HZDR
HELMHOLTZ ZENTRUM
DRESDEN ROSSENDORF

InvenioRDM at HZDR: RODARE

ROSSENDORF DATA REPOSITORY

DMA ST1 seminar, April 2024

Oliver Knodel // contact: o.knodel@hzdr.de

DRESDEN concept

Our Research Facility and our Large Scale Research Infrastructures

The Helmholtz-Zentrum Dresden - Rossendorf

— Employees approx. 1,470. Thereof 670 scientists.

— HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

Research Fields

— Energy, Health and Matter.

ELBE – Center for High-Power Radiation Sources

— Electron accelerator, free-electron lasers & THz source.

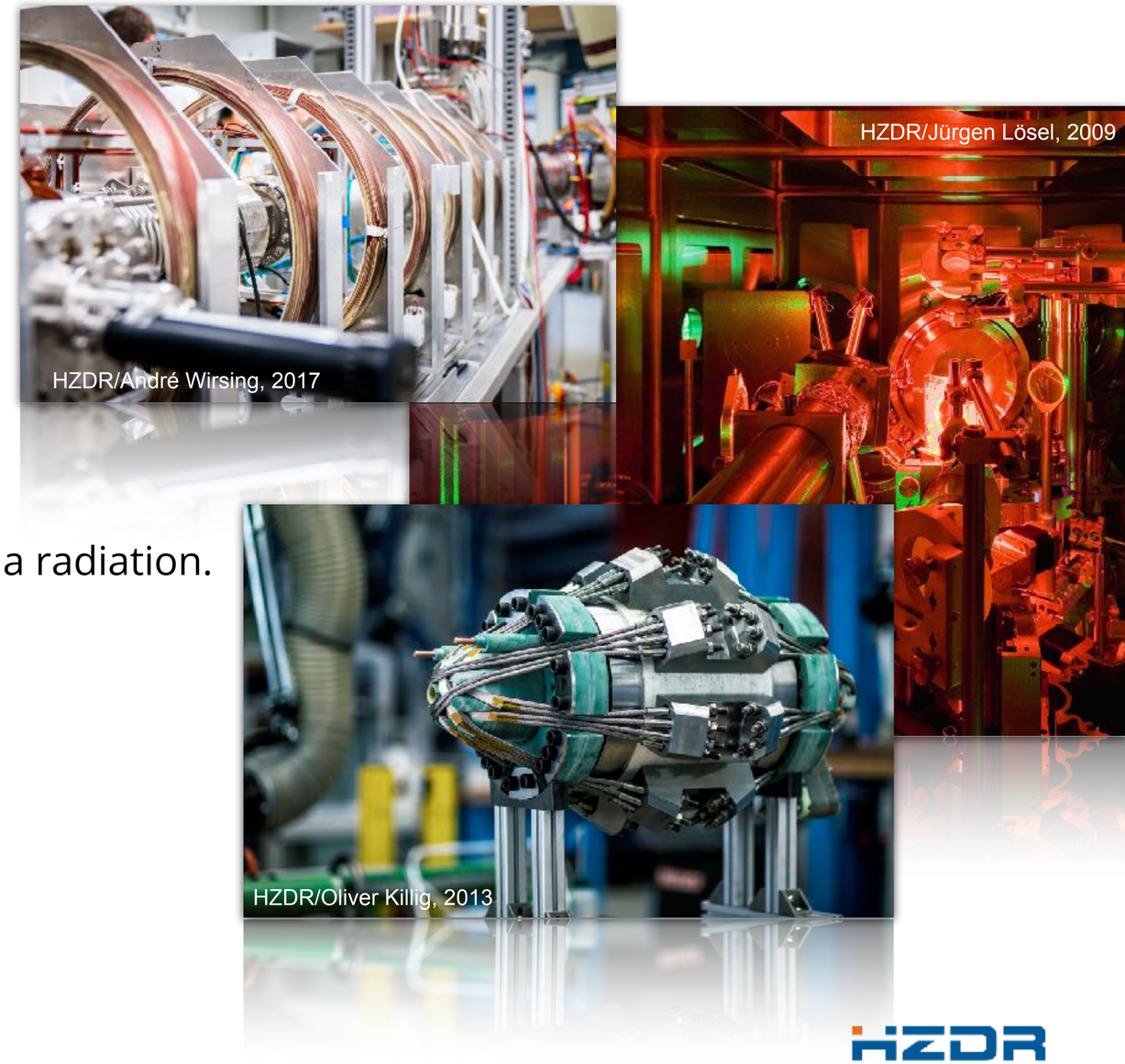
— Positrons, protons, neutrons as well as X-ray and gamma radiation.

Dresden High Magnetic Field Laboratory (HLD)

— Europe’s highest pulsed magnetic fields.

