

# Research and Metadata Management at HZDR

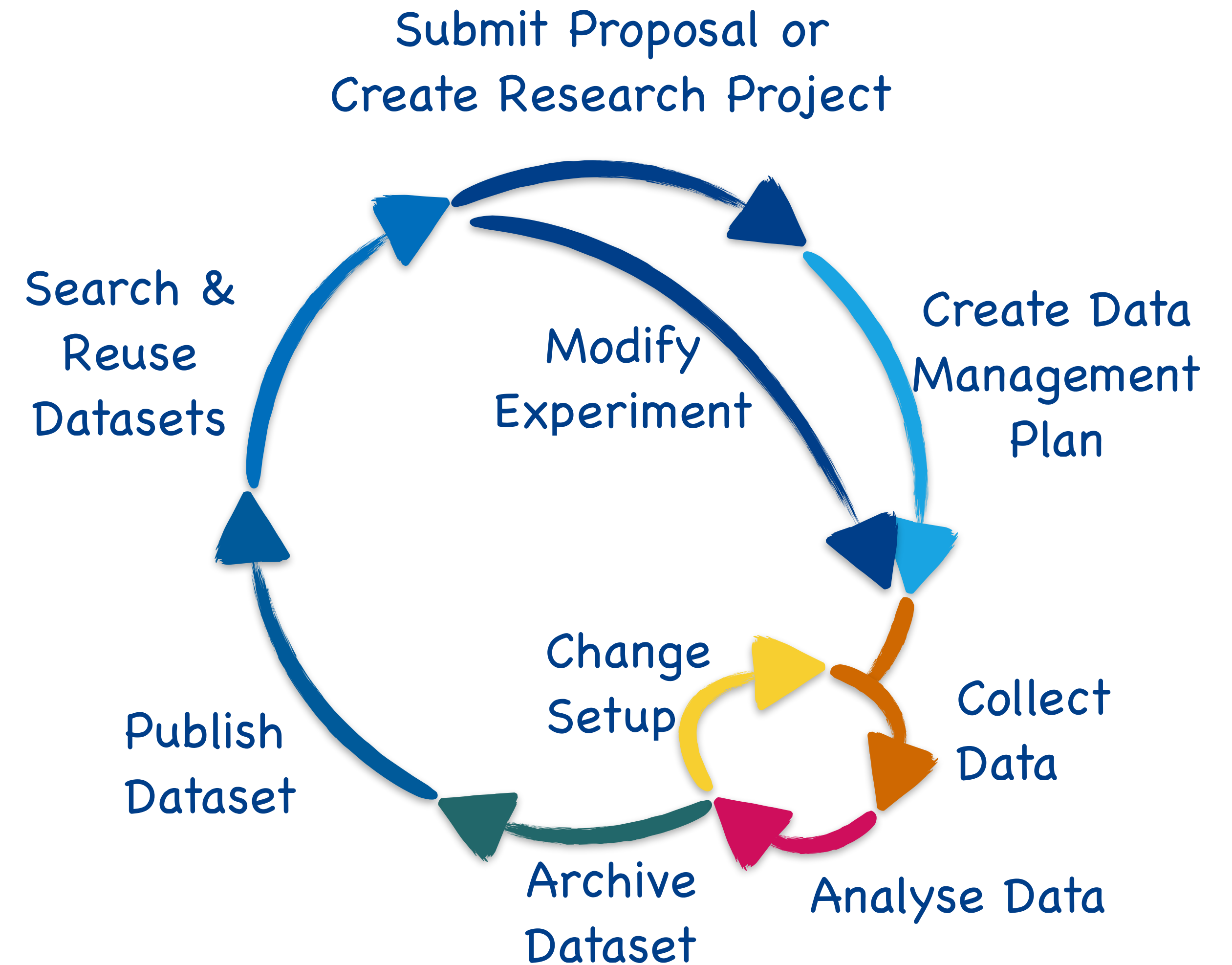


# Scientific Data and Project Lifecycle at HZDR

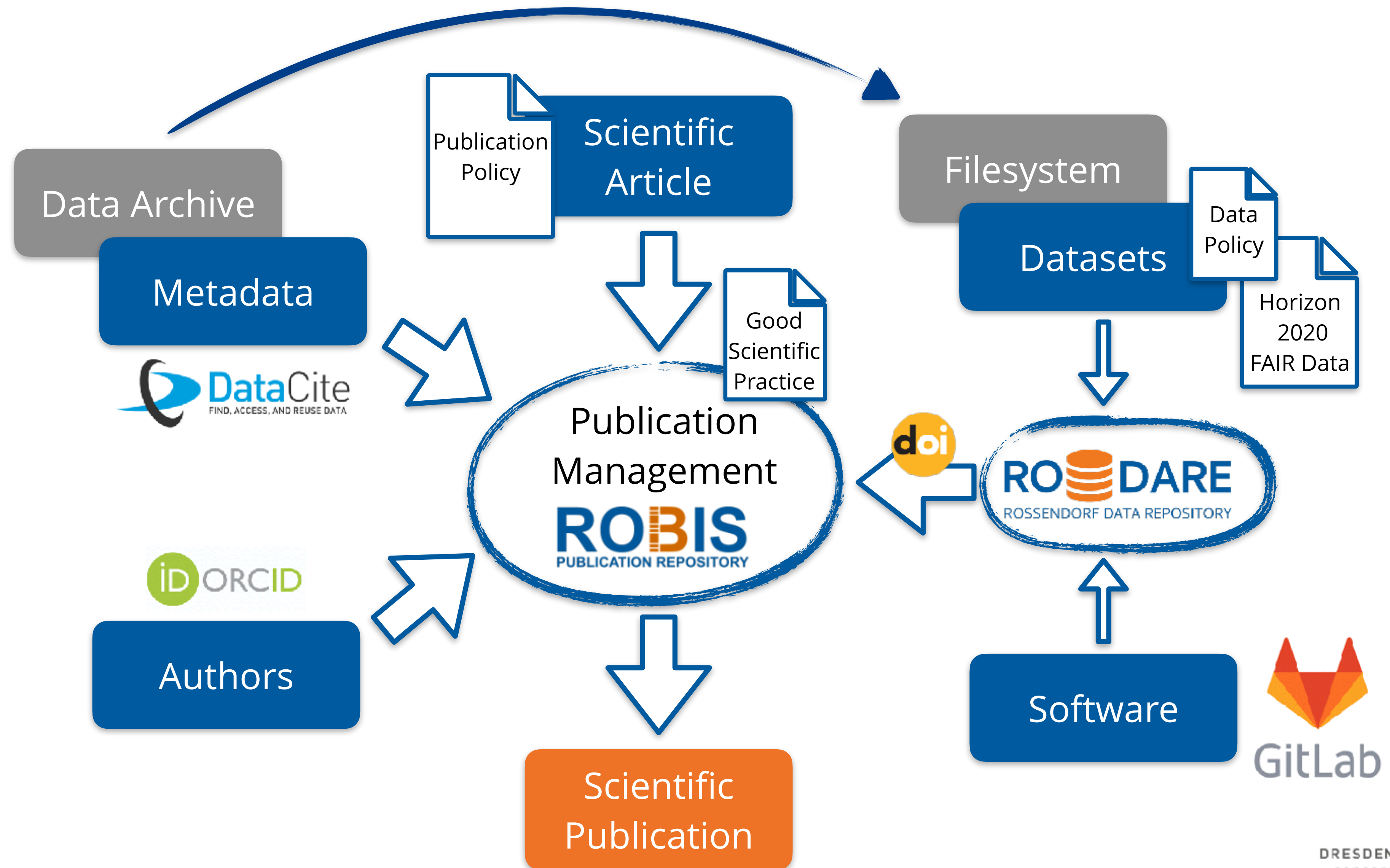
— In a typical project lifecycle we have multiple kinds of systems and **datasets** or **data products**:

- Proposal (Title, Authors, ...),
- Data Management Plan (Datasets, ...),
- Documentation (Experiment specific, ...),
- Workflows (Source Code, Files, ...),
- Source Code (Authors, License, ...),
- Instrument and protocol information,
- Sample or test object,
- Environment,
- Raw data from the experiment,
- Post-processed Data and
- Licenses.

— All of these datasets are referred to each other in different ways.



# HZDR Publication Ecosystem



# Our Contribution to Support FAIR Research: RODARE



February 2, 2021

## C++ & Python API for Scientific I/O with openPMD

3,961 views 211 downloads

Publication date: February 2, 2021  
DOI: 10.14278/rodare.798

Keywords: openPMD, Open Science, Open Data, HDF5, ADIOS, data, MPI, HPC, research, file-format, file-handling

Grants: European Commission, EUCALL - European Cluster of Advanced User Light Sources (654220)

Related identifiers: Cited by: 10.5281/zenodo.1167643, 10.5281/zenodo.1069534, 10.5281/zenodo.33674

Communities: OpenAIRE, RODARE

License (for files): GNU Lesser General Public License v3.0 only

Versions: Version 0.18.2 (Feb 2, 2021)

