Research and Metadata Management at HZDR

Research data and Metadata Management Workshop - Health // 21 May 2021
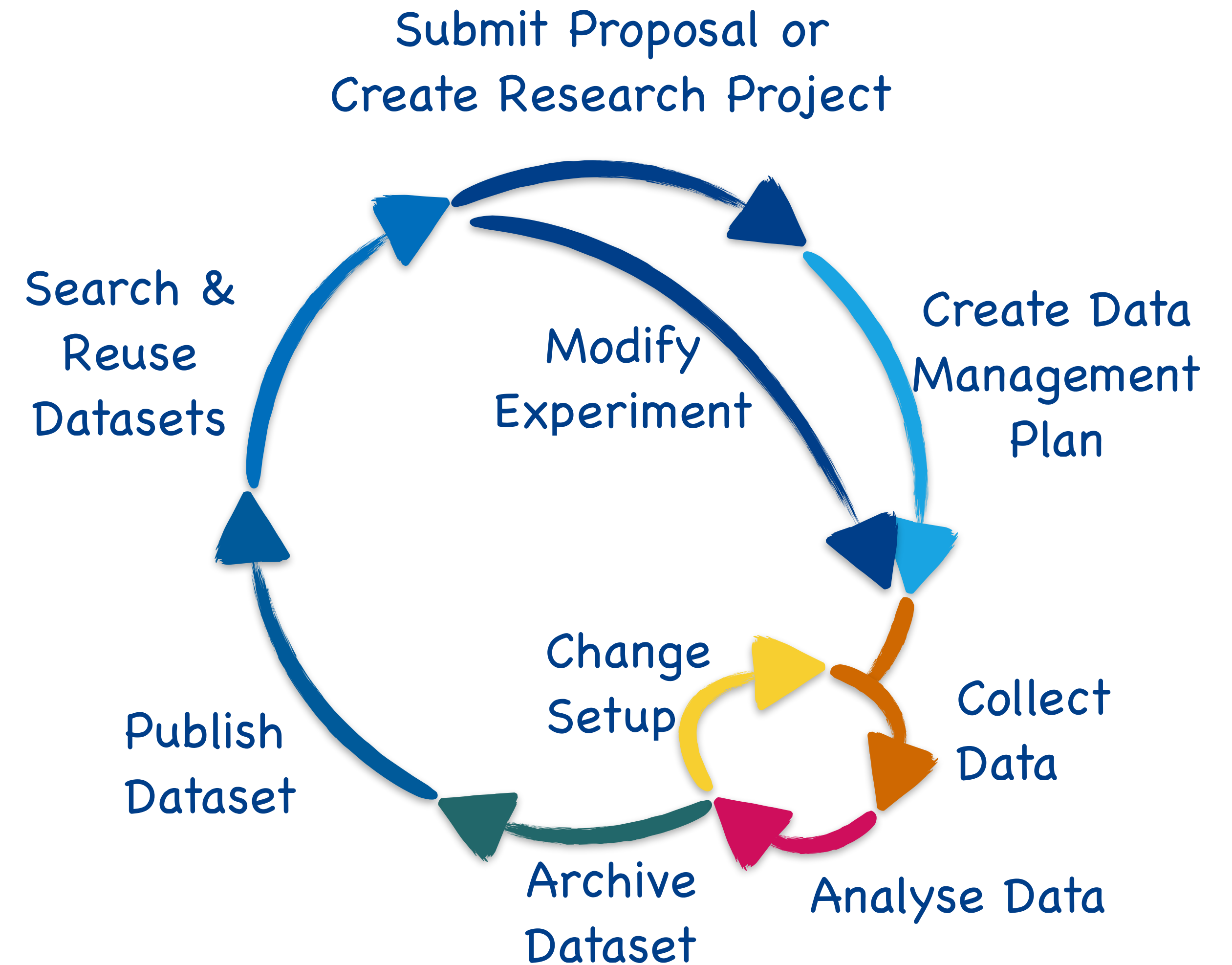Oliver