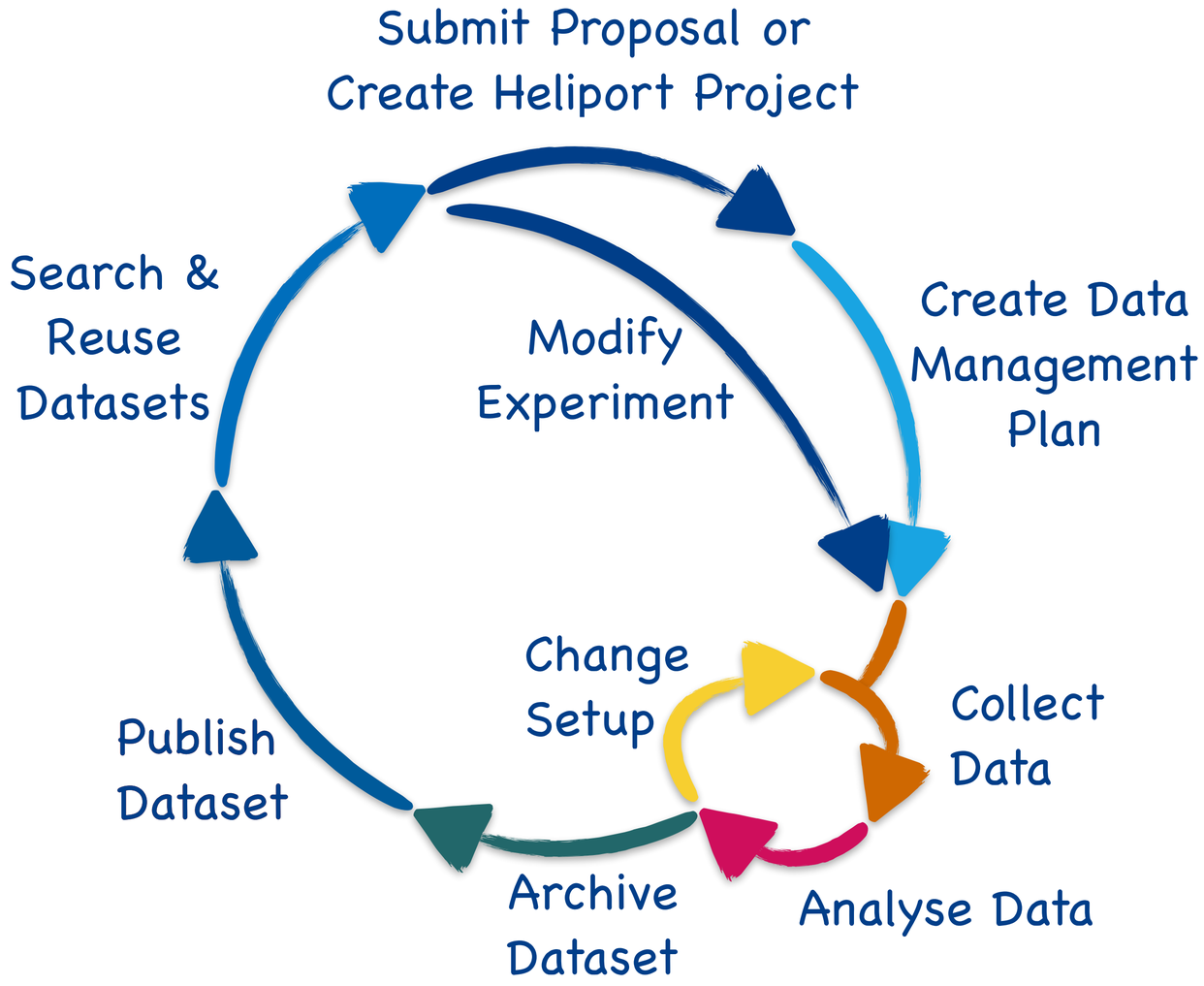
Ion Beam Center (IBC)

— Nanoscale surface analysis and modification.



The Overall Objective: 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.
- Data is an essential aspect of research
- A FAIR data publication should be part of every journal publication



Our Motivation for Developing a Full Digital Data Lifecycle:

- The HZDR has a data policy since May 2018.
- Reasons for the development:
  - Establishment and legitimisation of research data management at the HZDR,
  - Legal framework for data management and publication.
- Foundation for the development of tools to support our scientists:

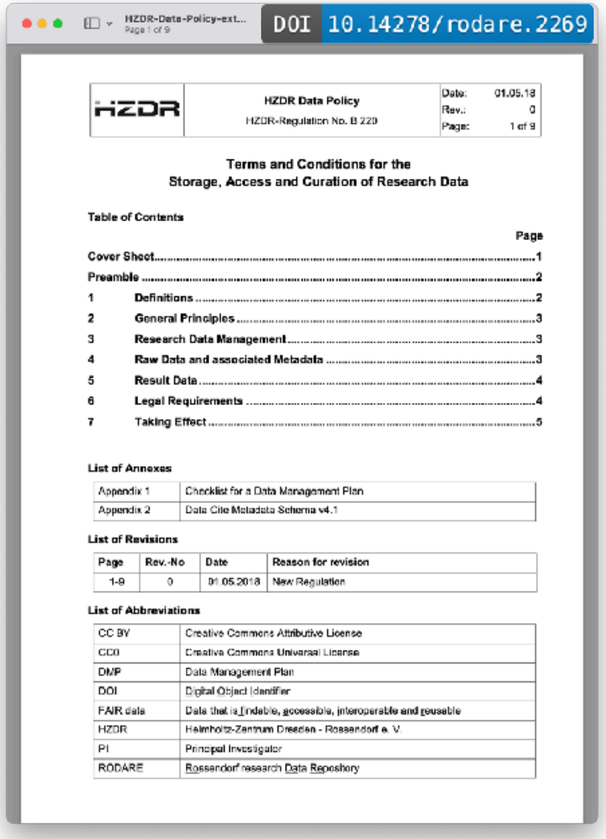
RODARE ROSSENDORF DATA REPOSITORY

ROBIS PUBLICATION REPOSITORY

HELIPORT HELmholtz Scientific Project W ORkflow PlaTform

The HZDR Data Policy

© Helmholtz-Gemeinschaft 2023, https://os.helmholtz.de/open-research-data/forschungsdaten-policies/



