

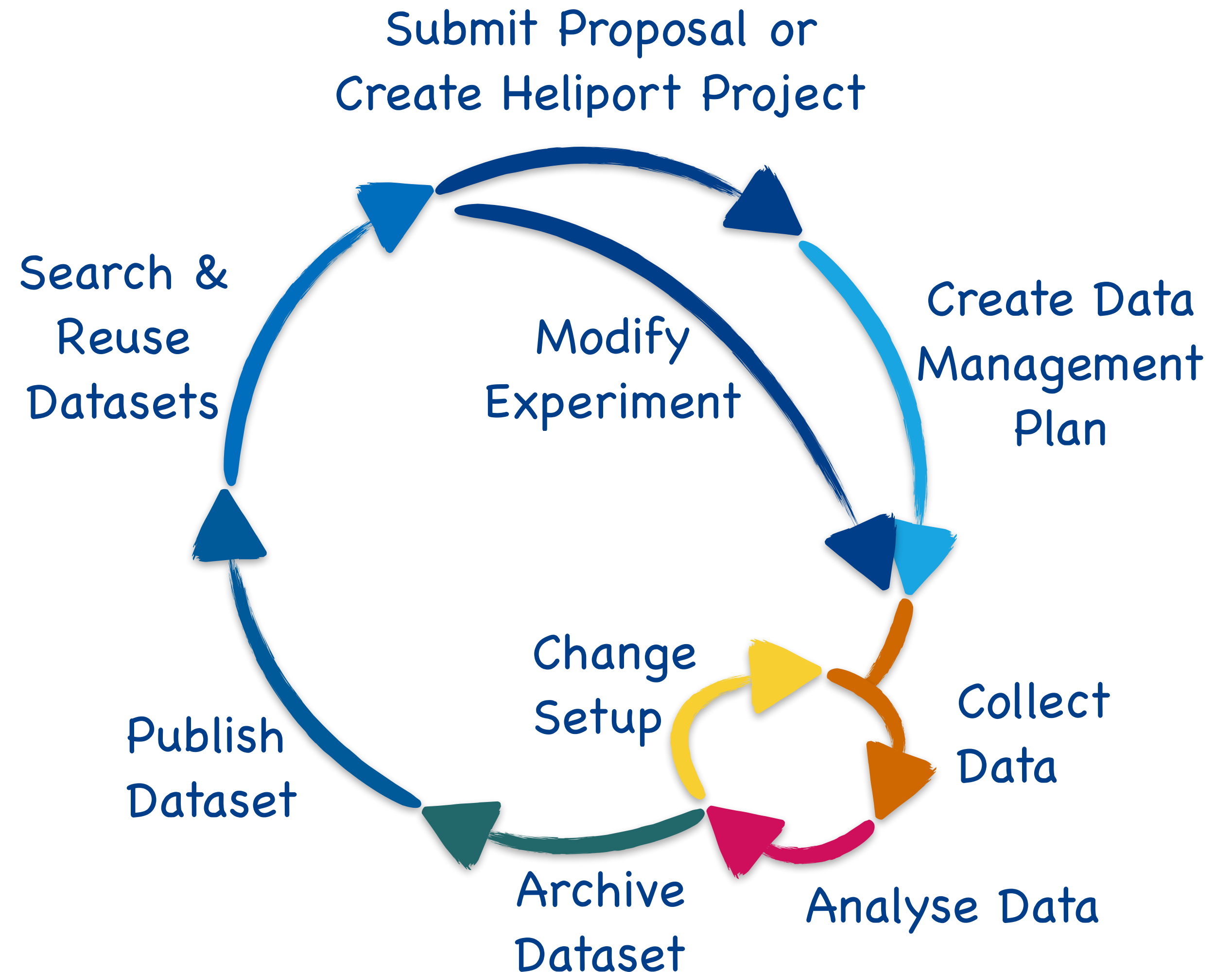
# Full Integrated Research Data Lifecycle – The Project HELIPORT

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



# 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

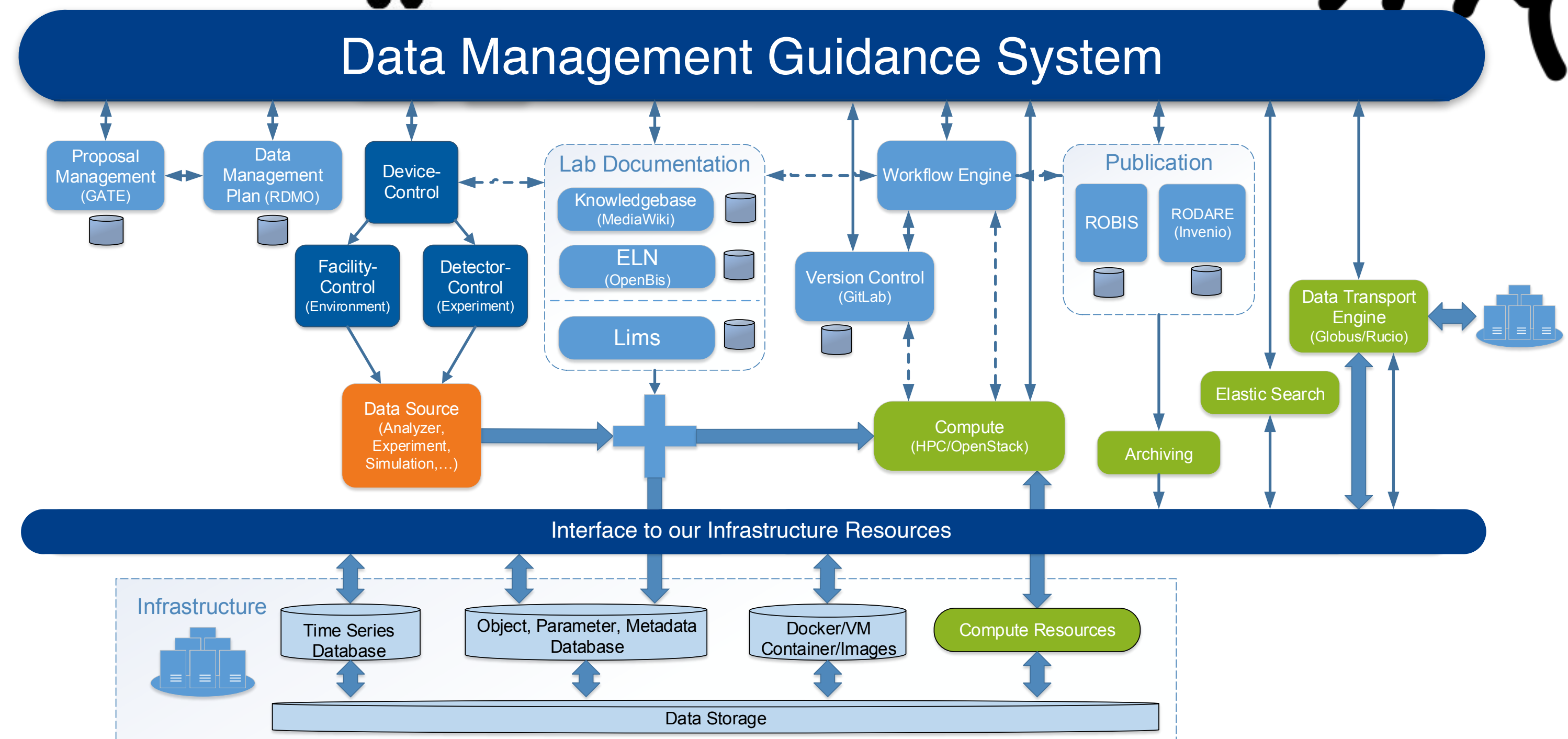
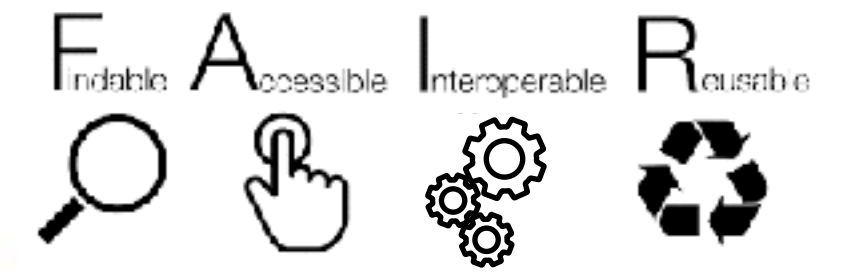
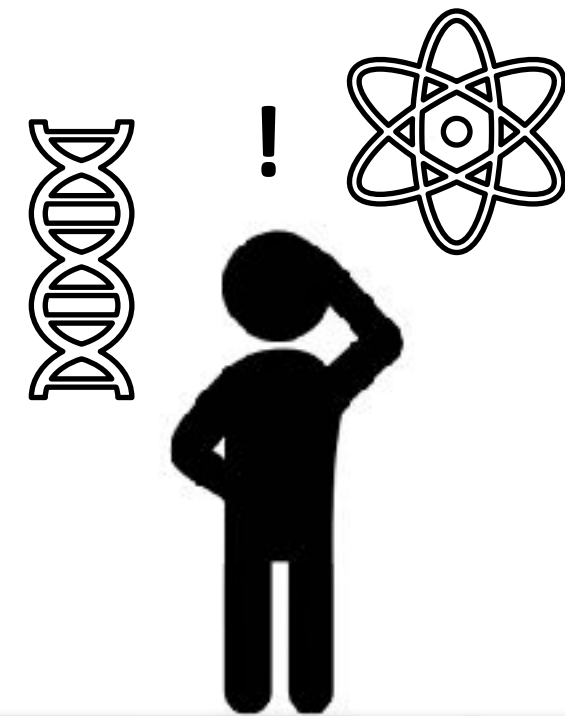
**F**indable, **A**ccessible, **I**nteroperable and **R**e-usable