Knodel // contact: o.knodel@hzdr.de
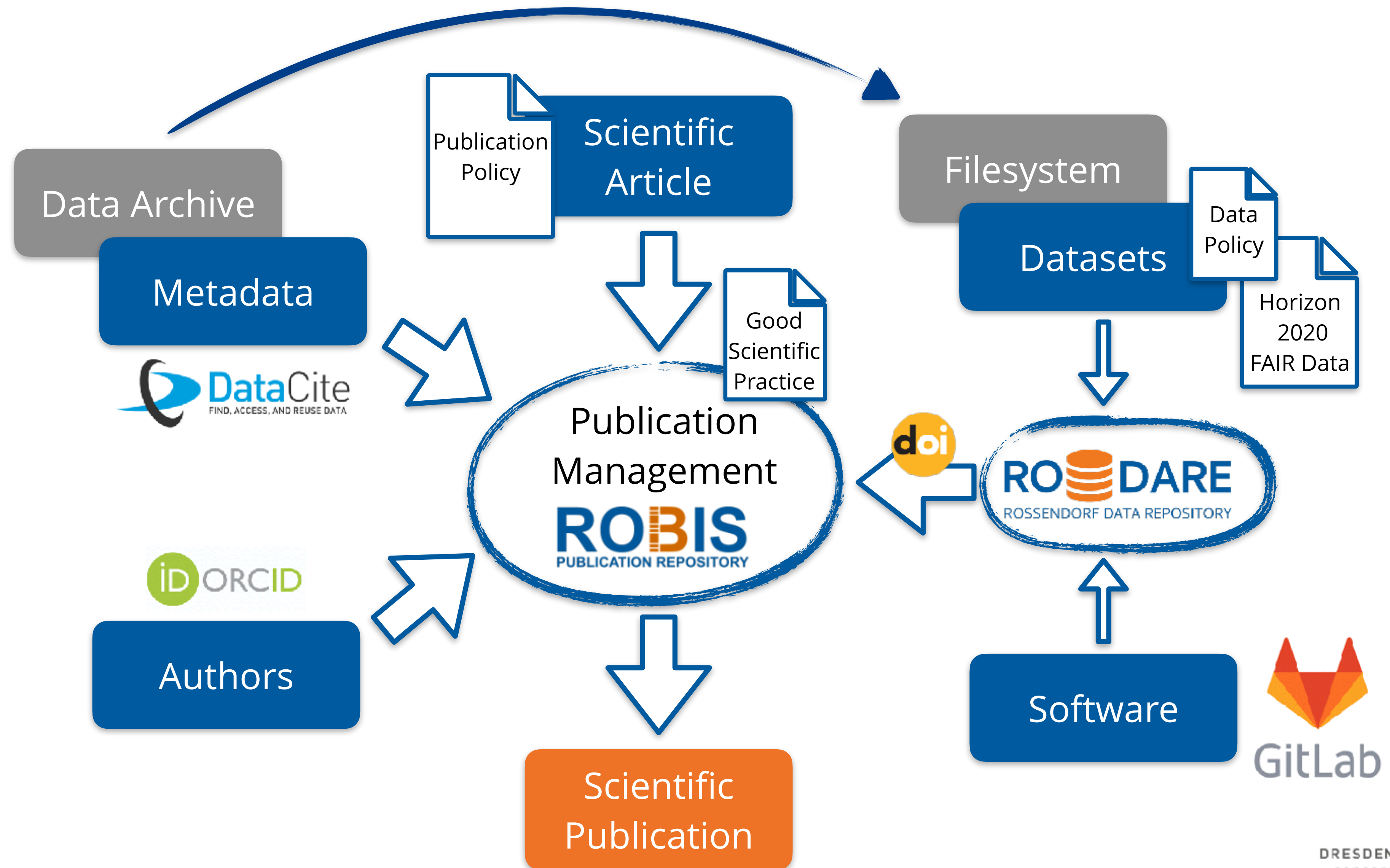
Scientific Data and Project Lifecycle at HZDR