Supported by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of two U.S. Department of Energy organizations (Office of Science and the National Nuclear Security Administration). Supported by the Consortium for Advanced Modeling of Particles Accelerators (CAMP), funded by the U.S. DOE Office of Science under Contract No. DE-AC02-05OR21424. This work was partially funded by the Center of Advanced Systems Understanding (CASUS), which is financed by Germany's Federal Ministry of Education and Research (BMBWF) and by the Saxon Ministry for Science, Culture and Tourism (SMWK) with tax funds on the basis of the budget approved by the Saxon State Parliament.

Preview: openPMD-api-0.18.2.zip

- openPMD-openPMD-api-7e74673
  - appveyor.yml (4.3 kB)
  - clang-tidy (920 Bytes)
  - dockerignore (2.0 kB)
  - editorconfig (530 Bytes)
  - github
    - ISSUE\_TEMPLATE
      - bug\_report.md (1.2 kB)
      - feature\_request.md (764 Bytes)
      - install\_problem.md (724 Bytes)
      - question.md (810 Bytes)
    - ci
      - sanitizer

September 8, 2020

## Chronic Inflammation Prediction for Inhaled Particles, the Impact of Material Cycling and Quarantining in the Lung Epithelium

Podipee, Rok

Contact person(s): Gregor Hlawacek, Nico Klingner

Work package leader(s): Rok Podipee

Correlative optical (STED) and ion (FIM) high-resolution images of lung epithelial cells interacting with metal nanoparticles where the mechanism of material cycling and quarantining is studied.

