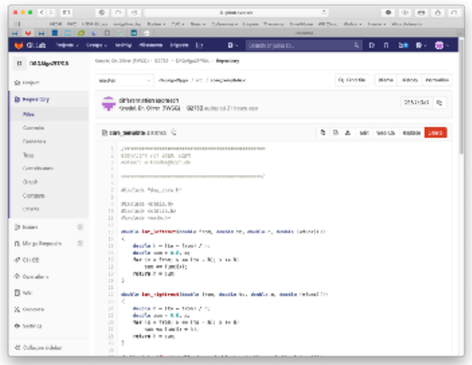
Scientific Software Development and Reproducible Workflows



Workflow Engine

Version Control

Compute (HPC, OpenStack)



GitLab

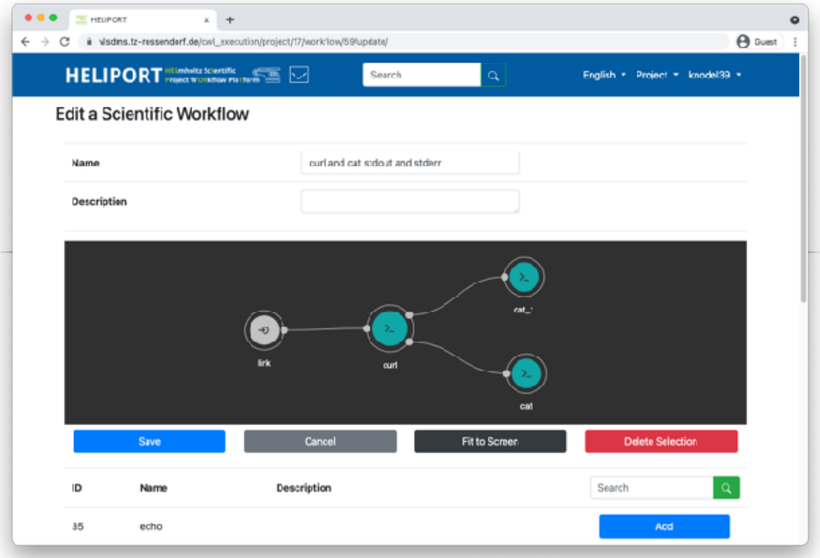
slurm workload manager

UNICORE

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

COMMON WORKFLOW LANGUAGE

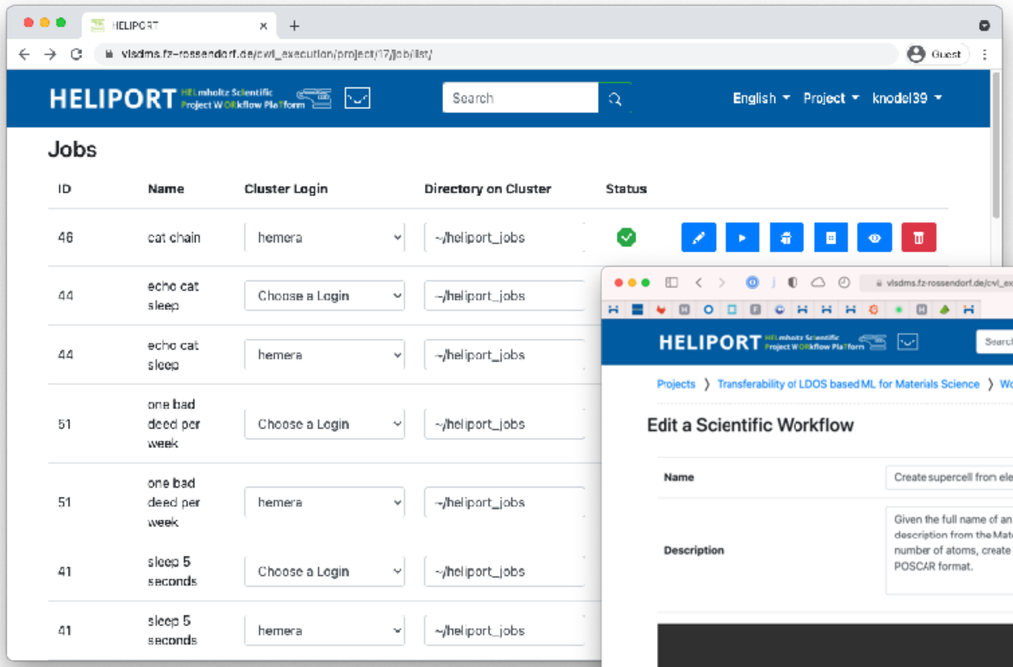
— Capsuling every step in a workflow adapts the **FAIR** principles.