FAIR data principles

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

# Our Observations and Experiences

- Our HZDR IT infrastructure can support various experiments, but it is complex...
- Scientists often don't know which services are available and how to use them.
- An overarching system guiding our scientists (and visitors) through the lifecycle of their research project (and our services) is inevitable.
- The concept of FAIR research becomes an important topic for our scientists.



# 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?

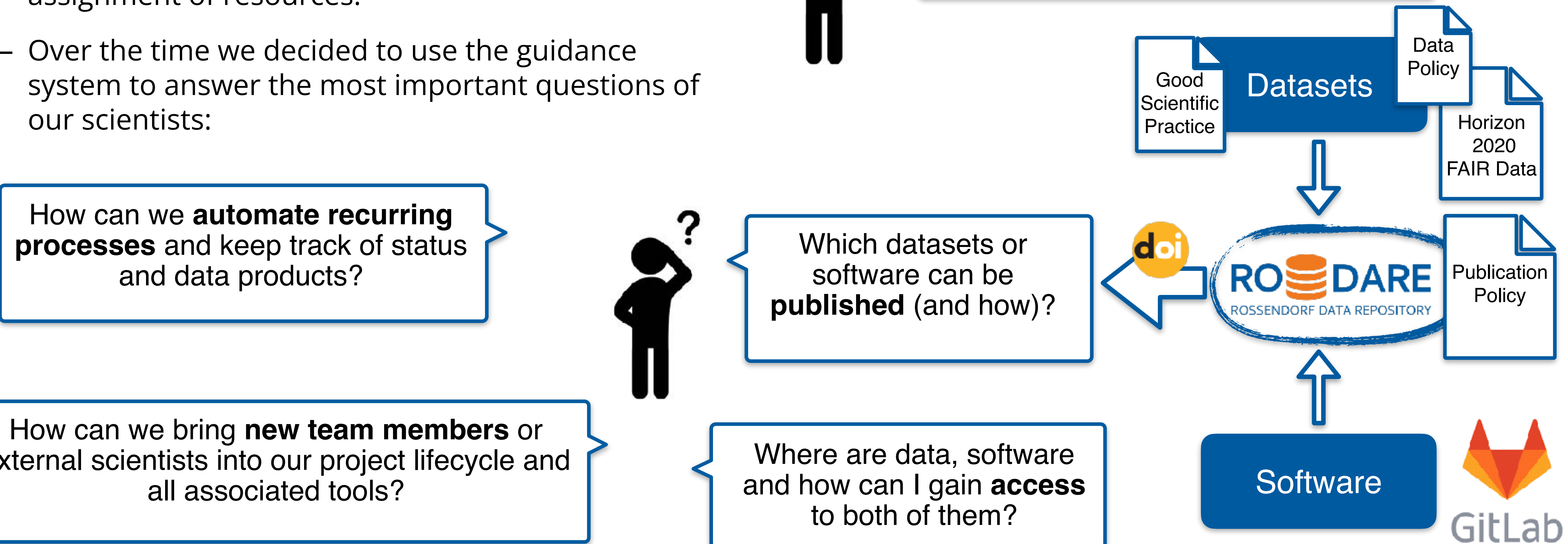


Which datasets or software can be **published** (and how)?

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

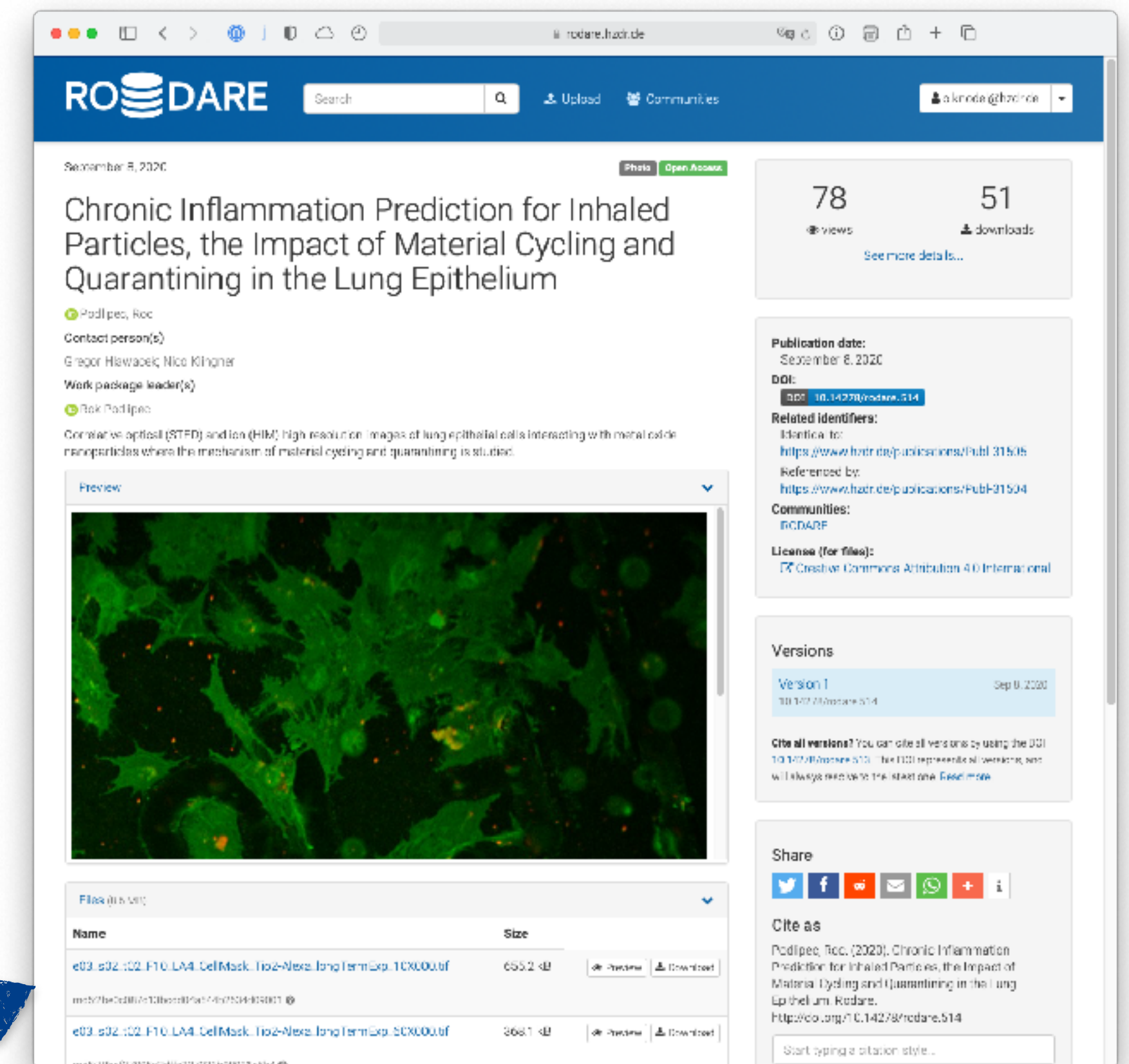
And how we can support them?!