Preview:

Files (8.6 MB)

Name	Size	Preview	Download
e03_s02_c02_F10_L1_A4_CellMask_Tin2_Alexa_LongTermExp_10X000.tif	658.2 kB		
m4523e0e08/d13beed04a544b2534d05001			
e03_s02_c02_F10_L1_A4_CellMask_Tin2-Alexa_LongTermExp_50X000.tif	568.1 kB		
m4573bef27205e3dfe227331b96921c9b4			

Referenced by: <https://www.hzdr.de/publications/Pub131504>

Communities: RODARE

License (for files): Creative Commons Attribution 4.0 International

Versions: Version 1 (10.14278/rodare.514) (Sep 8, 2020)

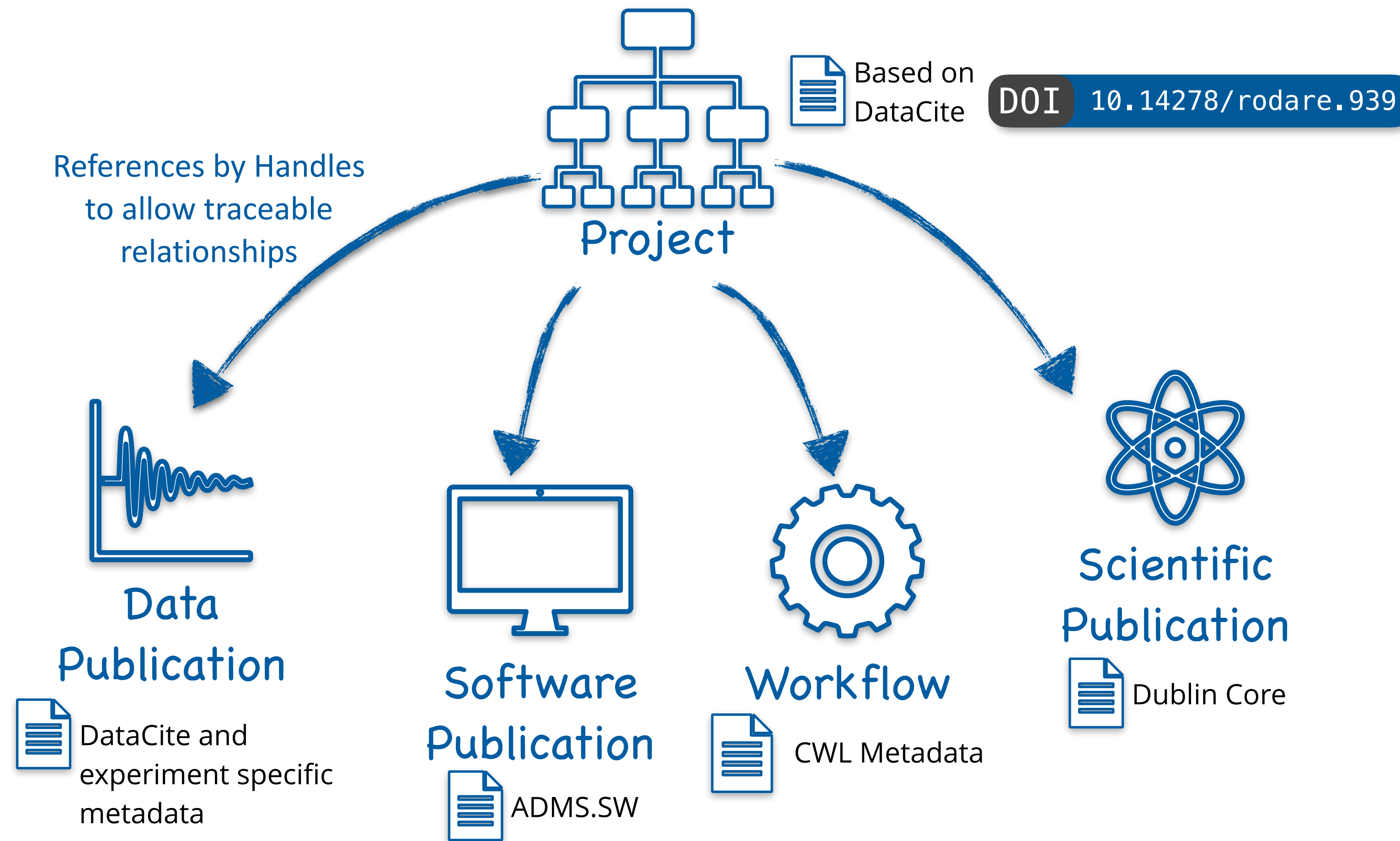
Cite all versions? You can cite all versions by using the DOI 10.14278/rodare.513. This DOI represents all versions, and will always resolve to the latest one. Read more.

Share:

Cite as: Podipee, Rok. (2020). Chronic Inflammation Prediction for Inhaled Particles, the Impact of Material Cycling and Quarantining in the Lung Epithelium. Rodare. <https://doi.org/10.14278/rodare.514>