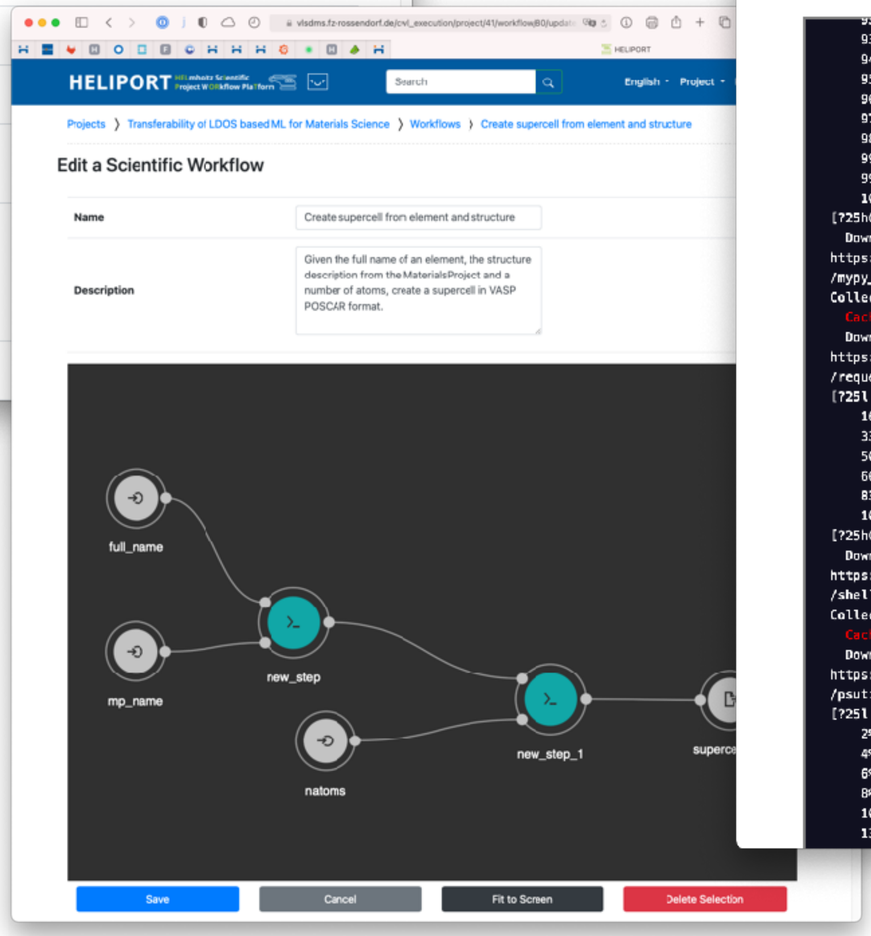


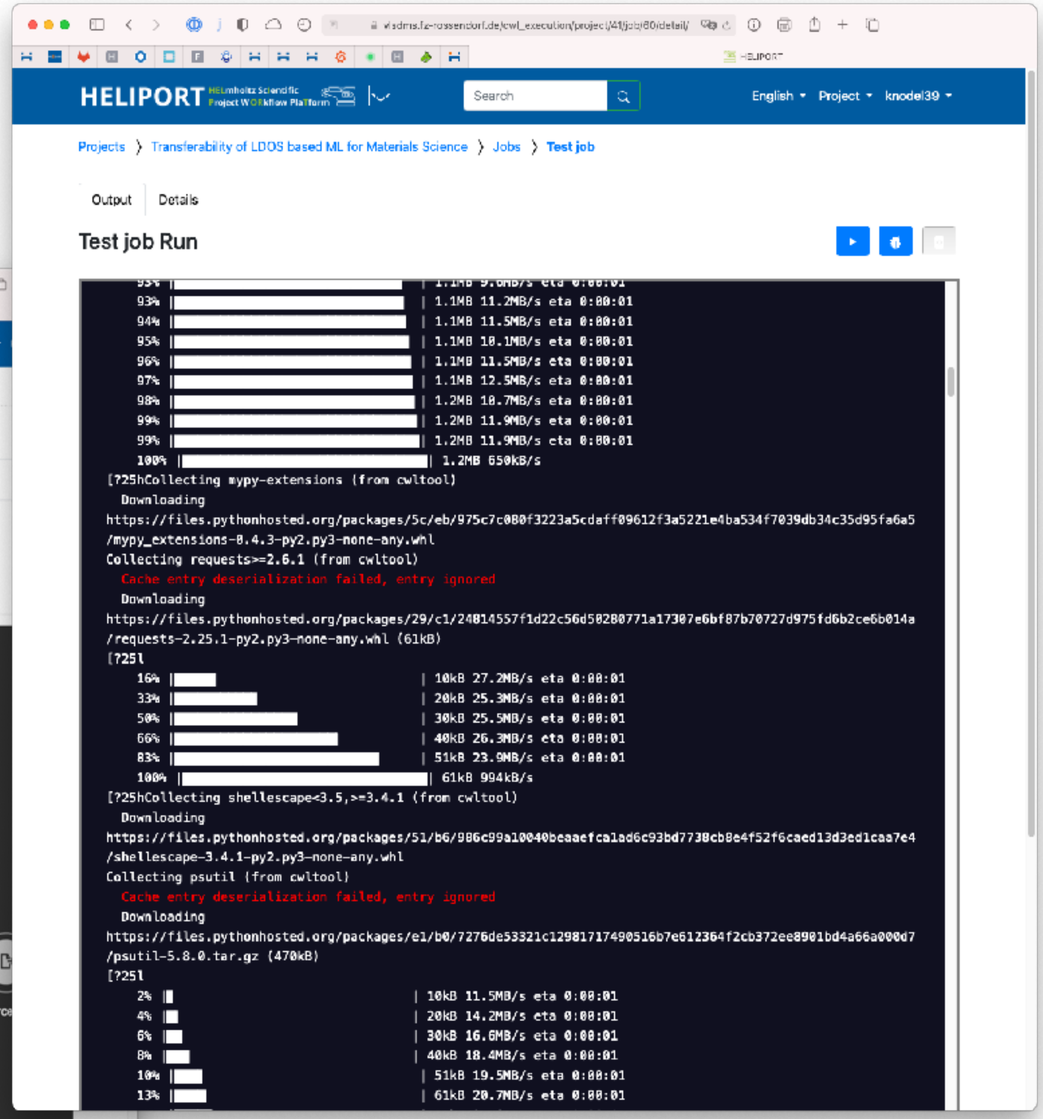
DRESDEN concept

HZDR

Heliport Tracks Workflows, their Execution and Metadata