— 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.

Submit Proposal or Create Research Project

Create Data Management Plan

Modify Experiment

Search & Reuse Datasets

Collect Data

Change Setup

Analyse Data

Publish Dataset

Archive Dataset

DRESDEN concept

HZDR

HZDR Publication Ecosystem

Our Contribution to Support FAIR Research: RODARE

https://rodare.hzdr.de

Metadata Ecosystem

Based on DataCite

DOI `10.14278/rodare.939`

Project

References by Handles to allow traceable relationships

Data Publication

DataCite and experiment specific metadata

Software Publication

ADMS.SW

Workflow

CWL Metadata

Scientific Publication

Dublin Core

Our Objective

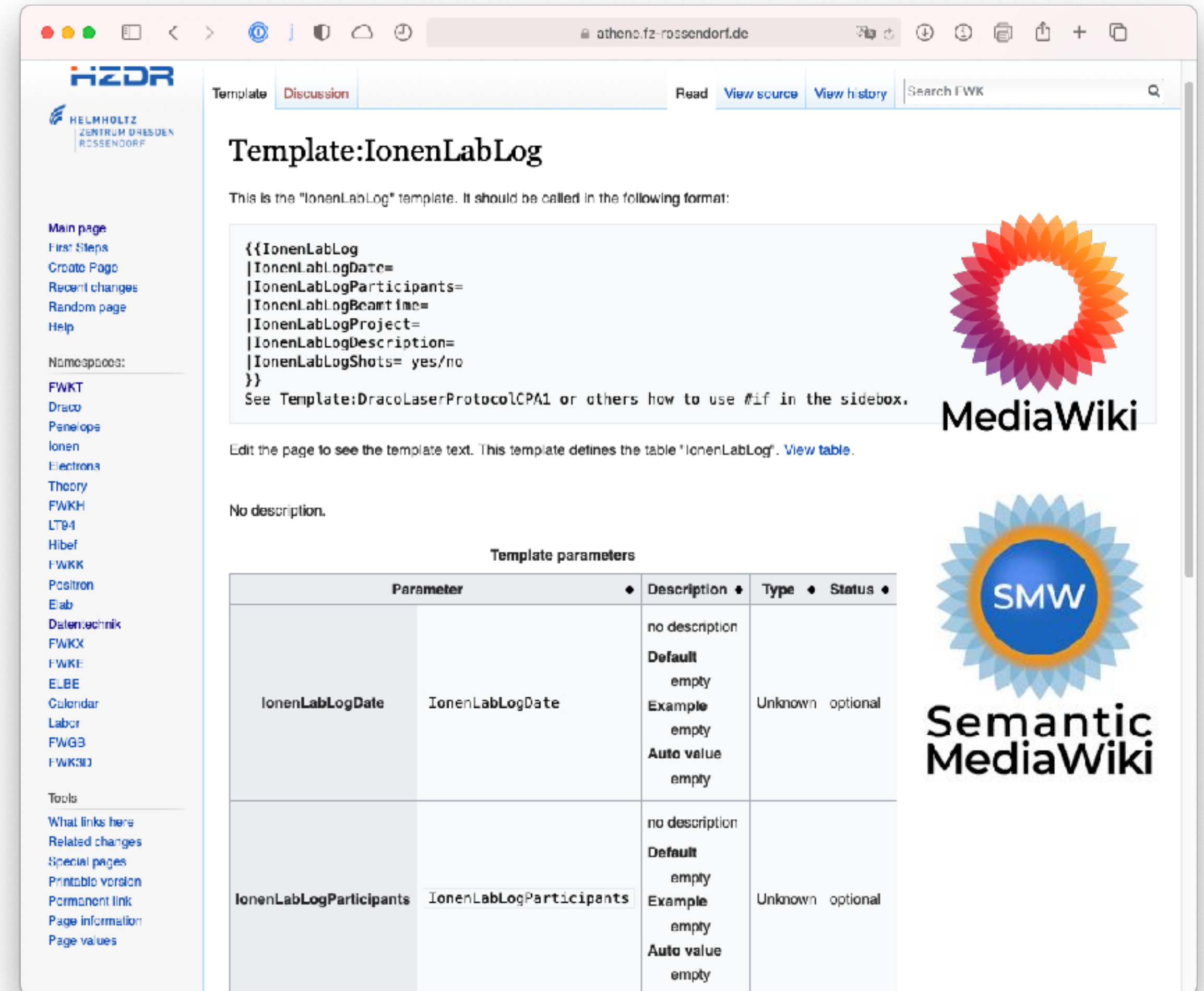— In all stages of an experiment we combine information about involved services with PIDs.

— 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

Findable Accessible Interoperable Reusable

DRESDEN concept   HZDR
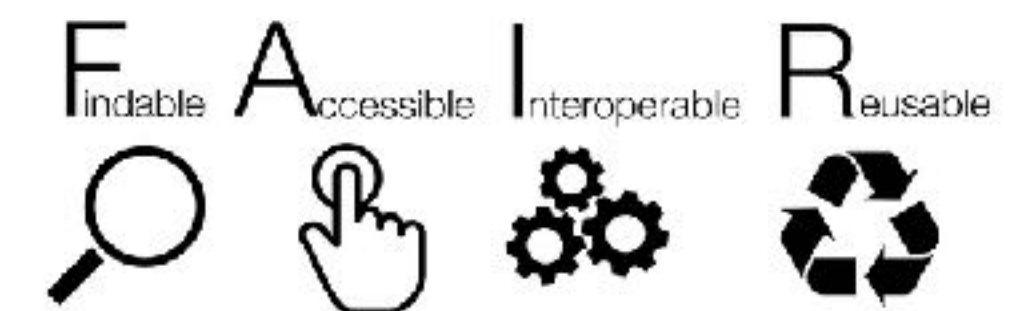
Experiment Specific Metadata — Still a Challenge…