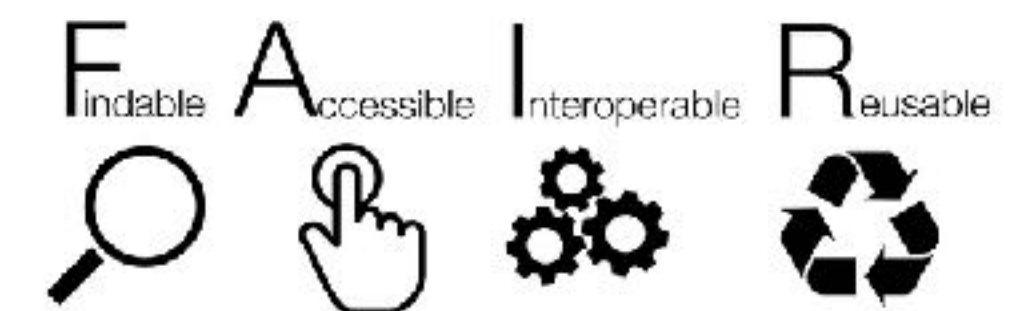
<https://rodare.hzdr.de>

# Metadata Ecosystem



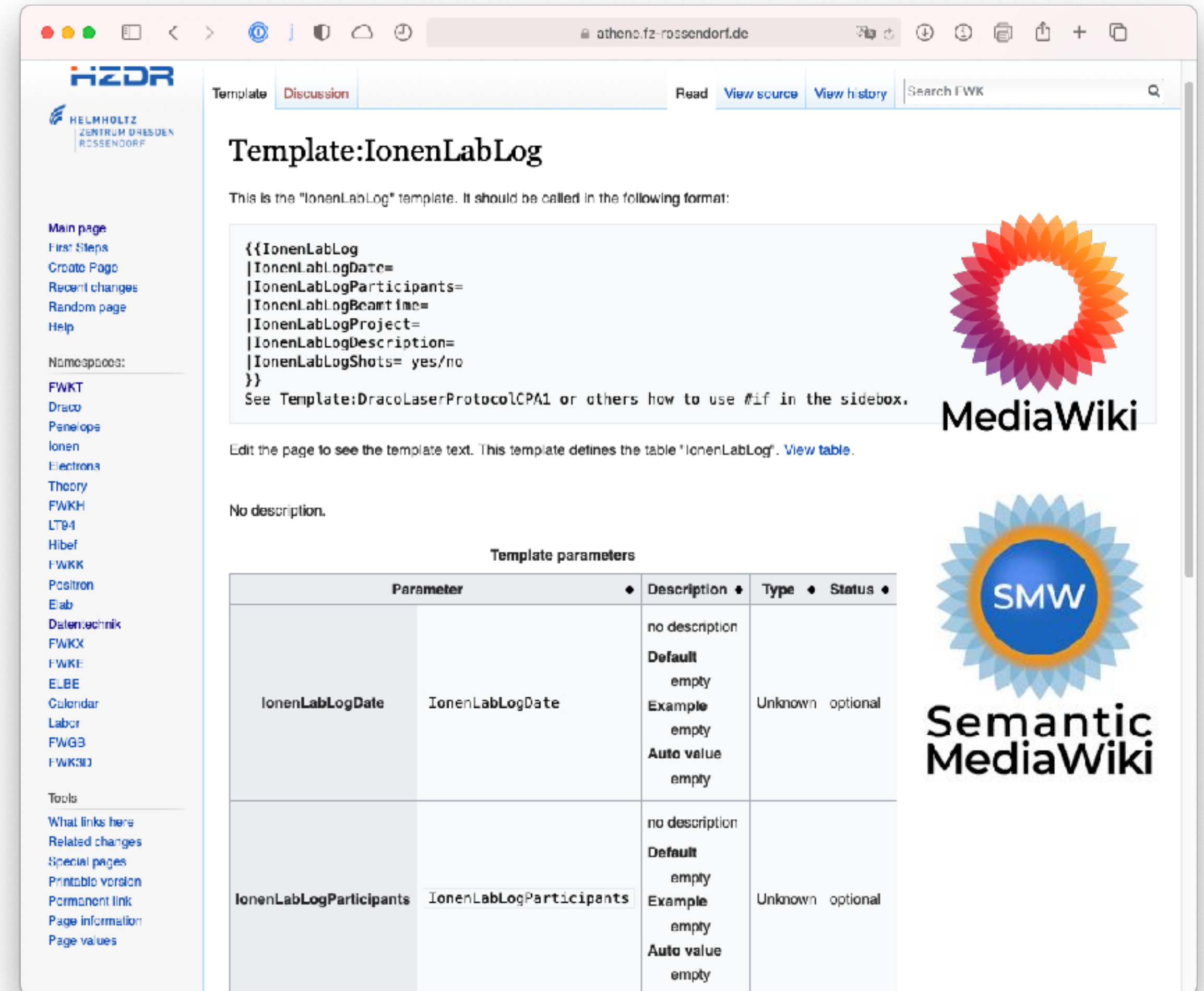
## Our Objective

- In all stages of an experiment we combine 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



# Experiment Specific Metadata — Still a Challenge...

- We need a flexible system to support our different research areas and data sources with an overarching metadata management.
- Electronic Lab Books for Better Meta-Data acquisition, structuring and management:
  - **Mediawiki:** Flexible and adaptable for our various experiments and easy to use.
  - **OpenBis:** Well-structured and flexible, but poor user acceptance at HZDR.
- The important message: The structured organization of all kinds of metadata is inevitable and the tool must be accepted by the users...



The screenshot shows a web browser window displaying the MediaWiki page for the template 'IonenLabLog'. The page is titled 'Template:IonenLabLog' and includes a navigation bar with 'Read', 'View source', and 'View history' options. The main content area shows the template code, a description, and a table of template parameters. The table has columns for 'Parameter', 'Description', 'Type', and 'Status'. The parameters listed are 'IonenLabLogDate' and 'IonenLabLogParticipants', both with a description of 'no description' and a status of 'Unknown optional'.