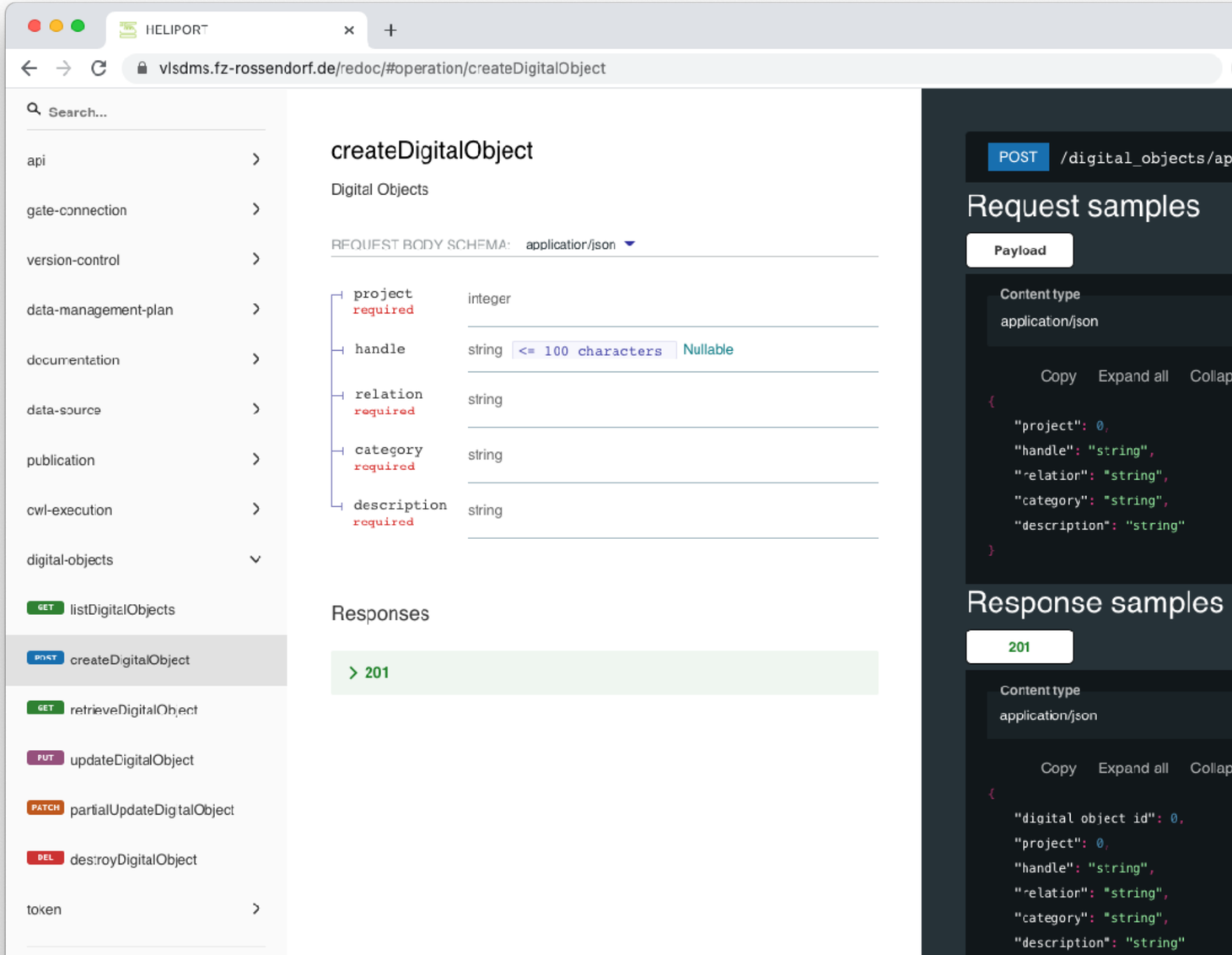






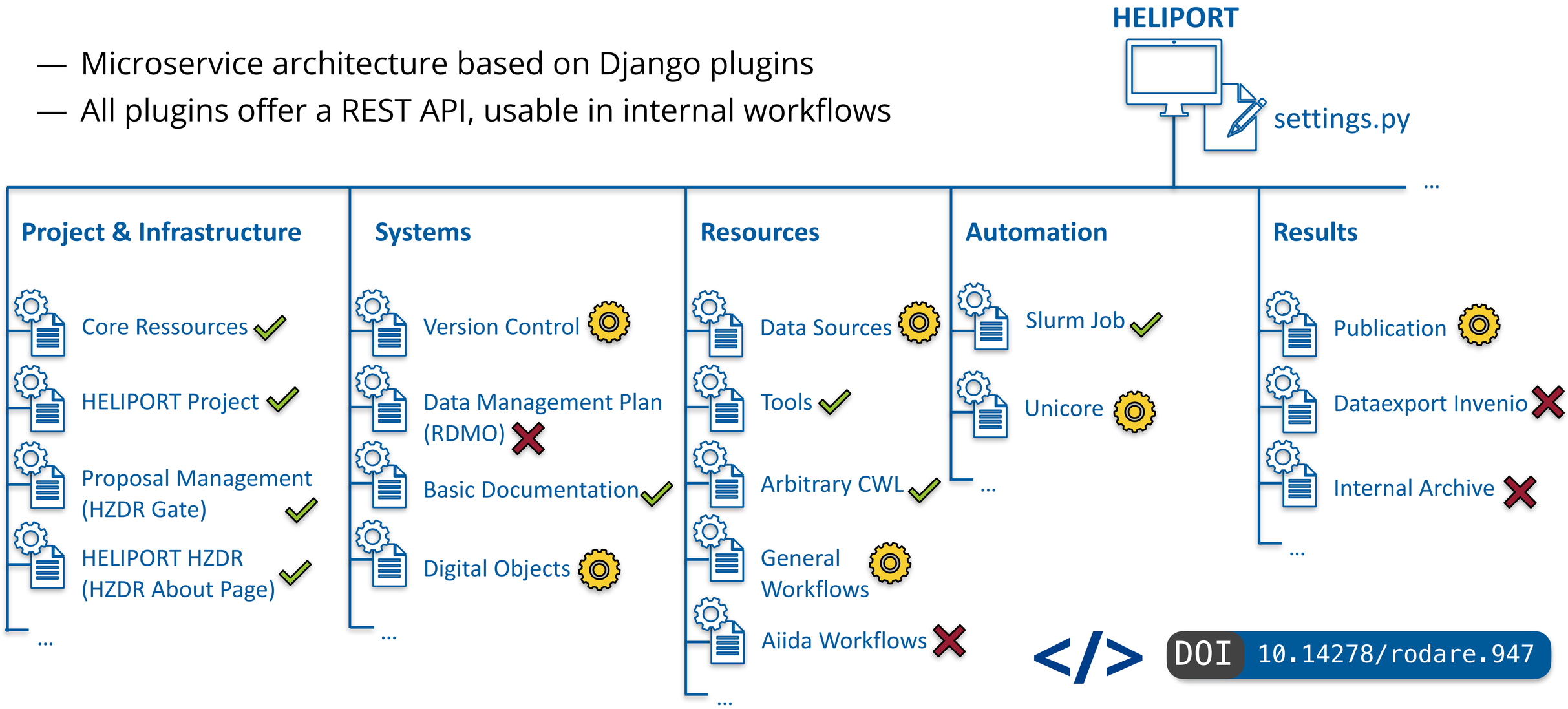
Heliport REST API

— 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.



Modular HELIPORT Design (Django Apps)

- Microservice architecture based on Django plugins
- All plugins offer a REST API, usable in internal workflows