— 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…

Research Data Management Plans based on Metadata

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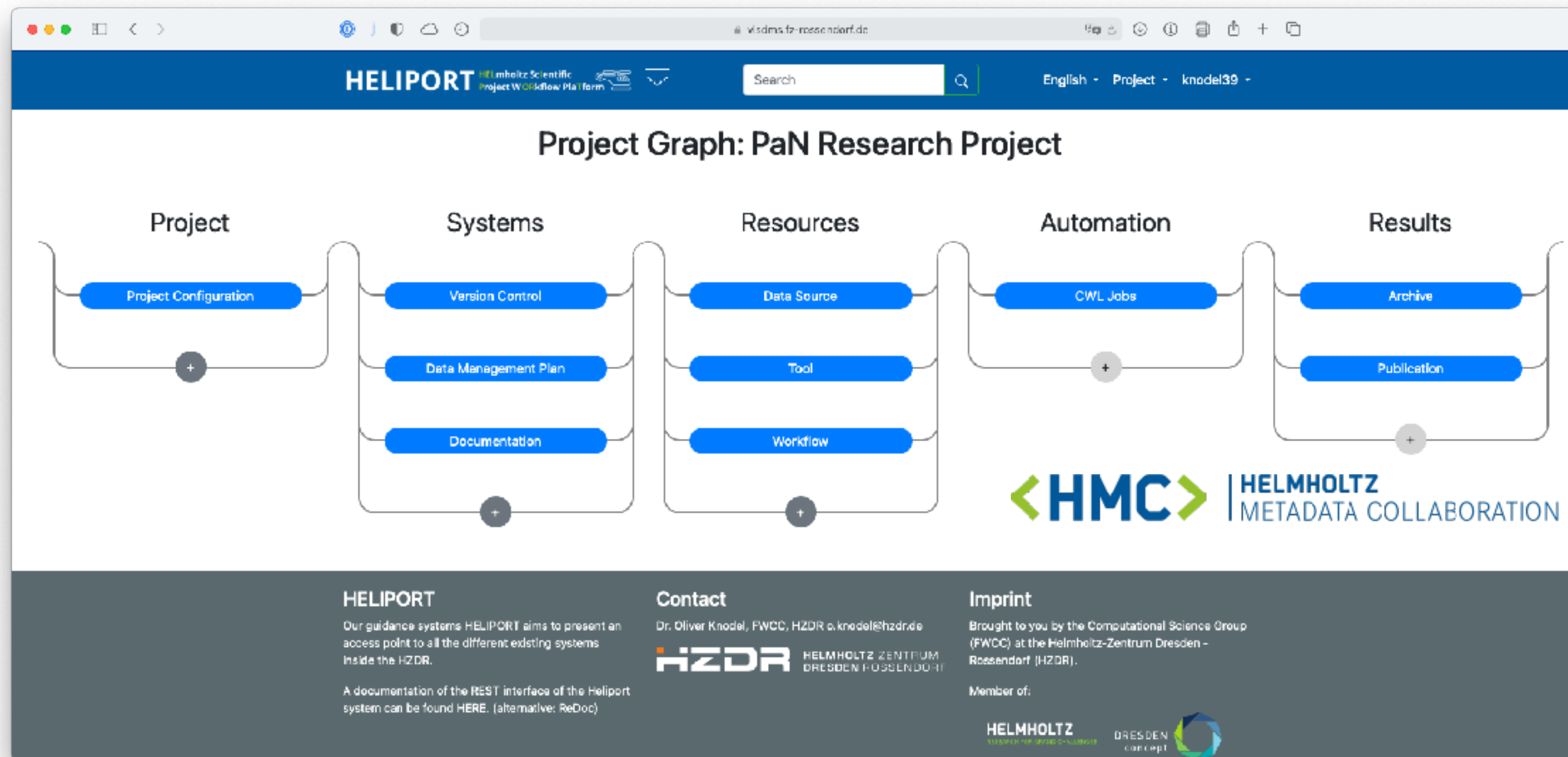