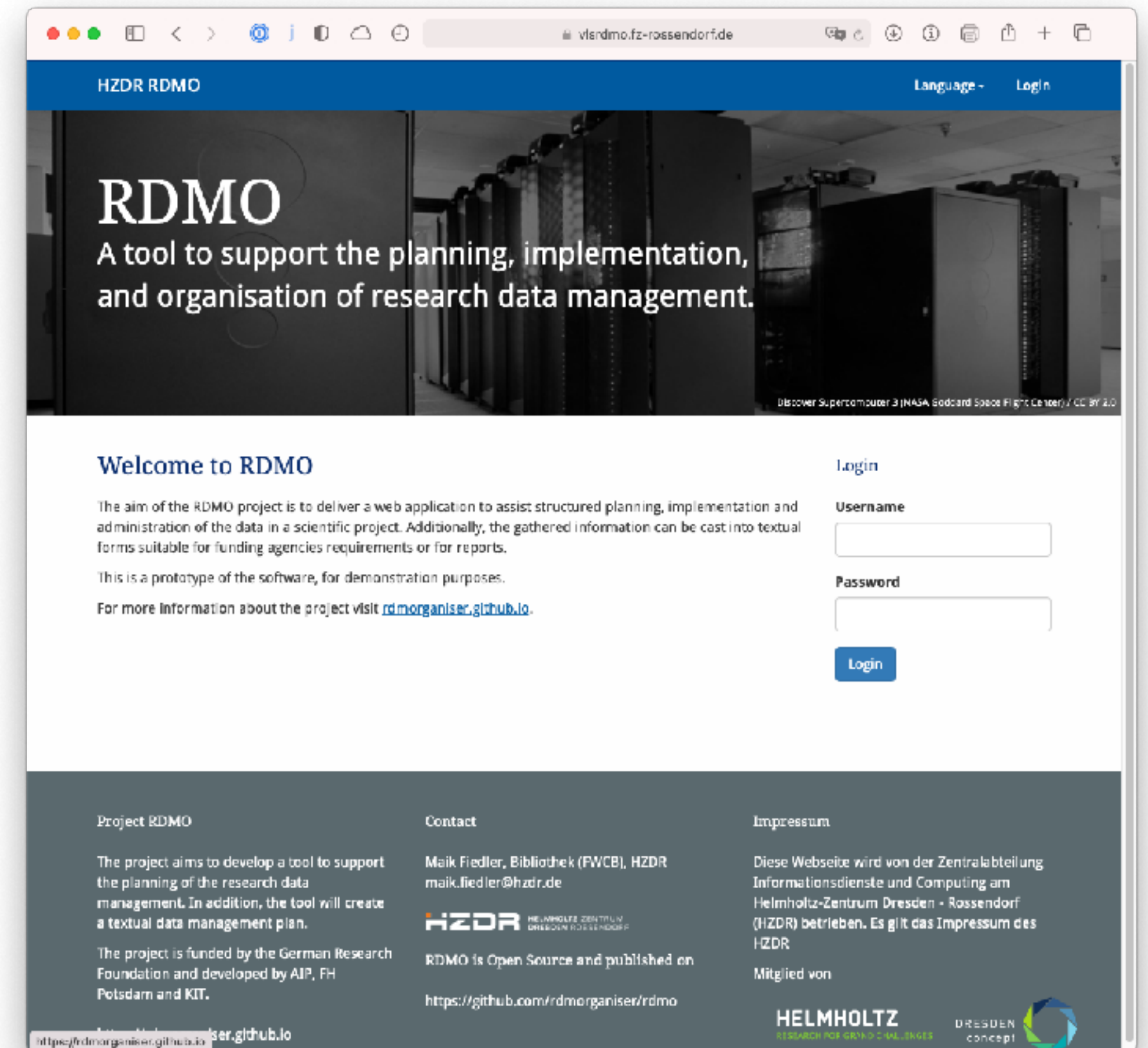
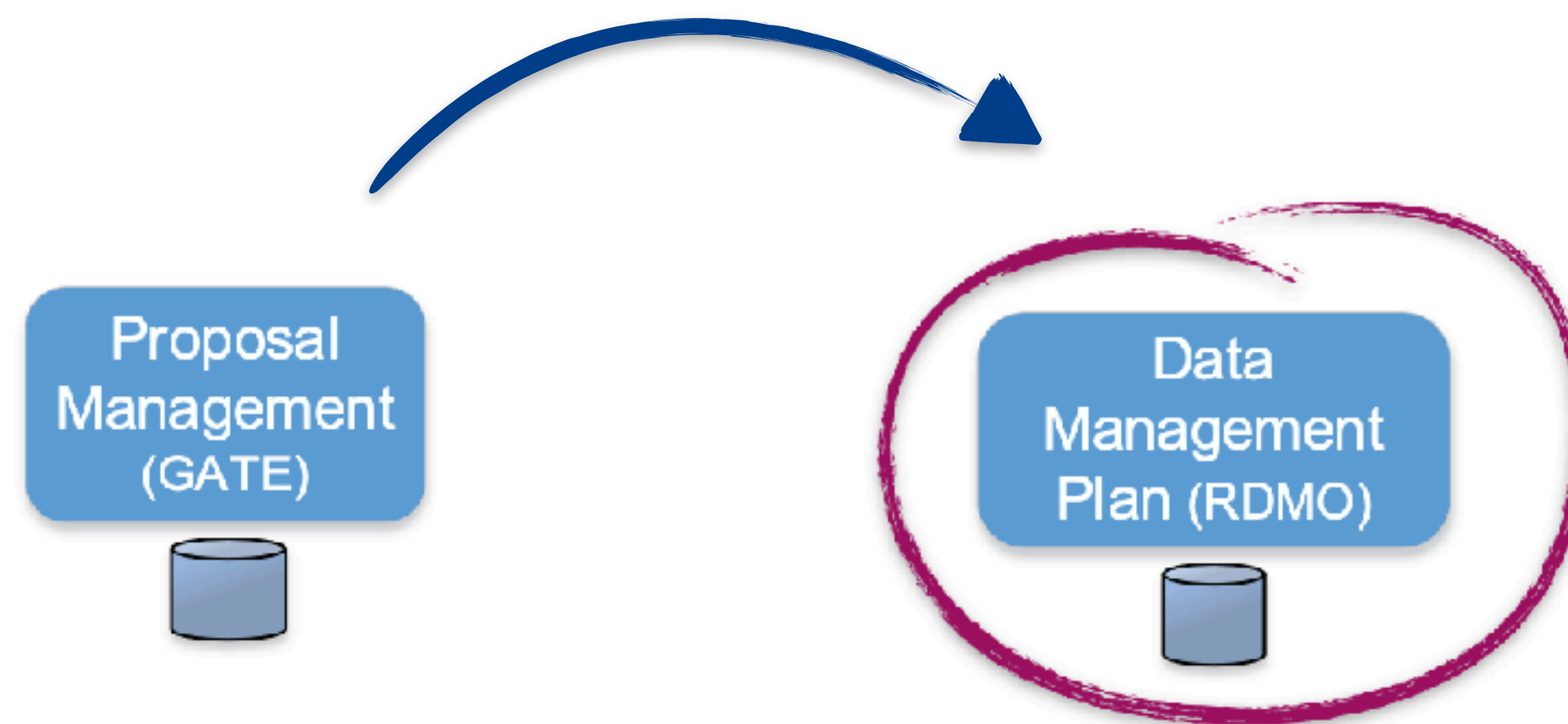
```
{{IonenLabLog
|IonenLabLogDate=
|IonenLabLogParticipants=
|IonenLabLogBeamtime=
|IonenLabLogProject=
|IonenLabLogDescription=
|IonenLabLogShots= yes/no
}}
```