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

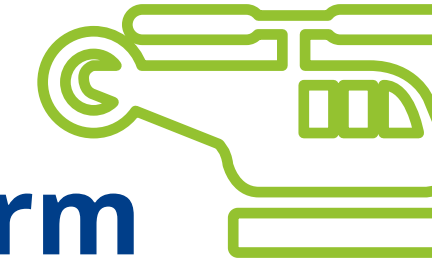


## How we went on with our ~~“Data Management Guidance System”~~

- 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:



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 09:14:34.295524+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://ddd",
      "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": ""
    }
  ]
}

```

# 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 HELIPOINT",
  "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** HELMholtz Scientific Project Workflow Platform

Search English

### Project Properties

HZDR-ID	HZDR.FWCC.2021.95018
Handle	<a href="https://hdl.handle.net/20.500.12865/HZDR.FWCC.2021.95018">20.500.12865/HZDR.FWCC.2021.95018</a>
Digital Object ID	73
uuid	8fab8a14-0f2f-414d-bbe0-747c38129bc4
serialization url	<a href="https://vlsdms.fz-rossendorf.de/project/9/serialize/">https://vlsdms.fz-rossendorf.de/project/9/serialize/</a>
Owner	Voigt, Martin (FWCC) - 141575
Created	May 18, 2021, 1:03 p.m.
Group	FWCC
Title	An Example Project
Description	This Project has the sole purpose of demonstrating the serialization functionality of HELIPOINT using vocabularies like datacite, rdfs or custom terms.

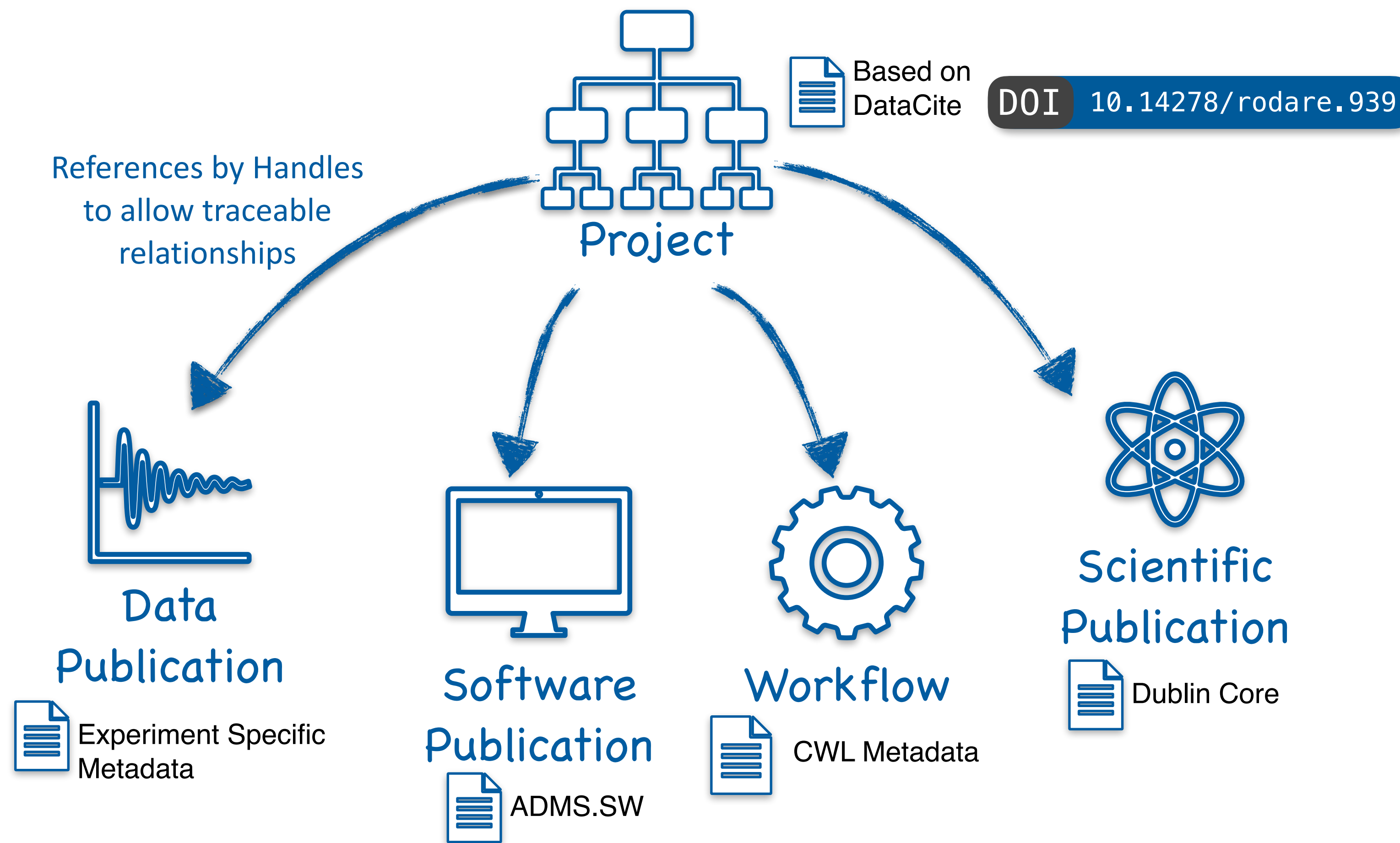
[Edit](#)

### Members

Name
Knodel, Dr. Oliver (FWCC) - 132739
Ufer, Robert (FWCC) - 141576

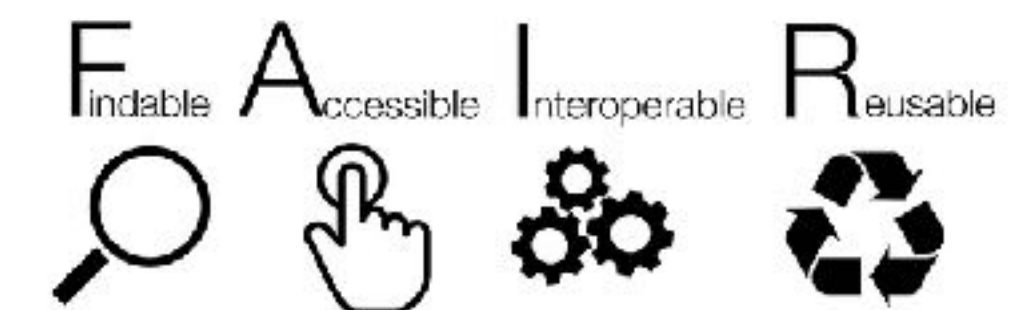


# Heliport Metadata Ecosystem



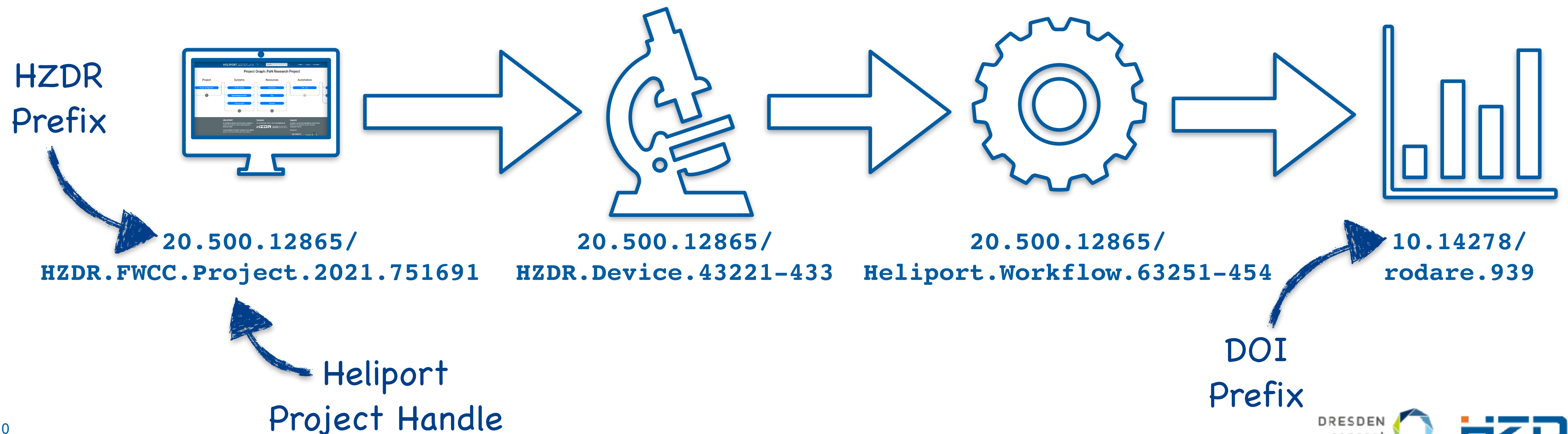
## Our Objective

- 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

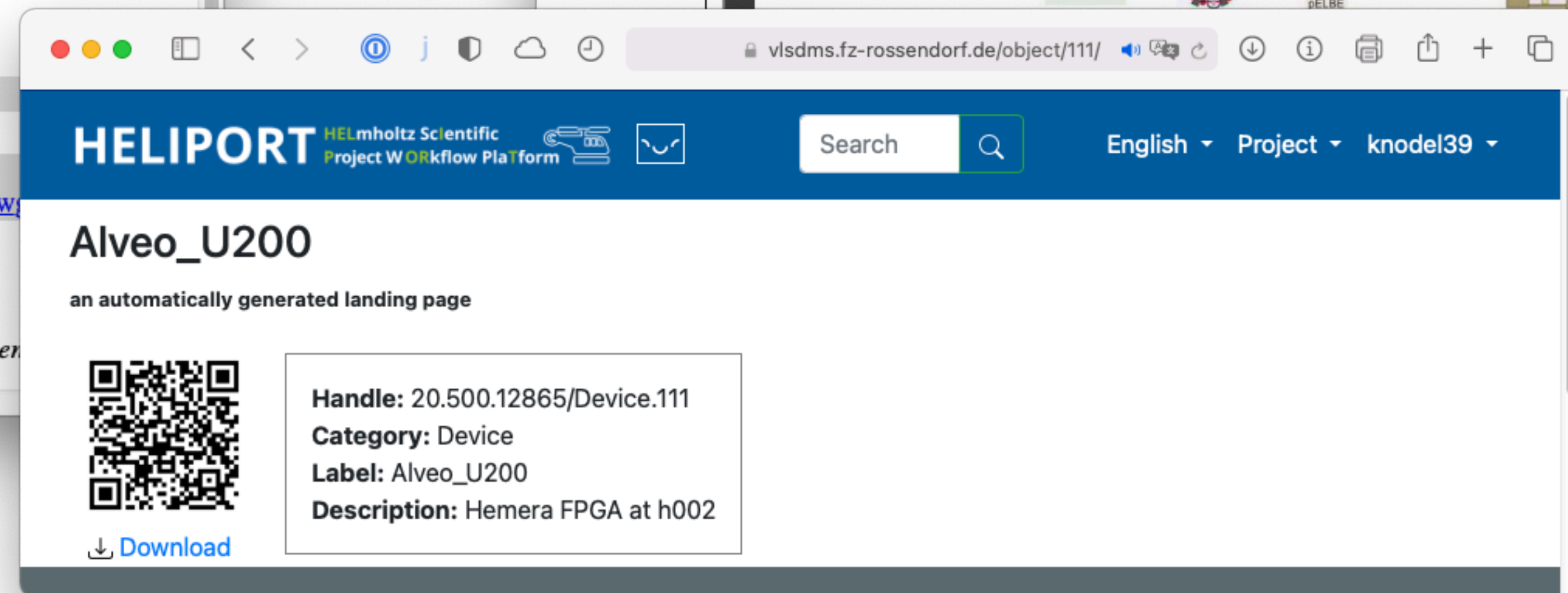
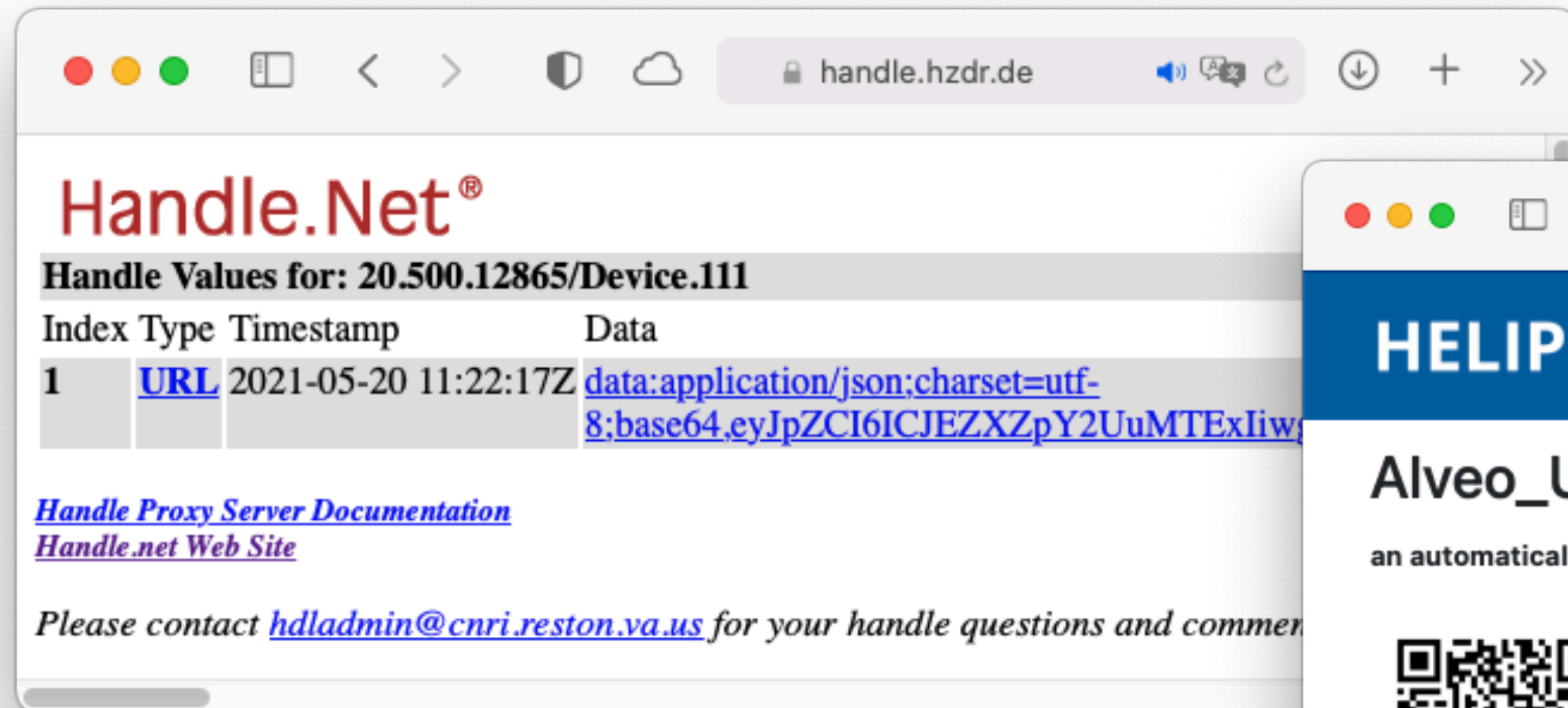
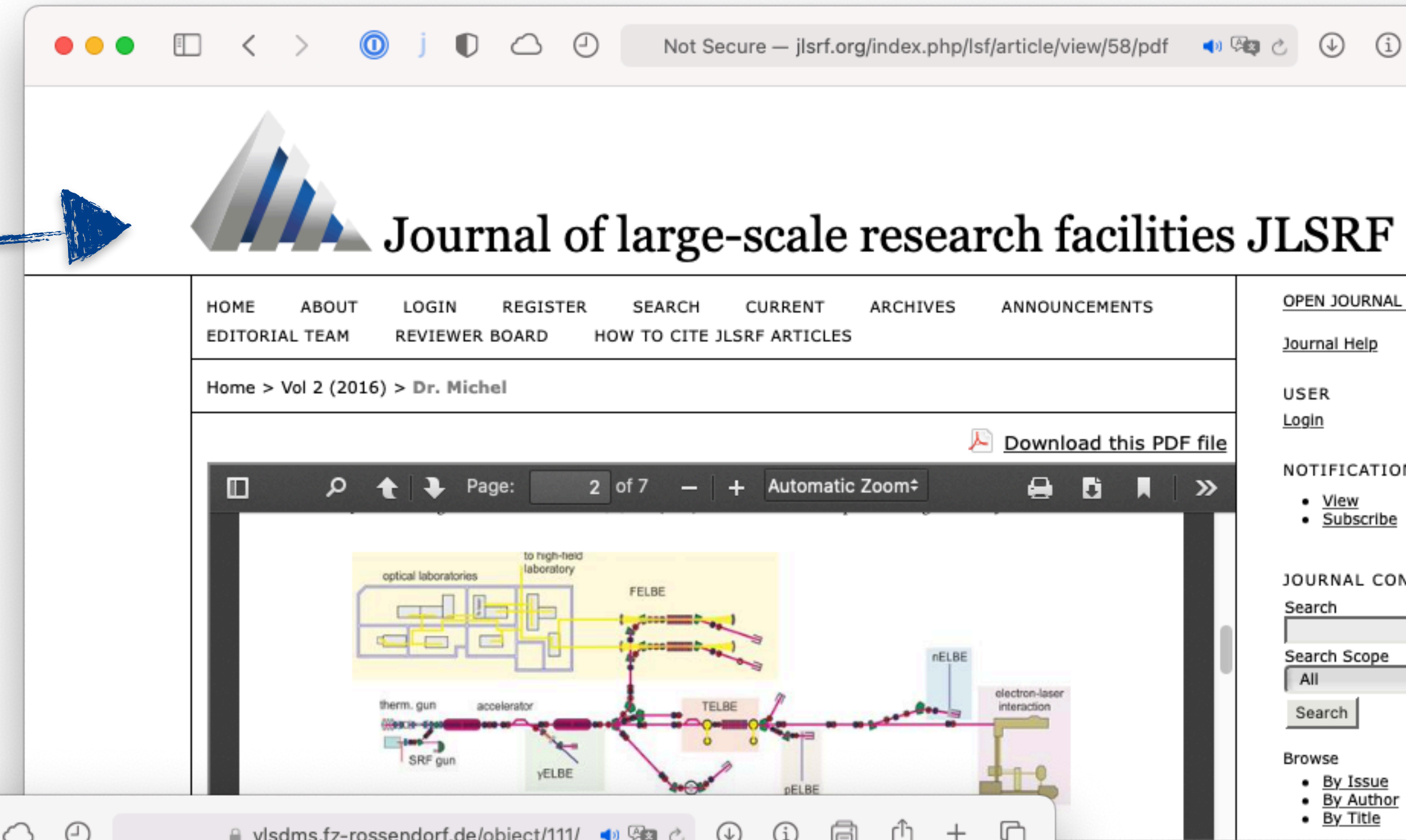
Heliport is linked with our local Handle-Server ([handle.hzdr.de](http://handle.hzdr.de)) **hdlenabled** and generates uniform PIDs (resolvable using [hdl.handle.net](http://hdl.handle.net)) from and for various systems and services. Associated information can be changed as needed without changing the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