— … it is possible to generate a Data Management Plan (DMP) and can keep it up to date during the whole lifecycle of the experiment.
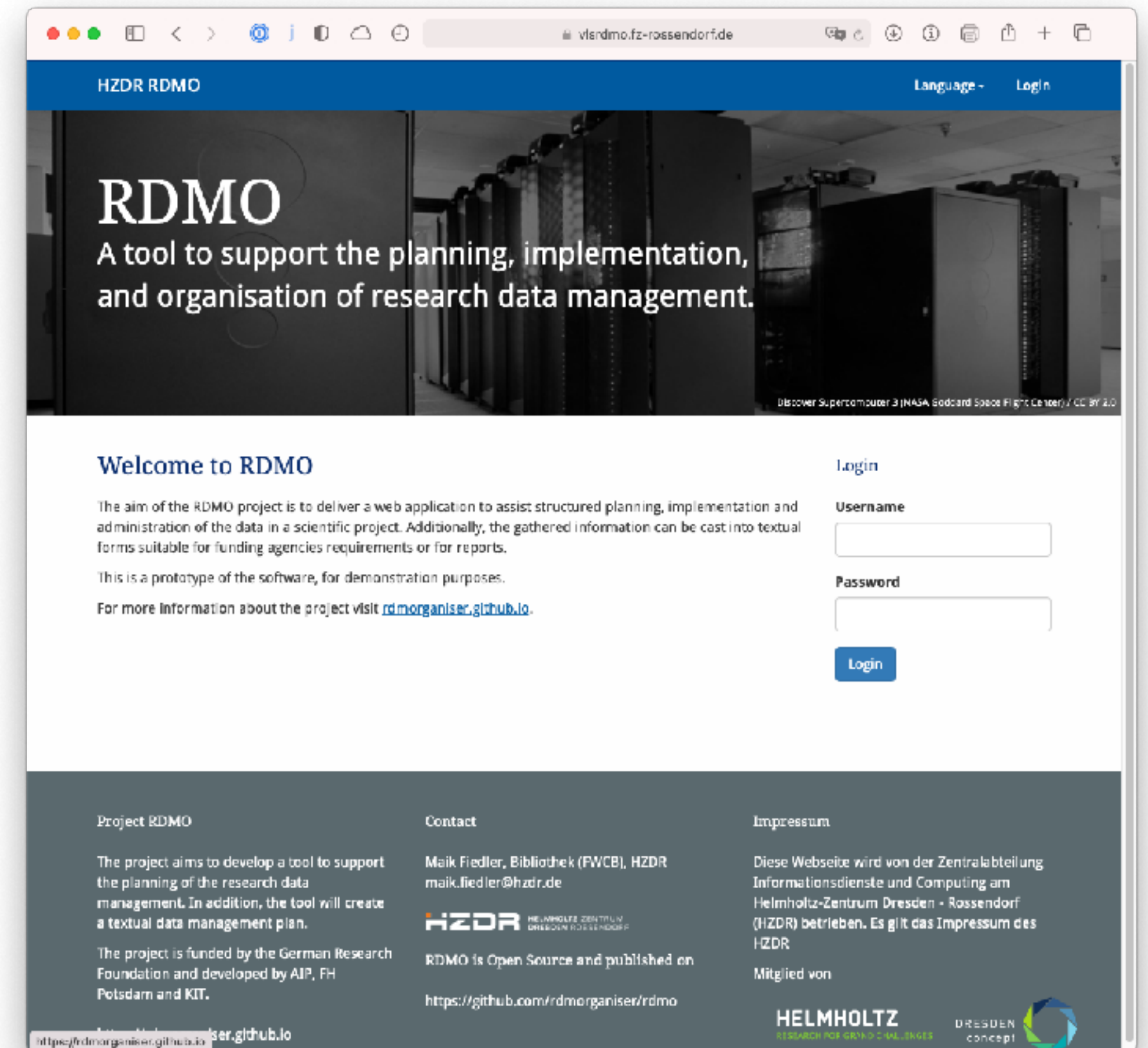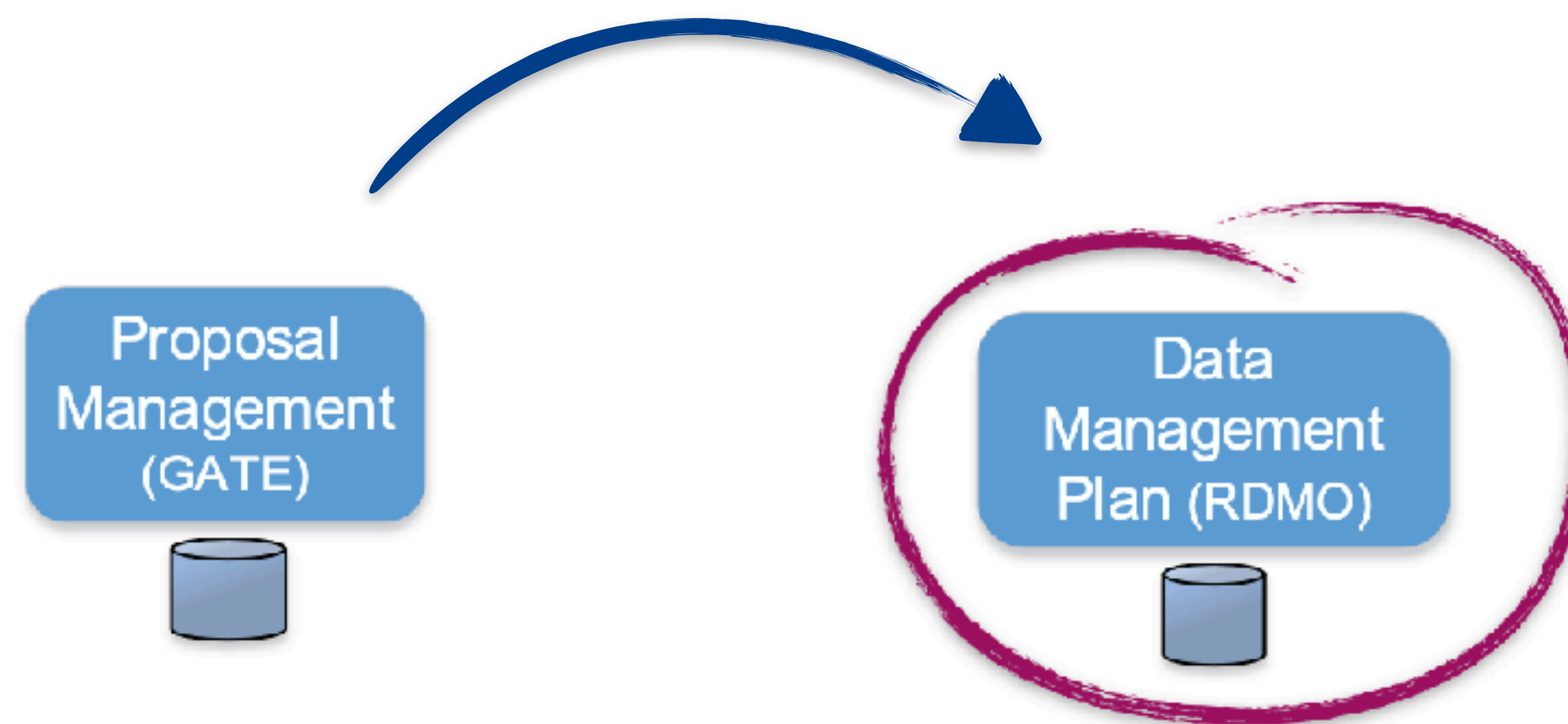
HELIPORT