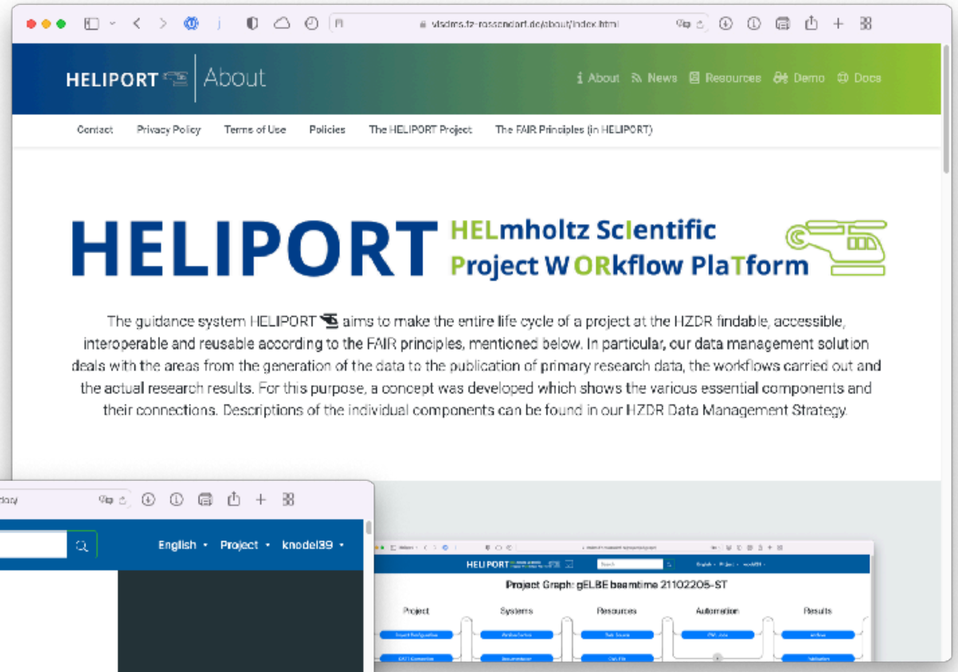


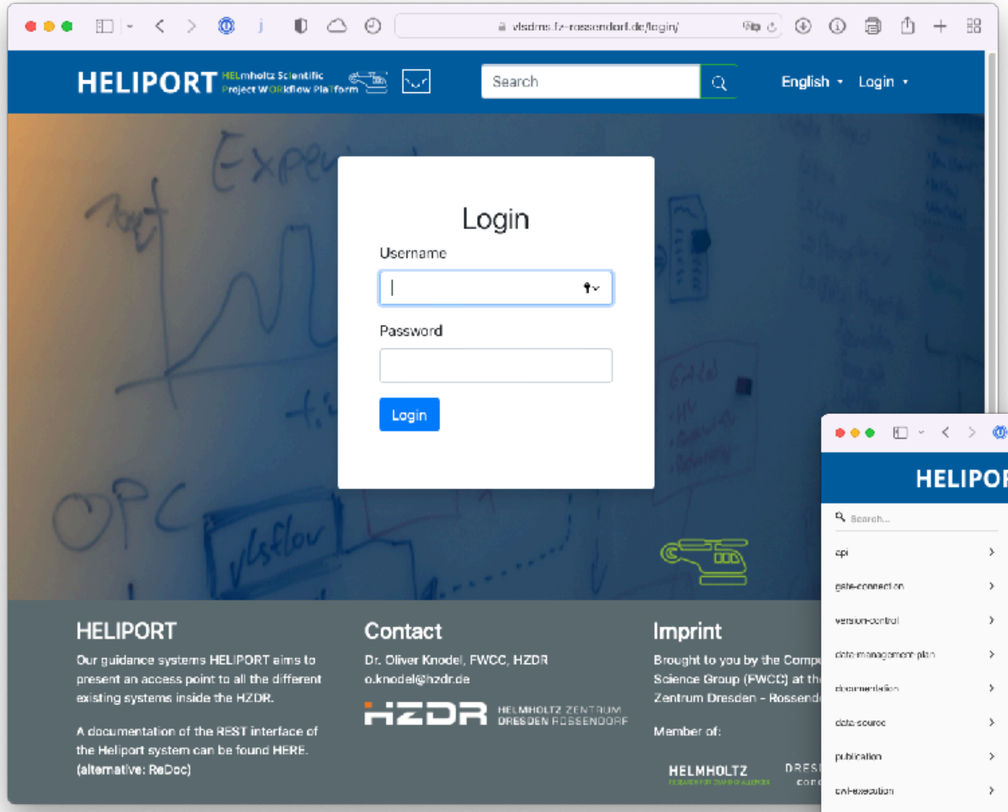
</> DOI 10.14278/rodare.947

HELIPORT JÜLICH FORSCHUNGSZENTRUM HI JENA Helmholtz Institute Jena HZDR

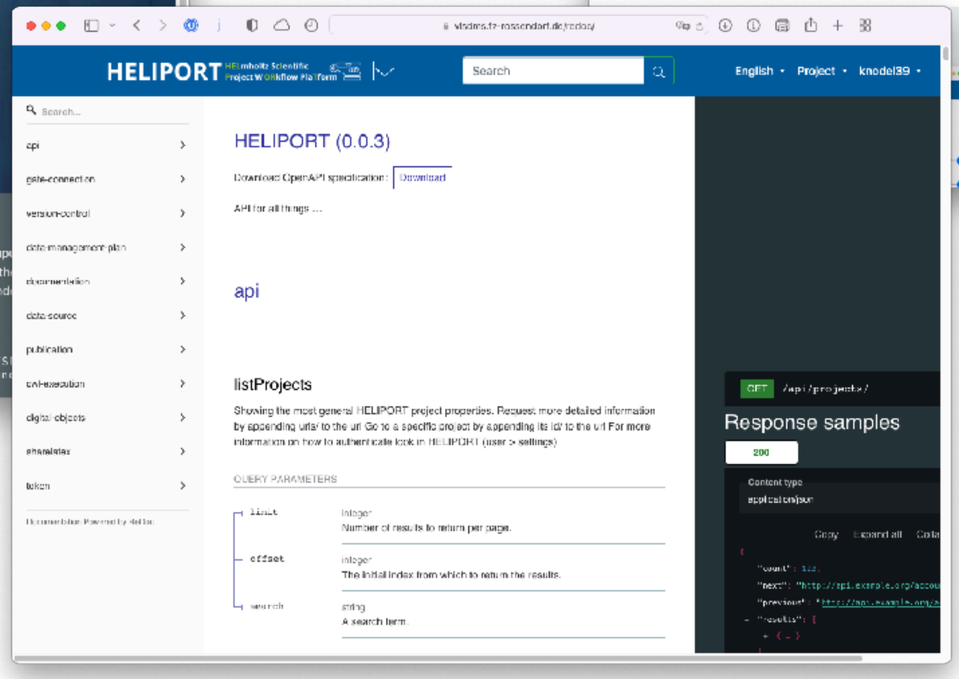
Resources (internal HZDR network)

vlsdms.fz-rossendorf.de/about





vlsdms.fz-rossendorf.de



vlsdms.fz-rossendorf.de/redoc

API Example

Generate API Token

Create a Token

On a terminal:

```
curl –X POST https://vlsdms.fz-rossendorf.de/token/ –d
“username=HZDR_USER&password=HZDR_PASSWORD"
```