See [Template:DracoLaserProtocolCPA1](#) or others how to use #if in the sidebox.

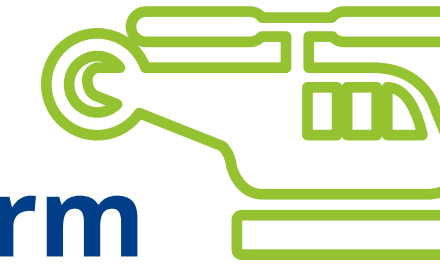
Parameter	Description	Type	Status
IonenLabLogDate	IonenLabLogDate no description Default: empty Example: empty Auto value: empty	Unknown	optional
IonenLabLogParticipants	IonenLabLogParticipants no description Default: empty Example: empty Auto value: empty	Unknown	optional

# Research Data Management Plans based on Metadata

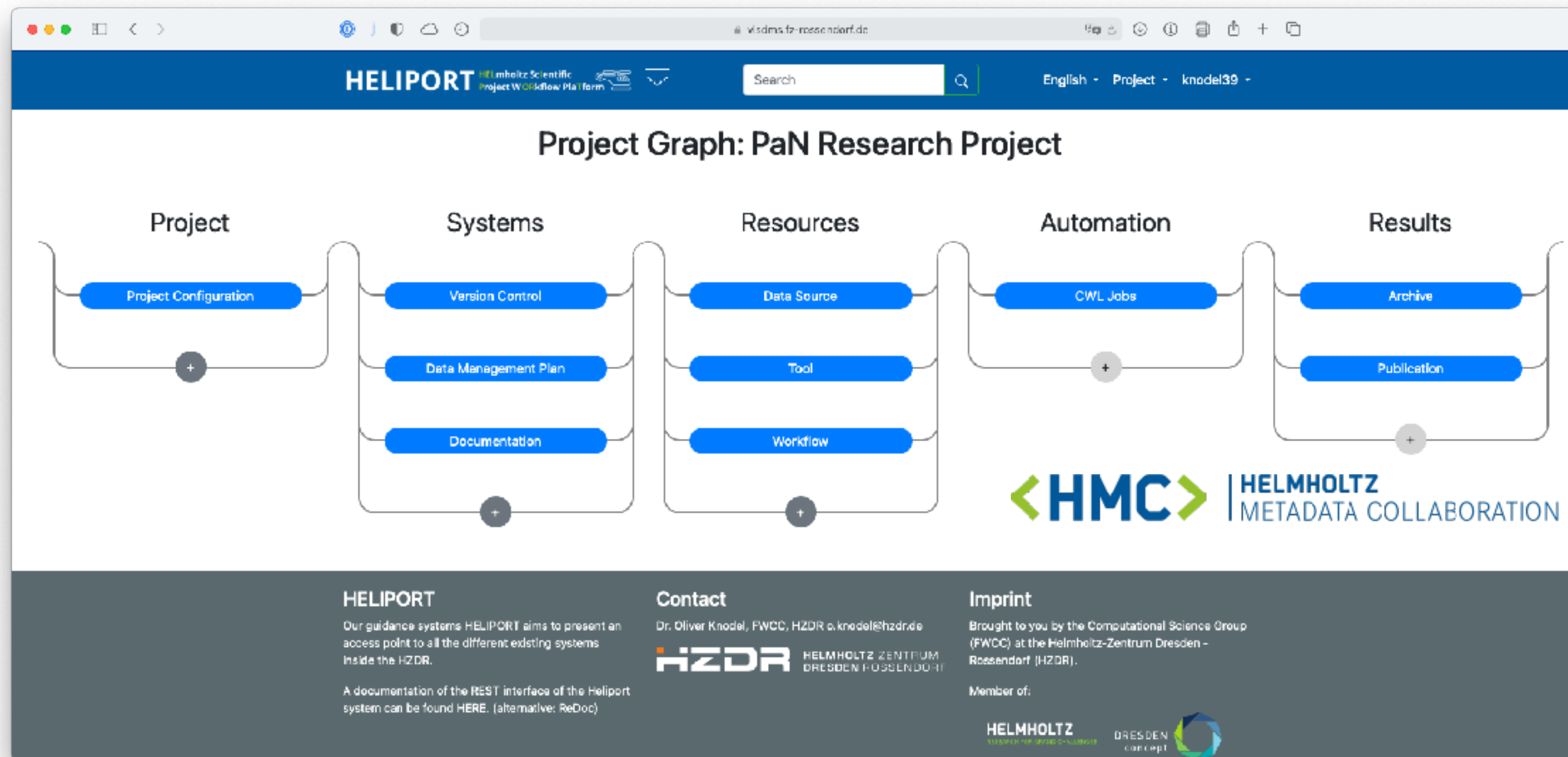
- Based on Research Facilities and subordinate detector we can estimate data products (size, format, metadata, ...).
- ... it is possible to generate a Data Management Plan (DMP) and can keep it up to date during the whole lifecycle of the experiment.