HELmholtz Sc**I**entific
Project W**OR**kflow Pla**T**form
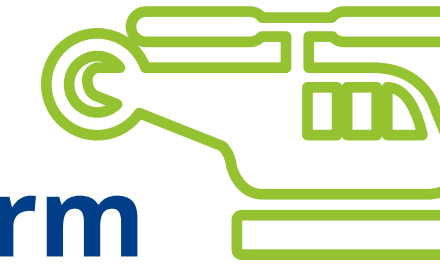
> The HELIPORT project aims at developing a platform which accommodates the **complete life cycle** of a scientific project and links all corresponding programs, systems, workflows and metadata schemas to create a more **FAIR** and comprehensible research project.
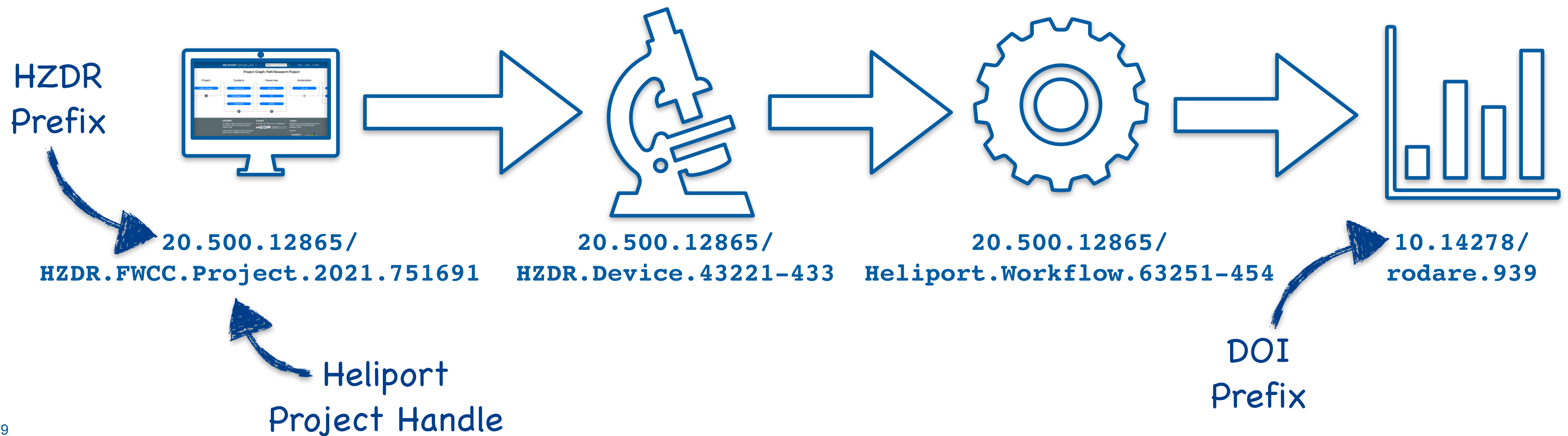
Handle Management Support at HZDR