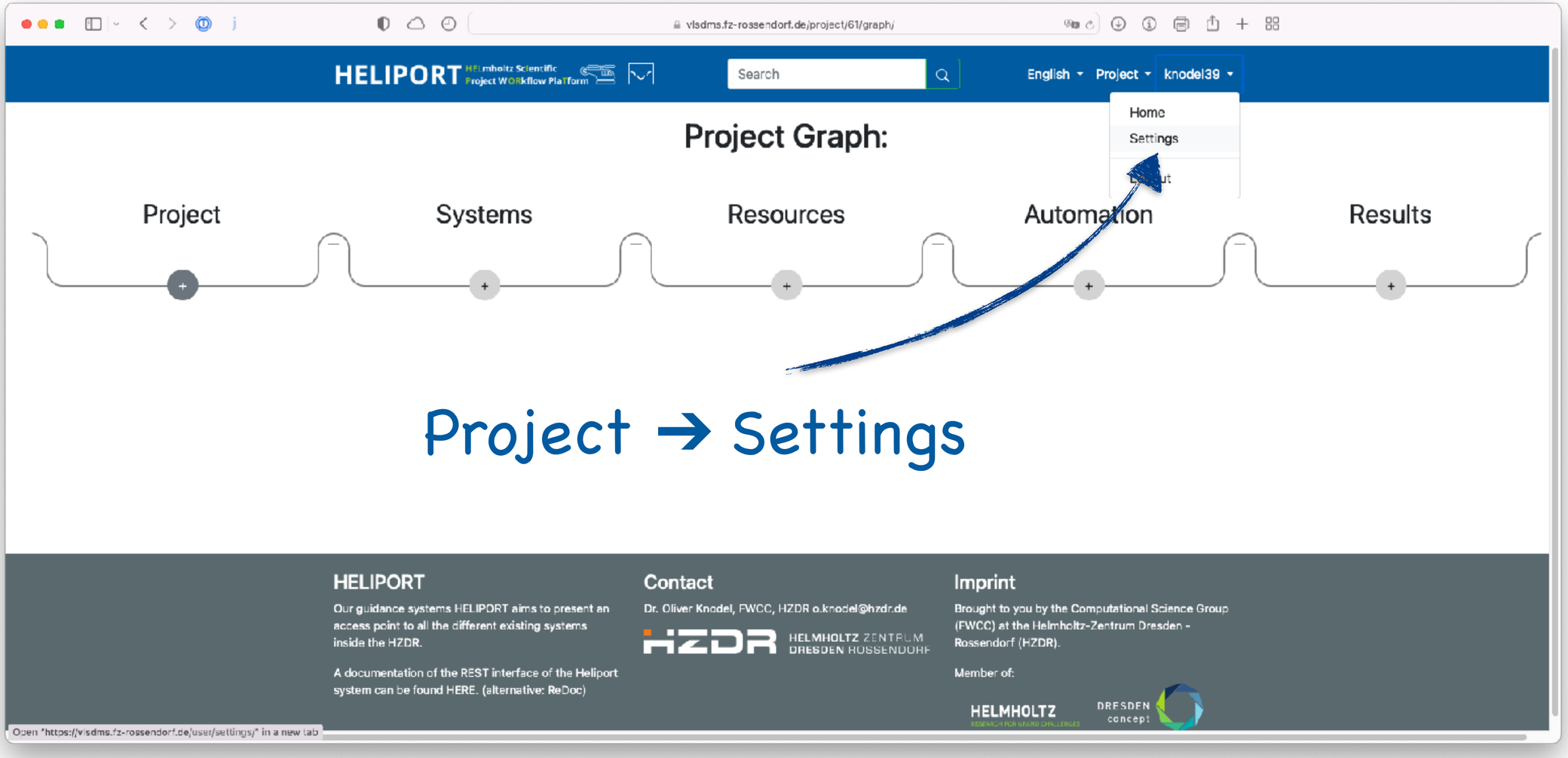
The response should look like this:

```
{“token":"9ff3ddjuiikl3l2mnn56pghsw6mwqh"}
```