HZDR-Data-Policy-ext... Page 1 of 9 DOI 10.14278/rodare.2269

HZDR | HZDR Data Policy, HZDR-Regulation No. B 220 | Date: 01.05.18, Rev.: 0, Page: 1 of 9

Terms and Conditions for the Storage, Access and Curation of Research Data

Table of Contents

List of Annexes

Appendix 1	Checklist for a Data Management Plan
Appendix 2	Data Cite Metadata Schema v4.1

List of Revisions

Page	Rev.-No	Date	Reason for revision
1-9	0	01.05.2018	New Regulation

List of Abbreviations

CC BY	Creative Commons Attributive License
CC0	Creative Commons Universal License
DMP	Data Management Plan
DOI	Digital Object Identifier
FAIR data	Data that is findable, accessible, interoperable and reusable
HZDR	Helmholtz-Zentrum Dresden - Rossendorf e. V.
PI	Principal Investigator
RODARE	Rossendorf research Data Repository

HZDR

HZDR Data Policy: Data Storage and Access

For research data the storage and safeguarding for at least ten (10) years corresponds to "good scientific practice" [5].

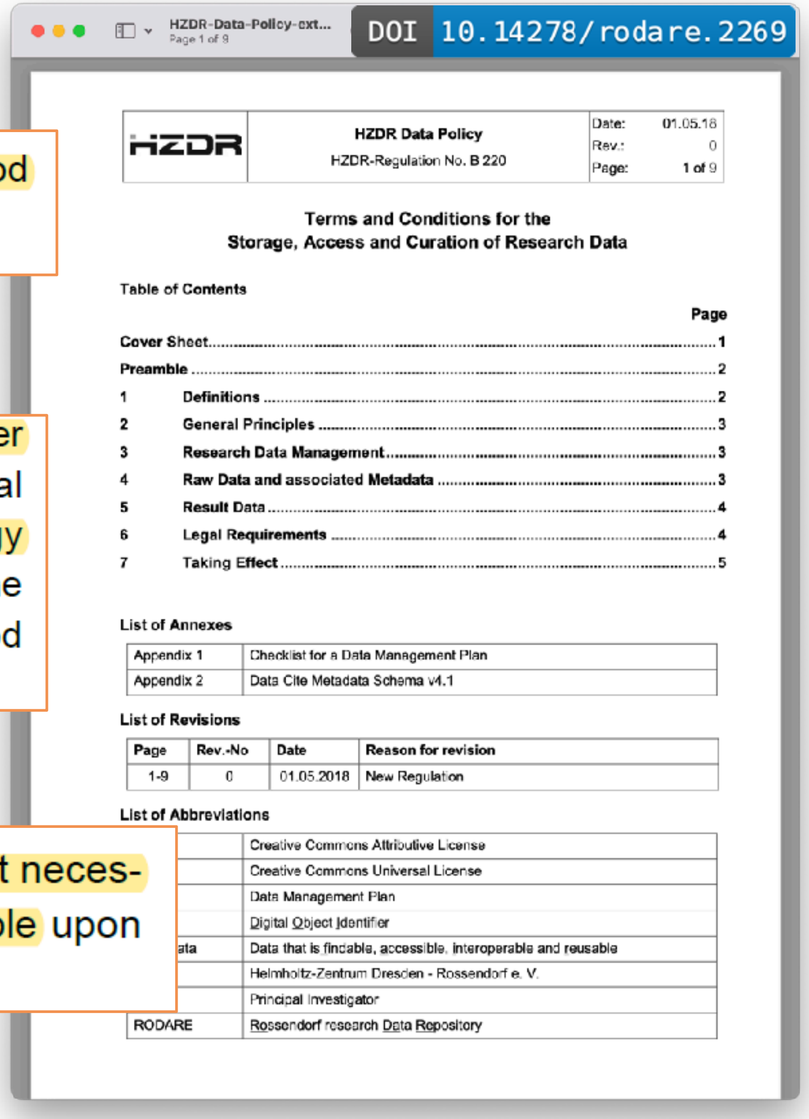
HZDR DATA Policy, 3 Research Data Management, page 3

Access to raw data and the associated metadata may be restricted to the members of the User Group for an embargo period of five (5) years after the end of the experiments of the individual research project. Thereafter, if legally permitted or if not necessarily required for Technology Transfer, it will be made openly accessible with HZDR acting as custodian. Any member of the Right Holder Group that wishes to maintain the restricted access to its data for a longer period will be required to file a corresponding request to the HZDR management.

HZDR DATA Policy, 4 Raw Data and associated Metadata, page 4

Access to Result Data is restricted to the User Group. As far as legally permitted or if not necessarily required for Technology Transfer, the Result Data may be made openly accessible upon request of the PI.