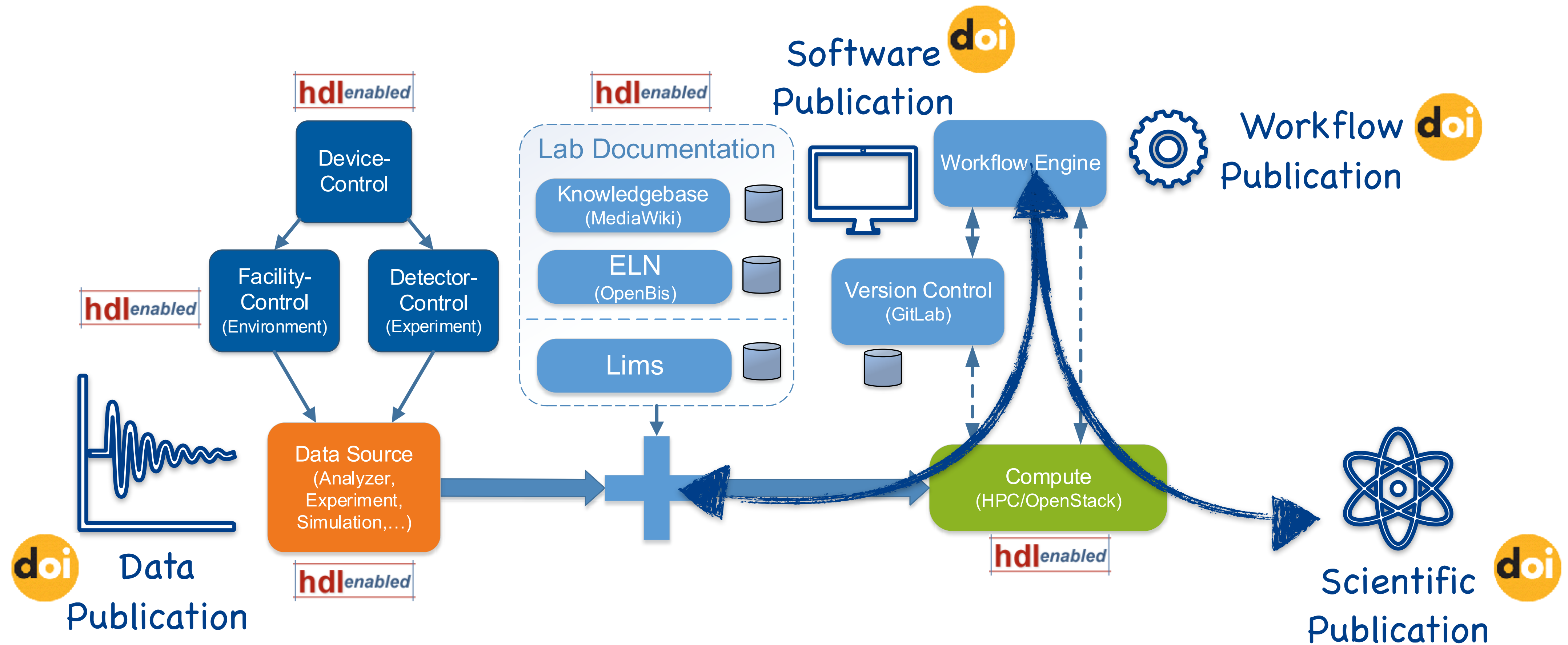
# The Role of Handles

- 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

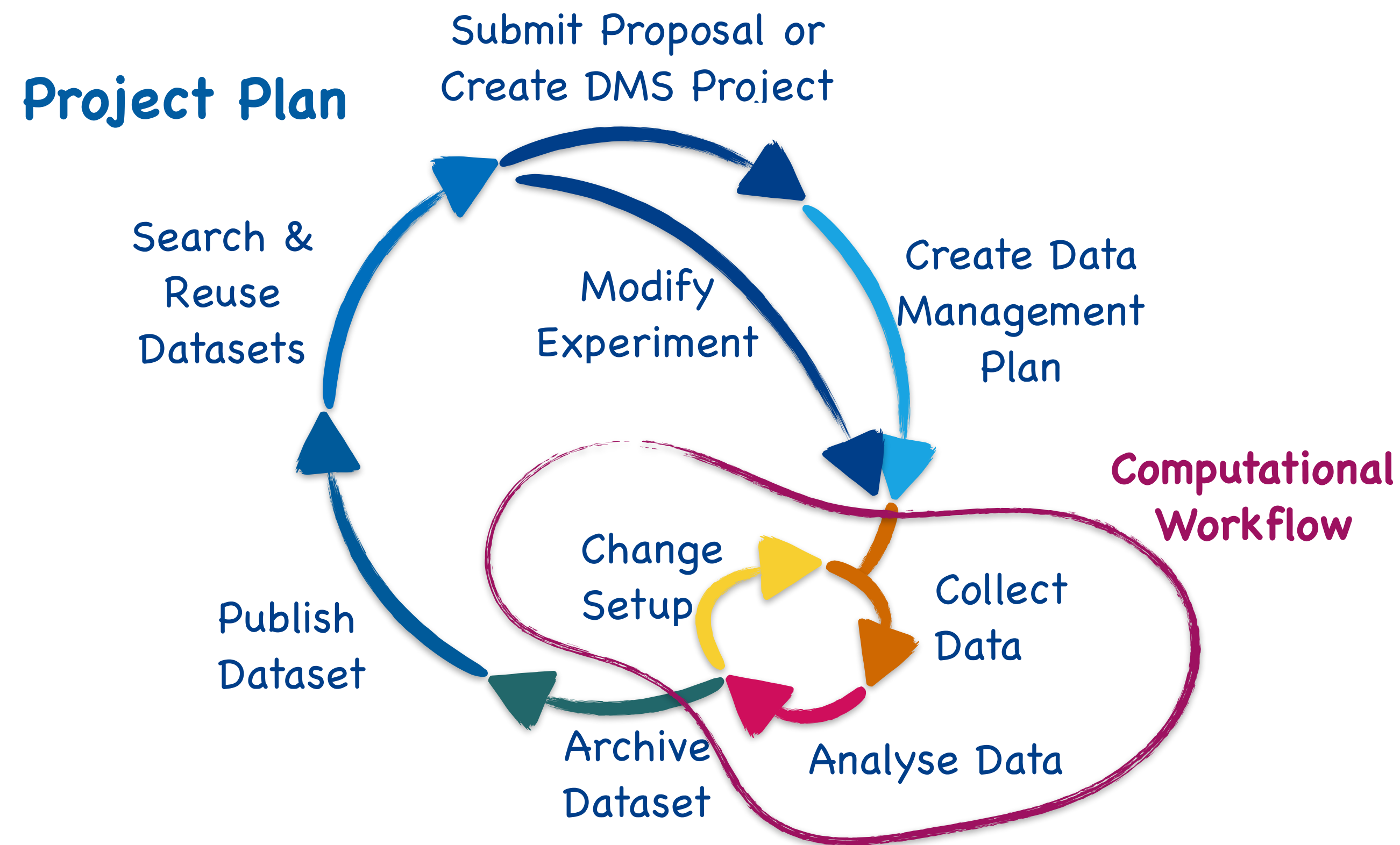


# Different Types of Publications based on our Infrastructure



# 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

ID	Name	Cluster Login	Directory on Cluster	Status
46	cat echo	Perseus	~/heliport_jobs	✓
44	echo cat sleep	Choose a Login	~/heliport_jobs	✓
44	echo cat sleep	Perseus	~/heliport_jobs	✓
51	one bad disc per work	Choose a Login	~/heliport_jobs	✗
51	one bad disc per work	Perseus	~/heliport_jobs	✗
41	sleep 5 seconds	Choose a Login	~/heliport_jobs	⚠
41	sleep 5 seconds	Perseus	~/heliport_jobs	⚠

Workflow Engine

Version Control

Compute (HPC, OpenStack)

UNICORE

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

- Documented and
- Reproducible



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

HELIPORT Edit a Scientific Workflow

Name: curl and cat echo:it and stderr

Description:

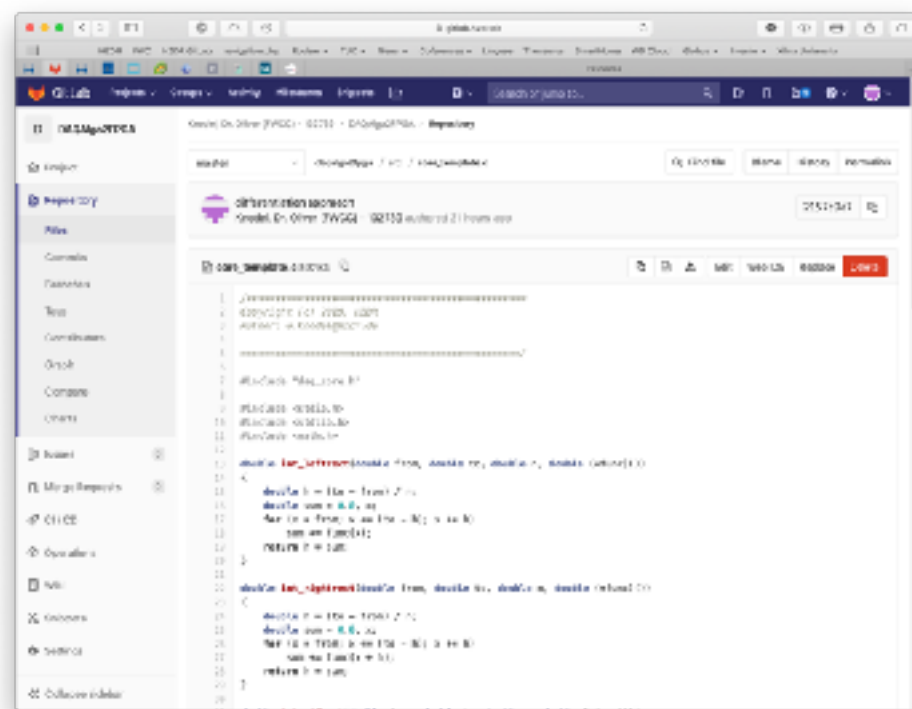
```