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.

HZDR
Prefix

`20.500.12865/`
`HZDR.FWCC.Project.2021.751691`

Heliport
Project Handle

`20.500.12865/`
`HZDR.Device.43221-433`

`20.500.12865/`
`Heliport.Workflow.63251-454`

`10.14278/`
`rodare.939`

DOI
Prefix

DRESDEN concept

HZDR

Scientific Workflow

HELIPORT has an build-in Integration of Scientific Workflows
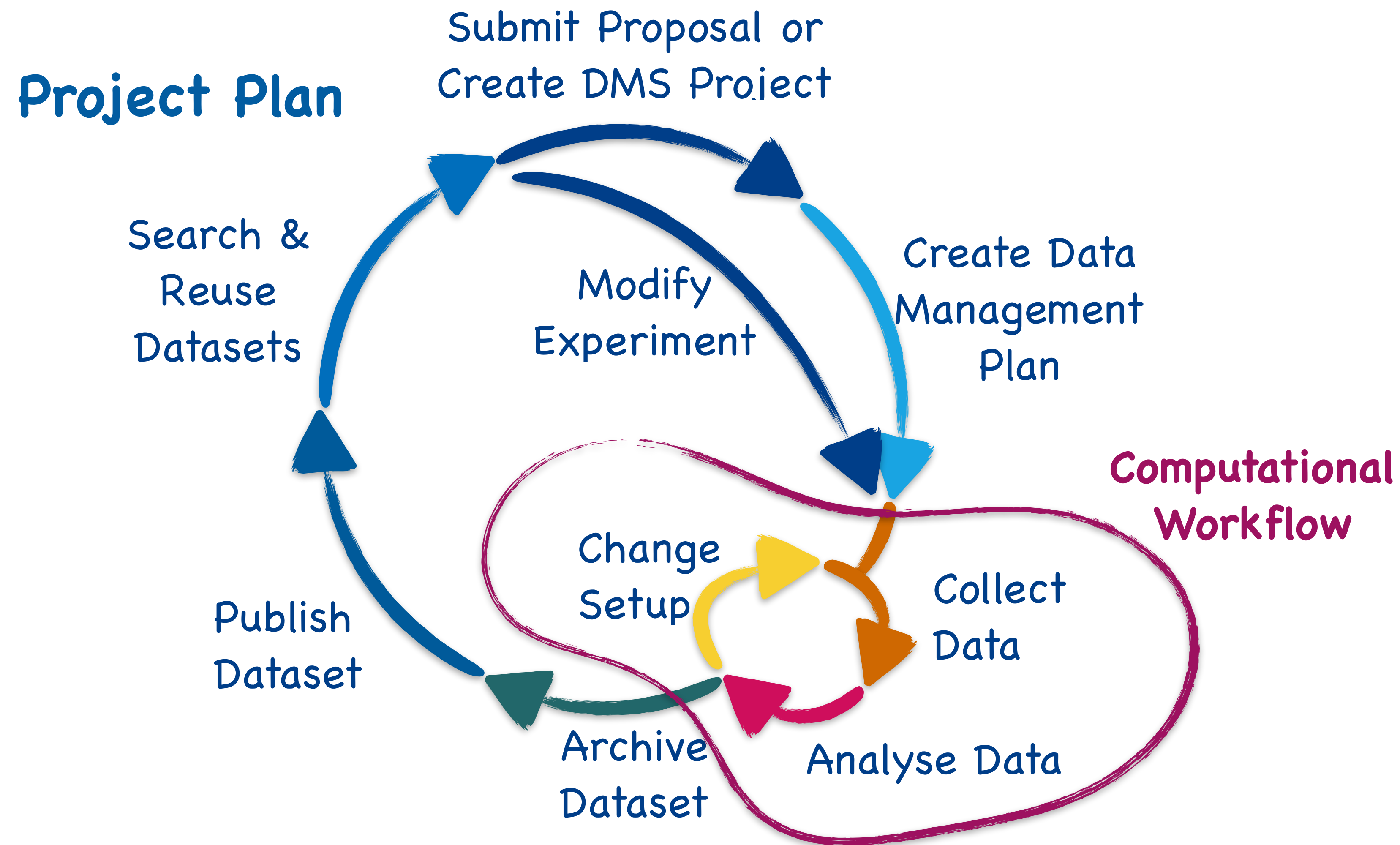
Scientific Software Development and Reproducible Workflows