HZDR DATA Policy, 5 Result Data, page 4


HZDR-Data-Policy-ext...
Page 1 of 9
DOI 10.14278/rodare.2269

HZDR Data Policy
HZDR-Regulation No. B 220
Date: 01.05.18
Rev.: 0
Page: 1 of 9

Terms and Conditions for the Storage, Access and Curation of Research Data

Table of Contents

List of Annexes

Appendix 1	Checklist for a Data Management Plan
Appendix 2	Data Cite Metadata Schema v4.1

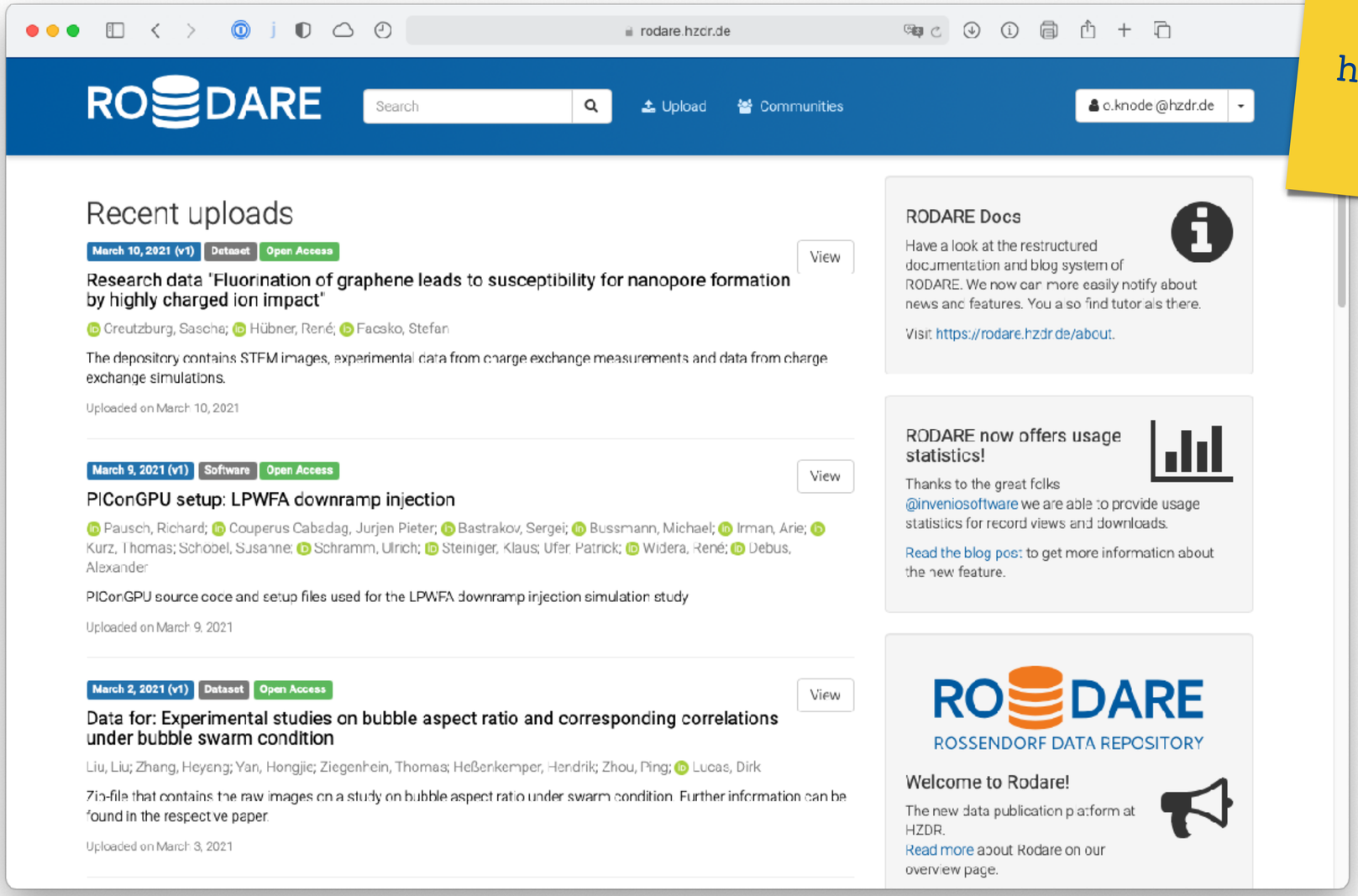
List of Revisions

Page	Rev.-No	Date	Reason for revision
1-9	0	01.05.2018	New Regulation

List of Abbreviations

	Creative Commons Attributive License
	Creative Commons Universal License
	Data Management Plan
	Digital Object Identifier
ata	Data that is findable, accessible, interoperable and reusable
	Helmholtz-Zentrum Dresden - Rossendorf e. V.
	Principal Investigator
RODARE	Rossendorf research Data Repository

Our Contribution to Support FAIR Research: RODARE



https://rodare.hzdr.de

Powered by:

INVENIO

Metadata:

DataCite
FIND, ACCESS, AND REUSE DATA

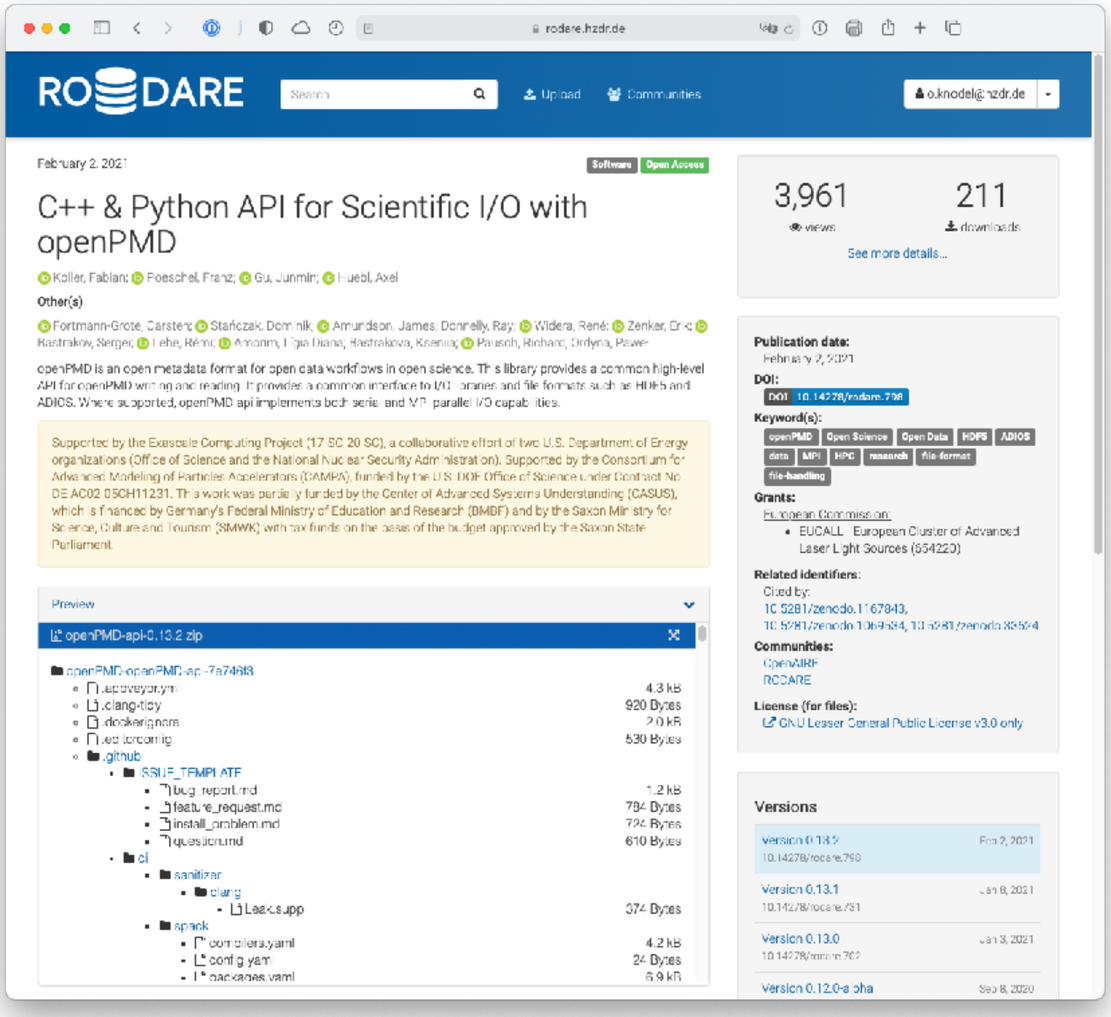
Registered in:

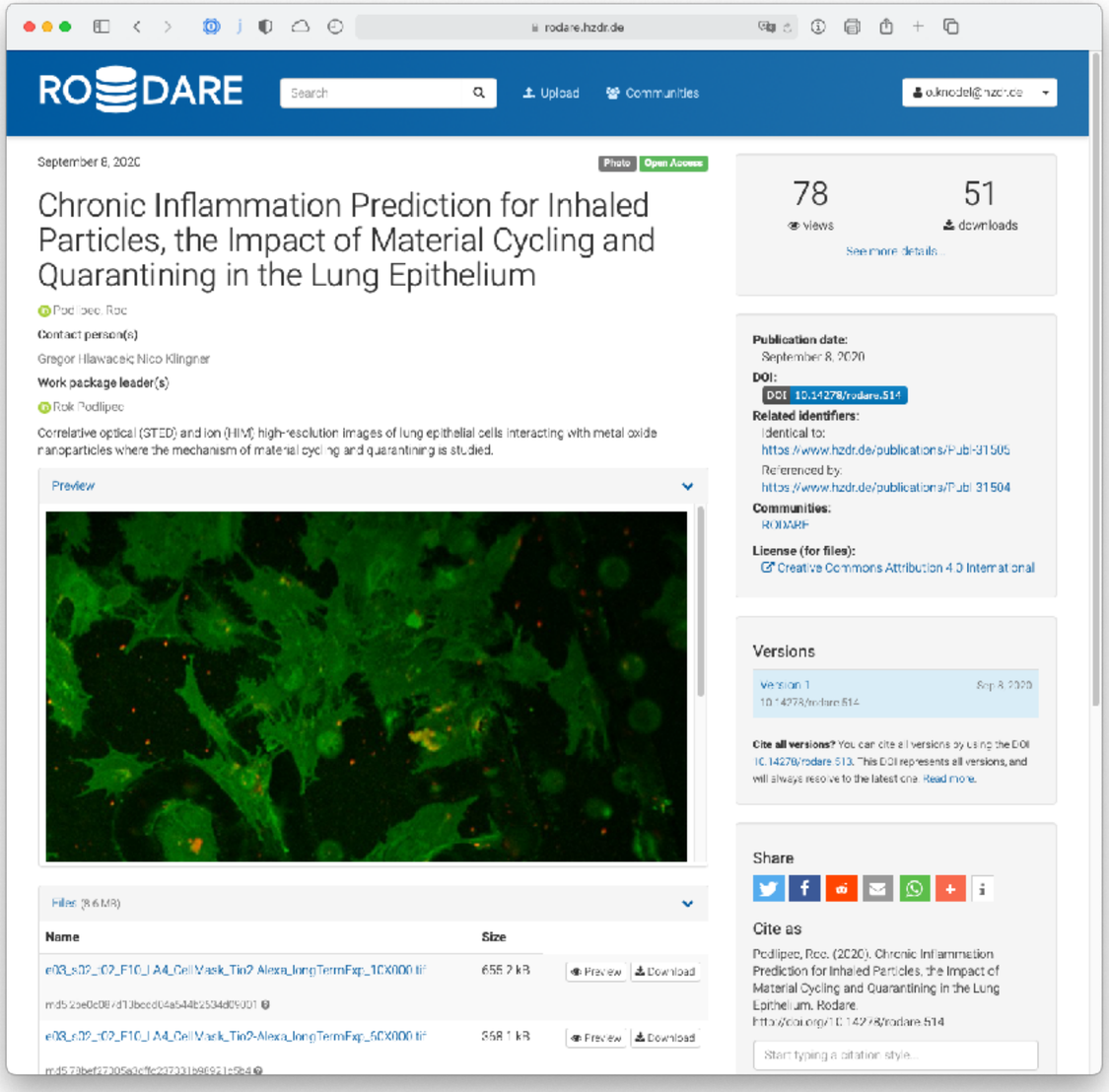
re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