graph LR
    link((link)) --> curl((curl))
    curl --> cat((cat))
    curl --> echo((echo))
  
```

Buttons: Save, Cancel, Fit to Screen, Delete Selection

ID	Name	Description
35	echo	

Buttons: Add



# Heliport Tracks Workflows, their Execution and Metadata

**Jobs Table:**

ID	Name	Cluster Login	Directory on Cluster	Status
46	cat chain	hemera	~/heliport_jobs	Success
44	echo cat sleep	Choose a Login	~/heliport_jobs	Pending
44	echo cat sleep	hemera	~/heliport_jobs	Pending
51	one bad deed per week	Choose a Login	~/heliport_jobs	Pending
51	one bad deed per week	hemera	~/heliport_jobs	Pending
41	sleep 5 seconds	Choose a Login	~/heliport_jobs	Pending
41	sleep 5 seconds	hemera	~/heliport_jobs	Pending

**Test job Run Output:**

```
95% | 1.1MB 9.0MB/s eta 0:00:01
93% | 1.1MB 11.2MB/s eta 0:00:01
94% | 1.1MB 11.5MB/s eta 0:00:01
95% | 1.1MB 10.1MB/s eta 0:00:01
96% | 1.1MB 11.5MB/s eta 0:00:01
97% | 1.1MB 12.5MB/s eta 0:00:01
98% | 1.2MB 10.7MB/s eta 0:00:01
99% | 1.2MB 11.9MB/s eta 0:00:01
99% | 1.2MB 11.9MB/s eta 0:00:01
100% | 1.2MB 650kB/s
[?25hCollecting mpy-extensions (from cwltool)
Downloading
https://files.pythonhosted.org/packages/5c/eb/975c7c080f3223a5cdaff09612f3a5221e4ba534f7039db34c35d95fa6a5
/mpy_extensions-0.4.3-py2.py3-none-any.whl
Collecting requests>=2.6.1 (from cwltool)
Cache entry deserialization failed, entry ignored
Downloading
https://files.pythonhosted.org/packages/29/c1/24014557f1d22c56d50280771a17307e6bf87b70727d975fd6b2ce6b014a
/requests-2.25.1-py2.py3-none-any.whl (61kB)
[?25l
16% | 10kB 27.2MB/s eta 0:00:01
33% | 20kB 25.3MB/s eta 0:00:01
50% | 30kB 25.5MB/s eta 0:00:01
66% | 40kB 26.3MB/s eta 0:00:01
83% | 51kB 23.9MB/s eta 0:00:01
100% | 61kB 994kB/s
[?25hCollecting shellescape<3.5, >=3.4.1 (from cwltool)
Downloading
https://files.pythonhosted.org/packages/51/b6/986c99a10040beaaefca1ad6c93bd7738cb8e4f52f6caed13d3ed1ca57e4
/shellescape-3.4.1-py2.py3-none-any.whl
Collecting psutil (from cwltool)
Cache entry deserialization failed, entry ignored
Downloading
https://files.pythonhosted.org/packages/e1/b0/7276de53321c12981717490516b7e612364f2cb372ee8901bd4a66a000d7
/psutil-5.8.0.tar.gz (470kB)
[?25l
2% | 10kB 11.5MB/s eta 0:00:01
4% | 20kB 14.2MB/s eta 0:00:01
6% | 30kB 16.6MB/s eta 0:00:01
8% | 40kB 18.4MB/s eta 0:00:01
10% | 51kB 19.5MB/s eta 0:00:01
13% | 61kB 20.7MB/s eta 0:00:01
```

**Edit a Scientific Workflow:**

Name: Create supercell from element and structure

Description: Given the full name of an element, the structure description from the MaterialsProject and a number of atoms, create a supercell in VASP POSCAR format.

Workflow Diagram:

```
graph LR
    full_name((full_name)) --> new_step((new_step))
    mp_name((mp_name)) --> new_step
    natoms((natoms)) --> new_step
    new_step --> new_step_1((new_step_1))
    new_step_1 --> supercell((supercell))
```

# 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 displays the Heliport REST API documentation interface. The browser address bar shows the URL: `vlsdms.fz-rossendorf.de/redoc/#operation/createDigitalObject`. The interface is divided into several sections:

- Left Sidebar:** A search bar and a list of API endpoints. The 'digital-objects' section is expanded, showing endpoints like `listDigitalObjects`, `createDigitalObject` (highlighted), `retrieveDigitalObject`, `updateDigitalObject`, `partialUpdateDigitalObject`, and `destroyDigitalObject`.
- Main Content Area:** Details for the `createDigitalObject` endpoint. It shows the request body schema for `application/json` with the following fields:
  - `project`: integer, required
  - `handle`: string, <= 100 characters, Nullable
  - `relation`: string, required
  - `category`: string, required
  - `description`: string, required
- Responses:** A section showing a response with status code `201`.
- Request samples:** A section showing a sample request payload:

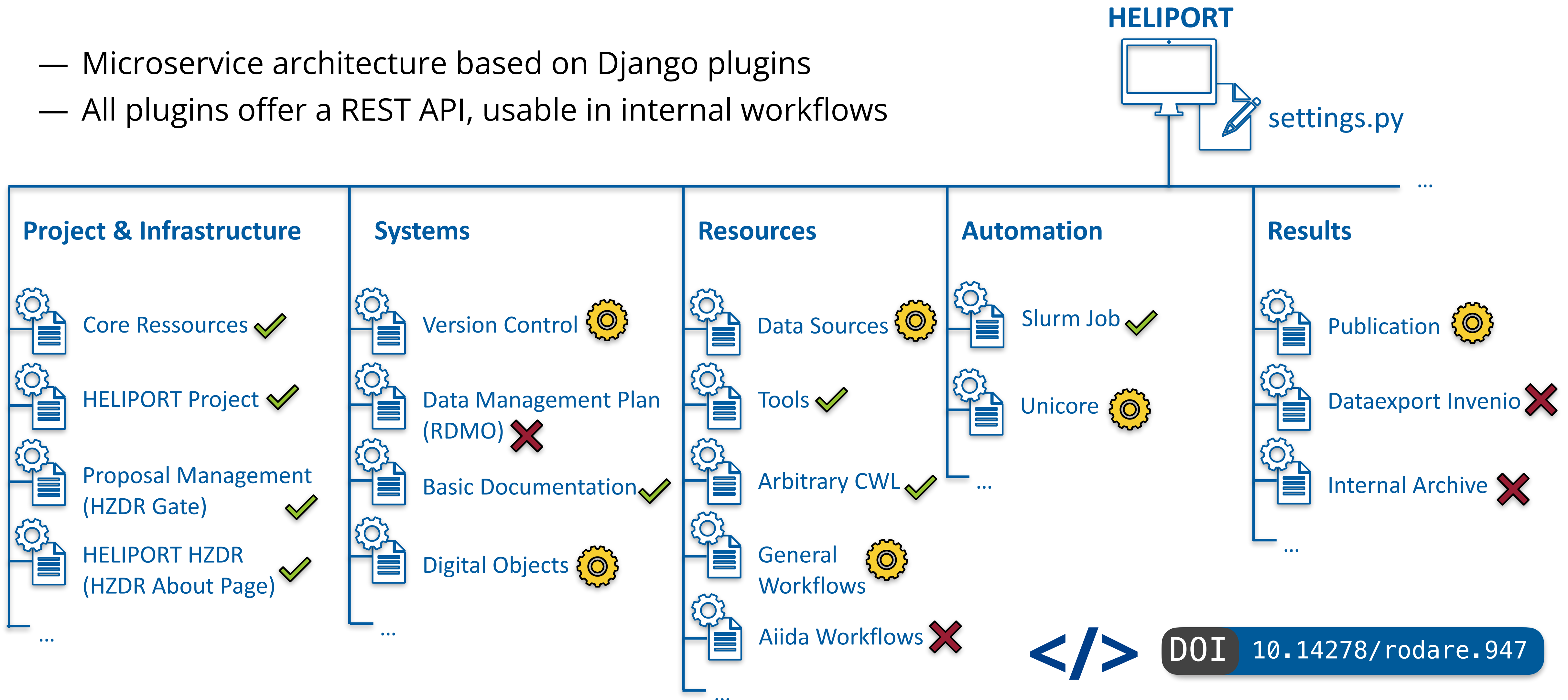
```
{  "project": 0,  "handle": "string",  "relation": "string",  "category": "string",  "description": "string"}
```
- Response samples:** A section showing a sample response with status code `201` and content type `application/json`. The response body is partially visible:

```
{  "digital object id": 0,  "project": 0,  "handle": "string",  "relation": "string",  "category": "string",  "description": "string"}
```



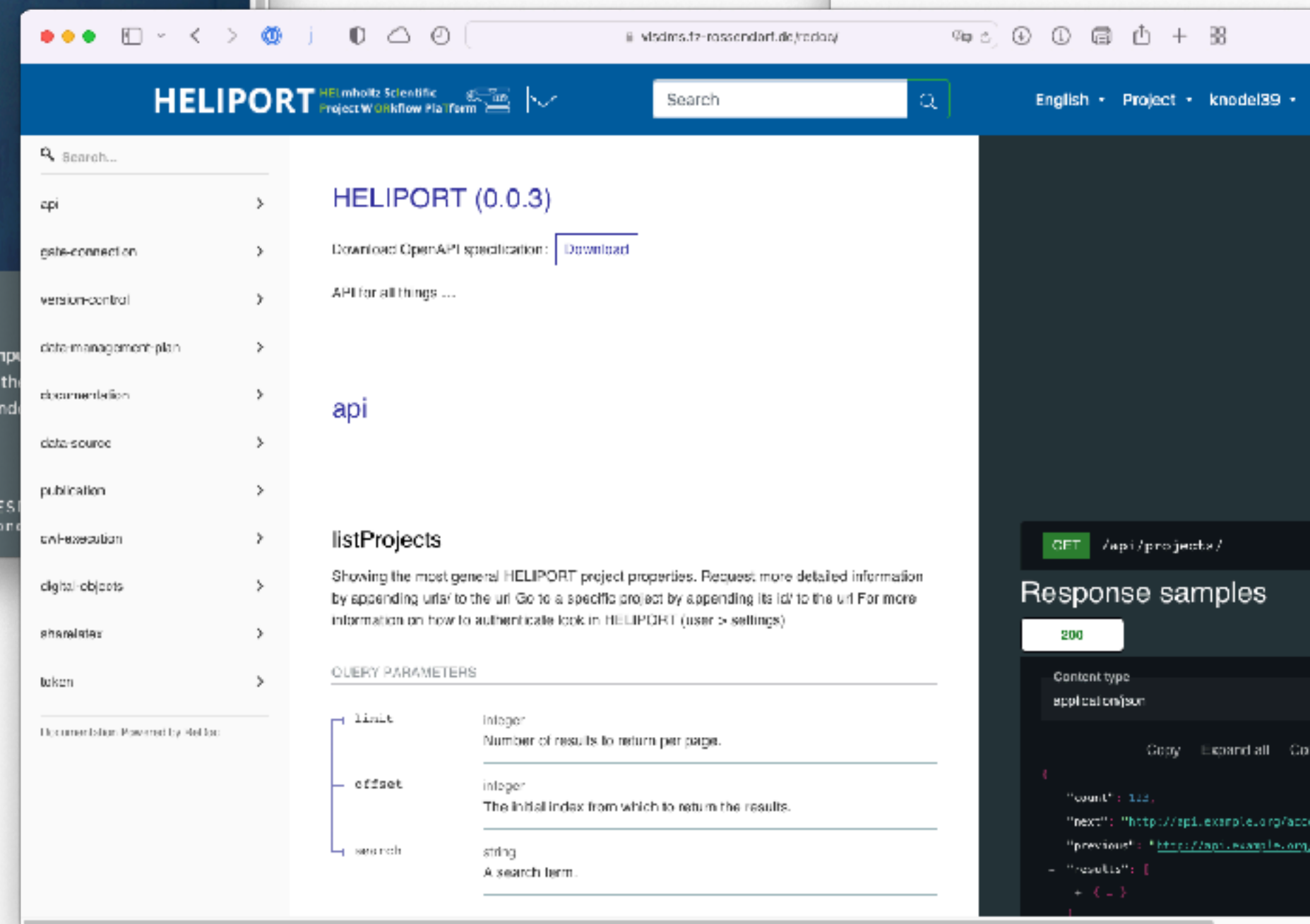
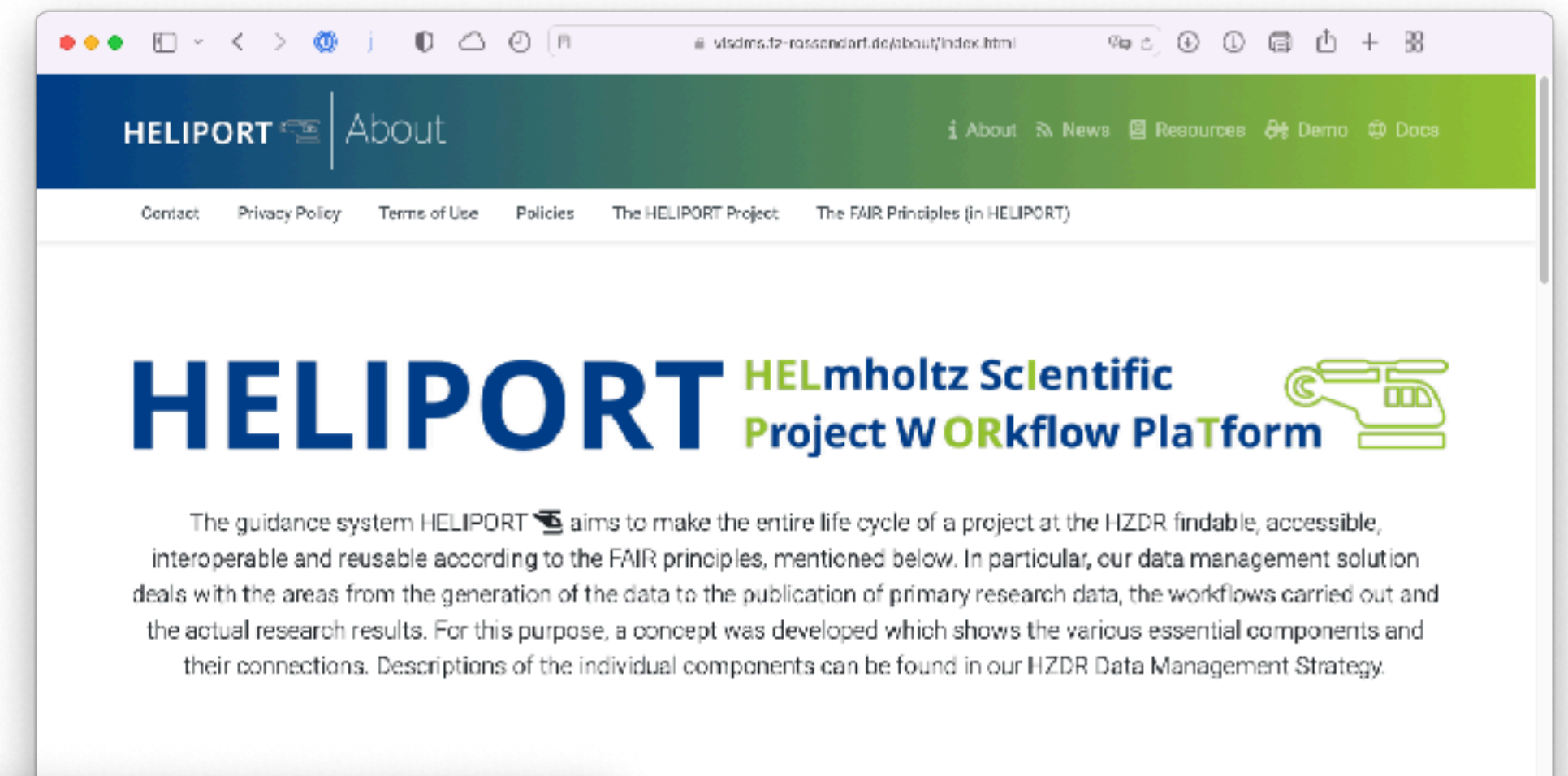
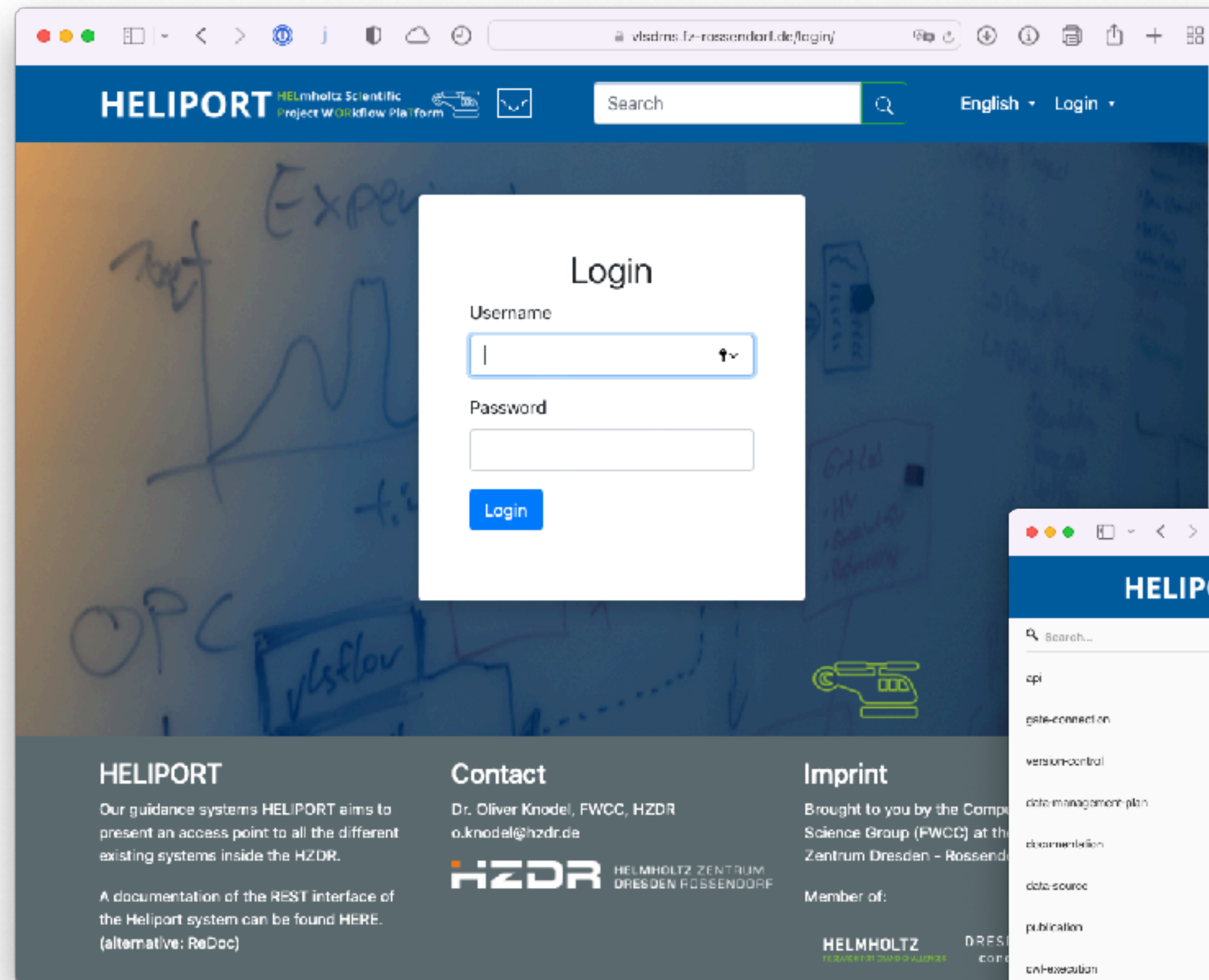
# Modular HELIPOINT Design (Django Apps)

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



# 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": "9ff3ddjuikl3l2mnn56pghsw6mwqh"}
```