# 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, workflows and metadata schemas to create a more **FAIR** and comprehensible research project.

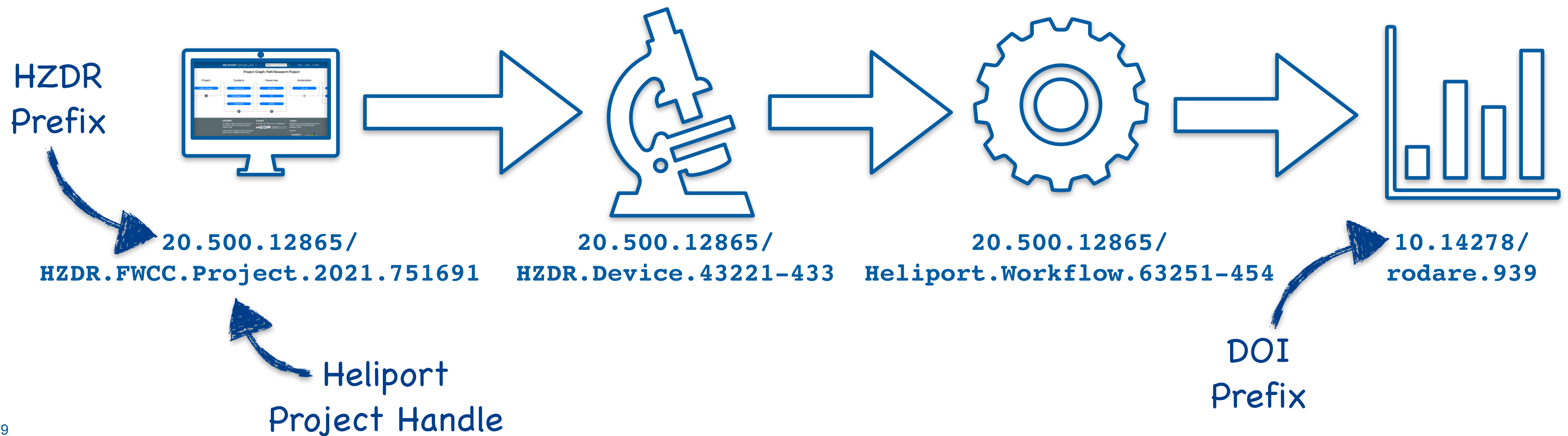


```
{
  "namespaces": {
    "datacite": "http://purl.org/spar/datacite/",
    "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
    "heliport": "https://heliport/schema/",
    "time": "http://www.w3.org/2006/time#",
    "dc": "http://purl.org/dc/terms/"
  },
  "heliport:project_id": 28,
  "datacite:hasIdentifier": "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://dddd",
      "datacite:hasDescription": "dddd"
    }
  ],
  "heliport:has_Documentation": [
    {
      "heliport:documentation_id": 7,
      "datacite:uri": "https://dddd",
      "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:has_Archive": [
    {
      "heliport:archive_id": 4,
      "datacite:hasDescription": "ret"
    }
  ],
  "heliport:has_Publication": [
    {
      "heliport:publication_id": 6,
      "datacite:hasDescription": ""
    }
  ]
}
```