http://doi.org/10.17616/R3BR40

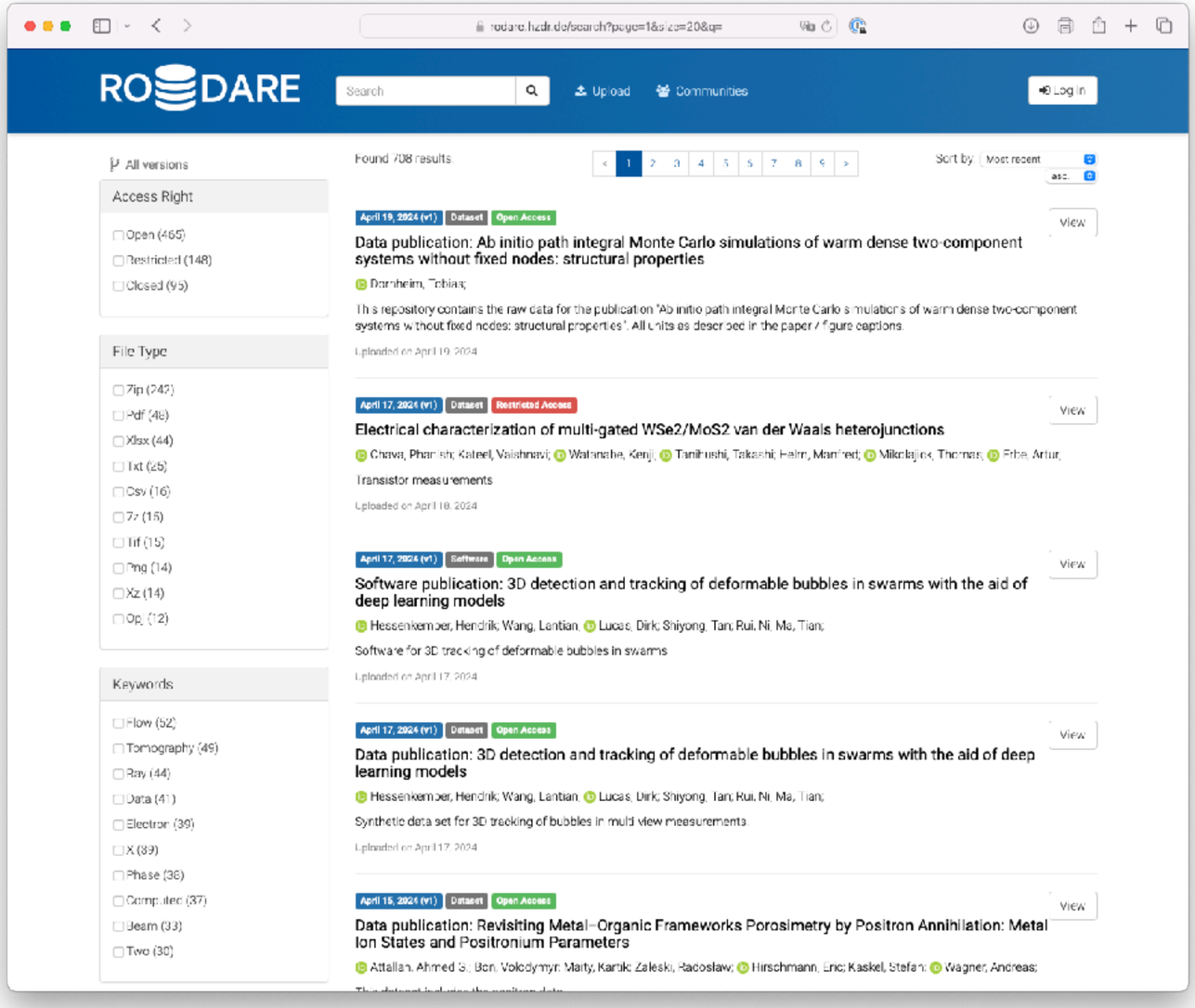
DRESDEN concept HZDR

Different Upload Types and Previews





Search and Filter on Metadata (DataCite)