## Read gate projects

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

# REST API Token

The screenshot shows a web browser window displaying the HELIPORT Project Graph interface. The URL in the address bar is `visdms.fz-rossendorf.de/project/61/graph/`. The page features a blue header with the HELIPORT logo, a search bar, and navigation options for language (English) and project (knodel39). The main content area is titled "Project Graph:" and displays a horizontal flow diagram with five nodes: Project, Systems, Resources, Automation, and Results. A blue arrow points from the "Automation" node to a dropdown menu that is open, showing options for Home, Settings, and Logout. Below the flow diagram, the text "Project → Settings" is displayed in a large blue font. The footer contains information about HELIPORT, contact details for Dr. Oliver Knodel, and logos for HELMHOLTZ and DRESDEN concept.

HELIPORT HELMHOLTZ Scientific Project Workflow Platform

Search English Project knodel39

Home Settings Logout

Project Systems Resources Automation Results

Project → Settings

HELIPORT  
Our guidance systems HELIPORT aims to present an access point to all the different existing systems inside the HZDR.  
A documentation of the REST interface of the Heliport system can be found HERE. (alternative: ReDoc)

Contact  
Dr. Oliver Knodel, FWCC, HZDR o.knodel@hzdr.de

IMPRINT  
Brought to you by the Computational Science Group (FWCC) at the Helmholtz-Zentrum Dresden - Rossendorf (HZDR).  
Member of:  
HELMHOLTZ DRESDEN concept

Open "https://visdms.fz-rossendorf.de/user/settings/" in a new tab

# REST API Token

The screenshot shows a web browser window with the URL `vlsdms.fz-rossendorf.de/user/settings/`. The page header includes the HELIPORT logo, a search bar, and navigation links for 'English', 'Project', and 'knodel39'. The main content area is titled 'Settings' and contains a 'User Information' section with the following details:

User-ID	knodel39
Surname	Knodel
Givenname	Oliver
E-Mail	o.knodel@hzdr.de
Group	FWCC
Institute	FZR
ORCID	<a href="https://orcid.org/0000-0001-8174-7795">https://orcid.org/0000-0001-8174-7795</a> <a href="#">Edit</a>

Below the user information is the 'Token Authentication' section. It contains the following text:

You can see your authentication token by clicking on the toggle button below. The token is used to authenticate the user when querying the REST api of the HELIPORT System. Keep your token secret!

example of how to use token: `curl -H "Authorization: Token 0123456789" https://vlsdms.fz-rossendorf.de/api/`

The token is displayed in a green box: `Token: 565Hj5322Th22nlfy44874e33`

Below the token is a blue 'Toggle Token' button. A blue arrow points from the text 'Toggle Token' on the left to this button.

At the bottom of the settings page is the 'Remote Server' section, which starts with the text: 'If you have files and wish to use them in HELIPORT. Connect to a machine here:'

Toggle  
Token