Read gate projects

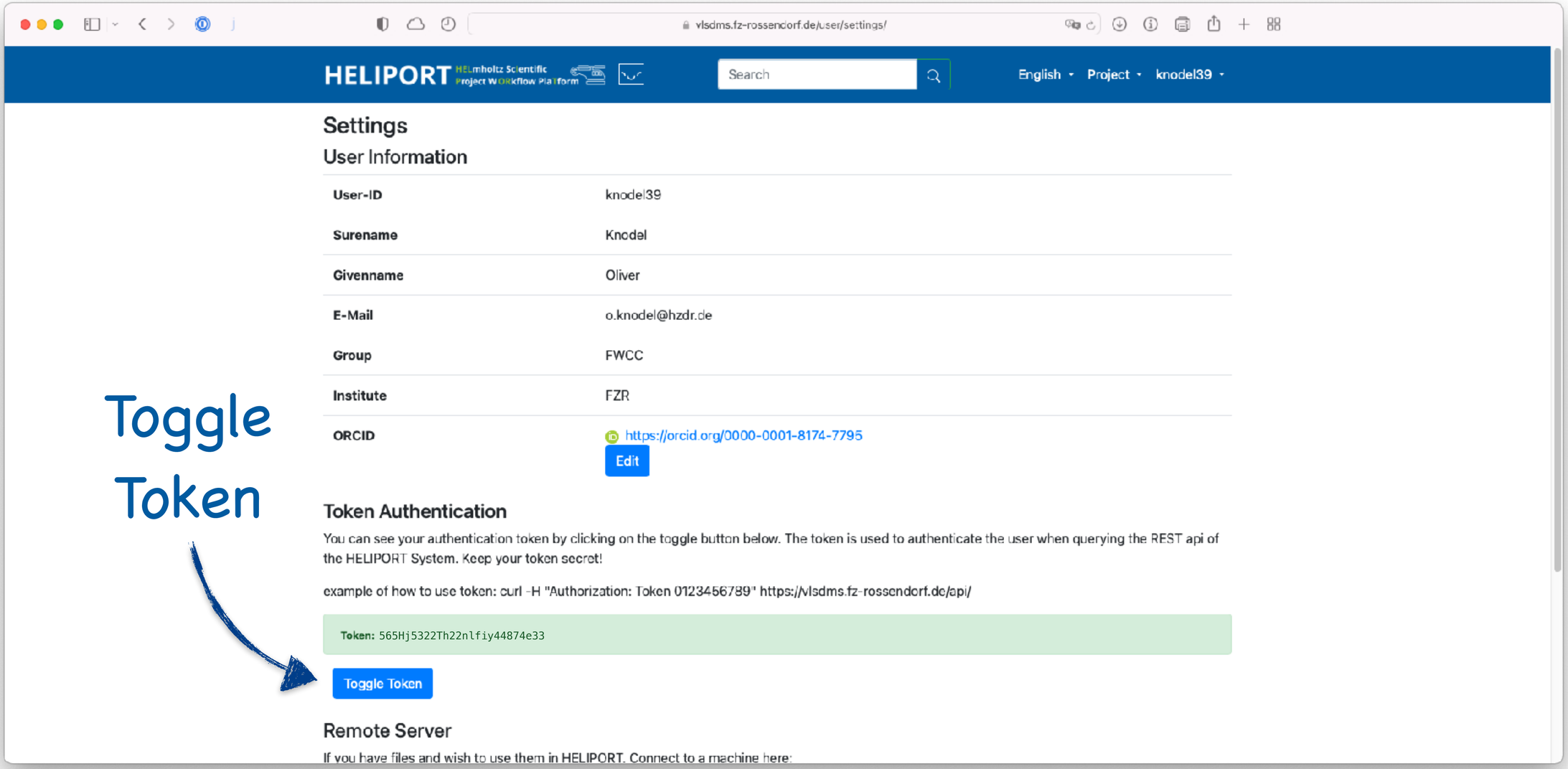
```
curl –H "Authorization: Token YOUR_TOKEN" https://vlsdms.fz-
rossendorf.de/gate_connection/api/gate_project/
```

REST API Token



DRESDEN concept HZDR

REST API Token



Toggle Token