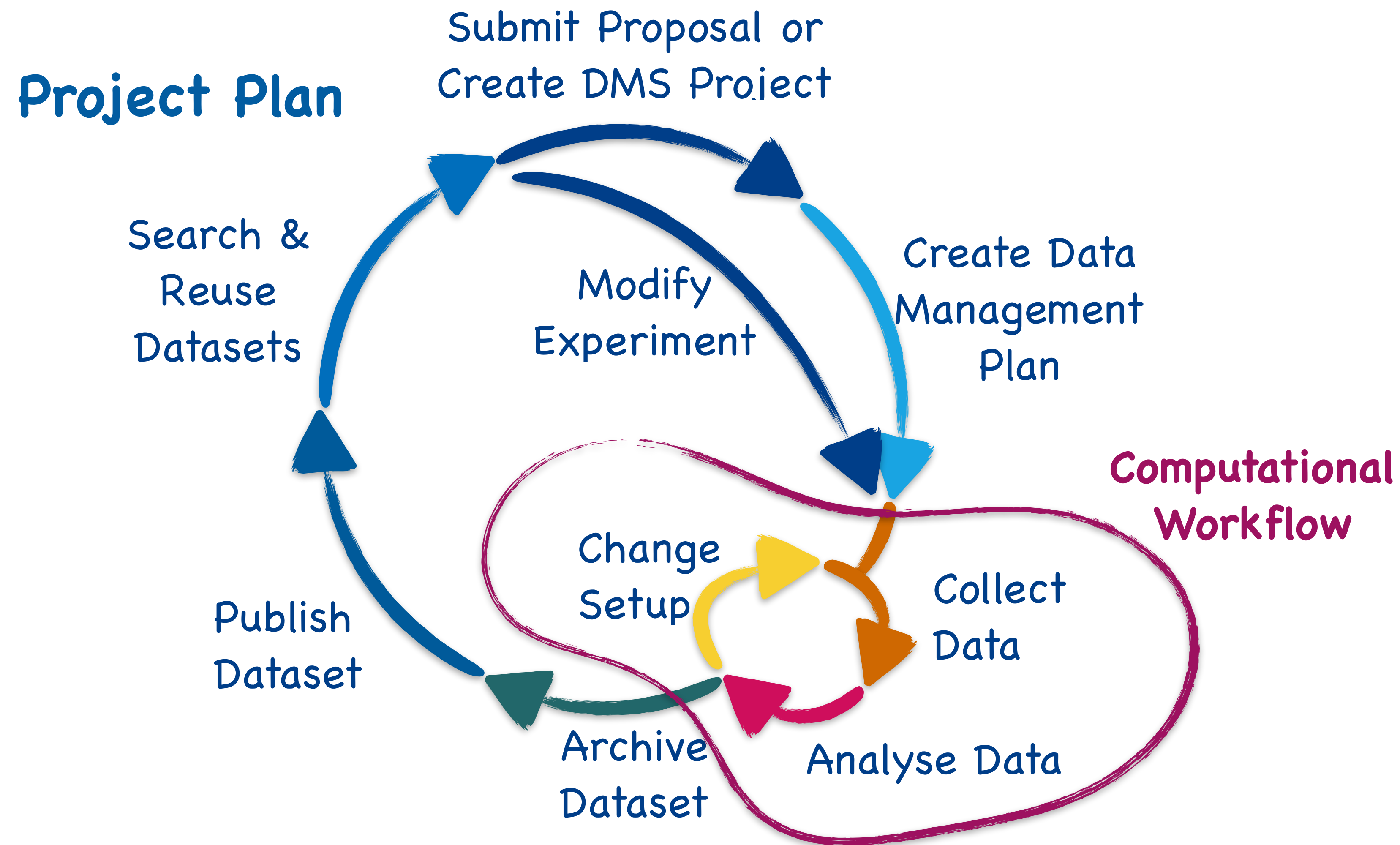
# Handle Management Support at HZDR

**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.



# Scientific Workflow

HELIPORT has an build-in Integration of Scientific Workflows



# Scientific Software Development and Reproducible Workflows

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

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



Workflow Engine

Version Control (GitLab)

Compute (HPC/OpenStack)



HELIPORT Edit a Scientific Workflow

Name: curl and cat stdout and stderr

Description:

Workflow diagram showing steps: link, curl, cat, cat\_

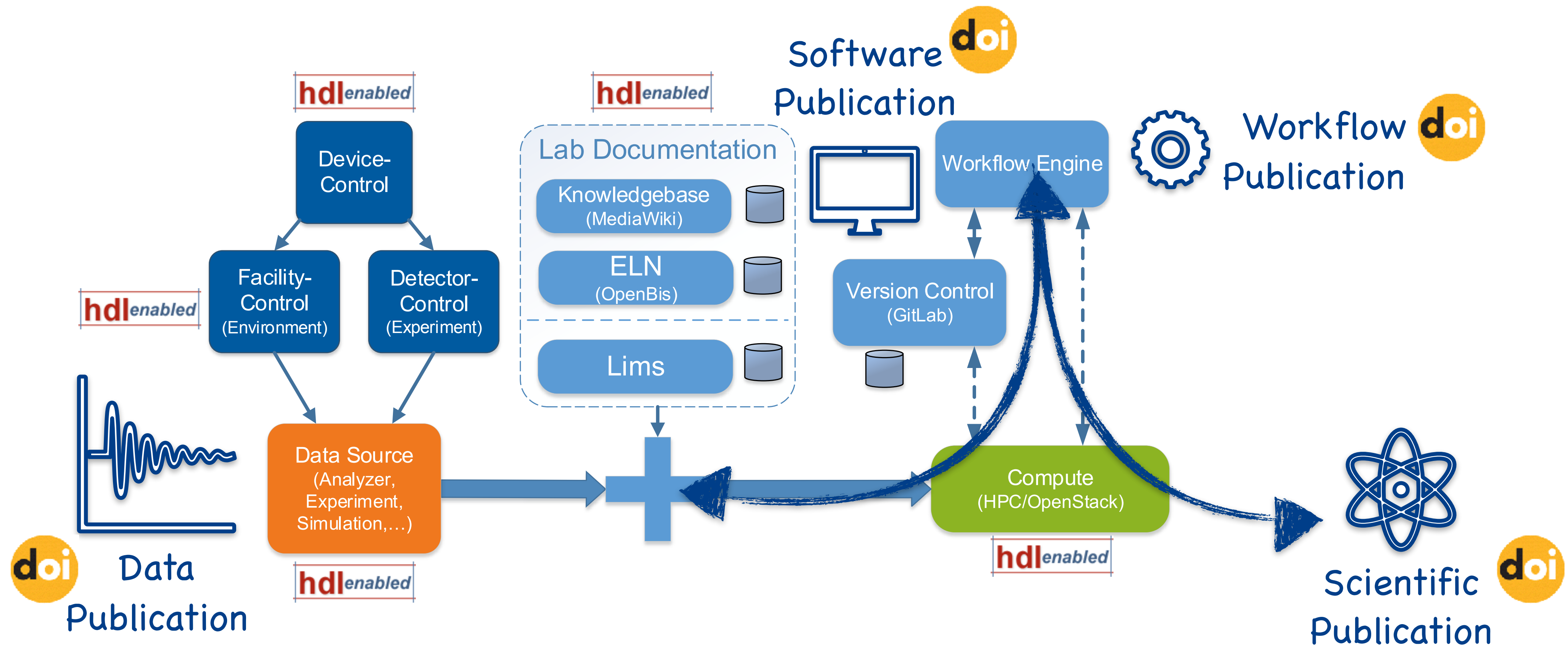
Buttons: Save, Cancel, Fit to Screen, Delete Selection

ID	Name	Description
35	echo	



# Different Types of Publications

An Example for a Complete FAIR Research Experiment



# Conclusions

- The metadata is distributed over many different schemes, systems and services.
- With PIDs or Handles we can link metadata schemes to bring all in relation to each other.
- For all facets of an experiment we need multiple different metadata schemes.
- A guidance system or infrastructure, gathering all the metadata from all systems is desirable and leads us towards a completely **FAIR** research Project.