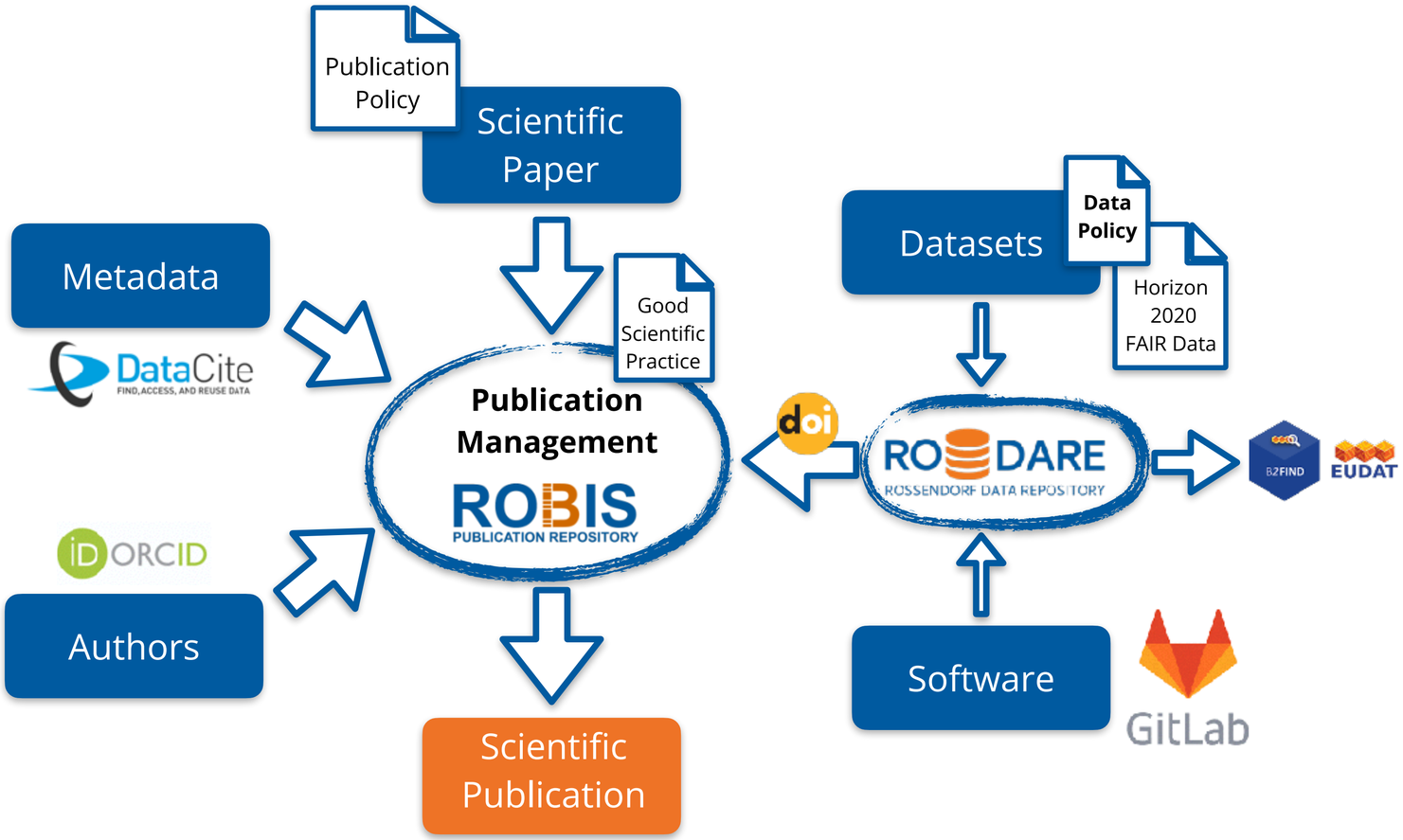



```
{
  "aggregations": {
    "access_right": {
      "buckets": [
        {
          "doc_count": 465,
          "key": "open"
        },
        {
          "doc_count": 148,
          "key": "restricted"
        },
        {
          "doc_count": 95,
          "key": "closed"
        }
      ],
      "doc_count_error_upper_bound": 0,
      "sum_other_doc_count": 0
    },
    "file_type": {
      "buckets": [
        {
          "doc_count": 242,
          "key": "zip"
        },
        {
          "doc_count": 48,
          "key": "pdf"
        },
        {
          "doc_count": 44,
          "key": "xlsx"
        },
        {
          "doc_count": 25,
          "key": "txt"
        },
        {
          "doc_count": 16,
          "key": "csv"
        },
        {
          "doc_count": 15,
          "key": "7z"
        },
        {
          "doc_count": 15,
          "key": "tif"
        },
        {
          "doc_count": 14,
          "key": "png"
        },
        {
```

Consequences for our Publication Systems

— For data publications we developed the Rossendorf Data Repository.