Workflow Engine

Version Control
(GitLab)

Compute
(HPC/OpenStack)

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

COMMON WORKFLOW LANGUAGE

GitLab

slurm
workload manager

UNICORE

DRESDEN concept

HZDR

Different Types of Publications
An Example for a Complete FAIR Research Experiment
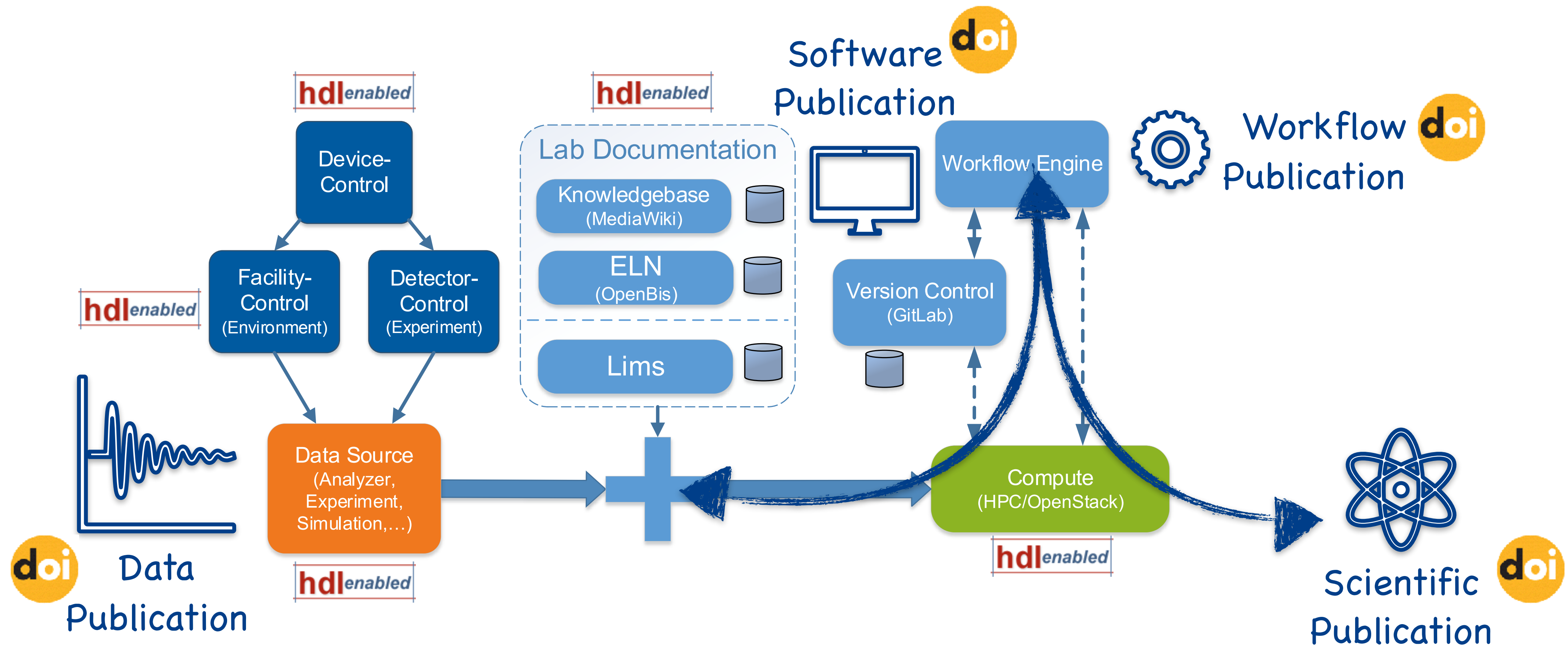
Conclusions

— 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.