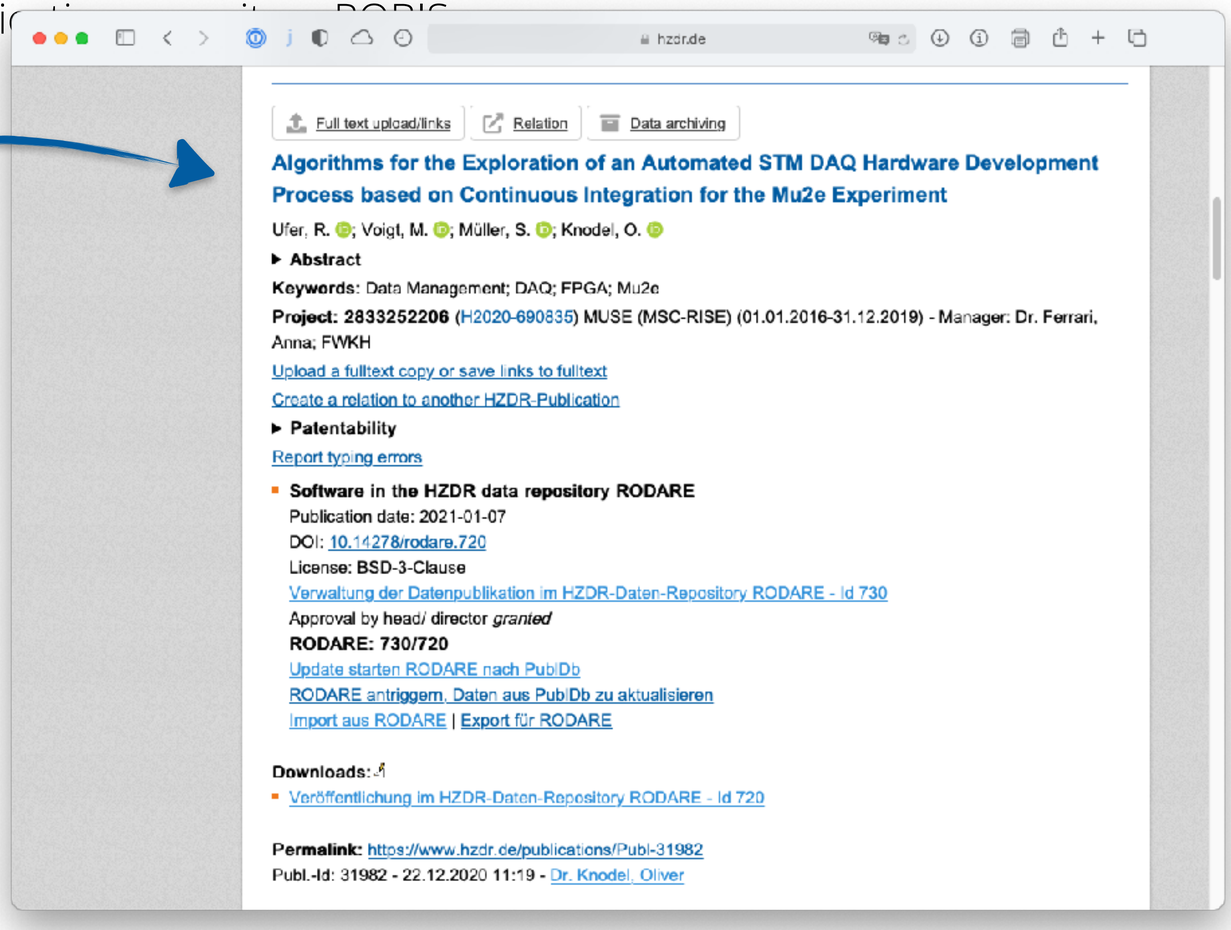
— A complex integration into the HZDR publication system ROBIS followed...



Integrated in the HZDR Publication Repository

- HZDR publications must be registered in our overall publi[illegible] ROBIS.
- Before publishing you will be redirected to ROBIS.

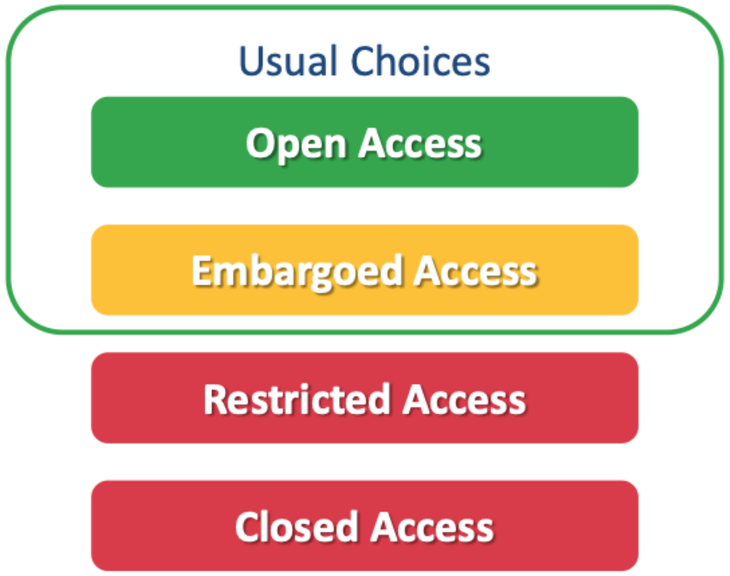
Register in HZDR publication repository



DRESDEN concept HZDR

Multiple Visibility Levels

— We know that not everything can be made publicly available. Therefore, RODARE supports four different levels of visibility for your uploads:



— The DOI landing page of a record with the corresponding metadata (DataCite) are always visible.

— The embargo period can be extended.

DOI Versioning

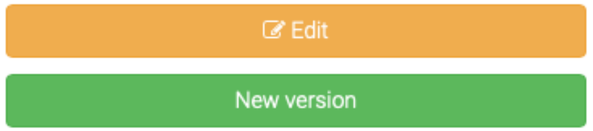
Upon first publishing 

two DOIs are **registered**:

— One DOI represents the **specific version** of the record,

— Another DOI represents **all versions** of the record.

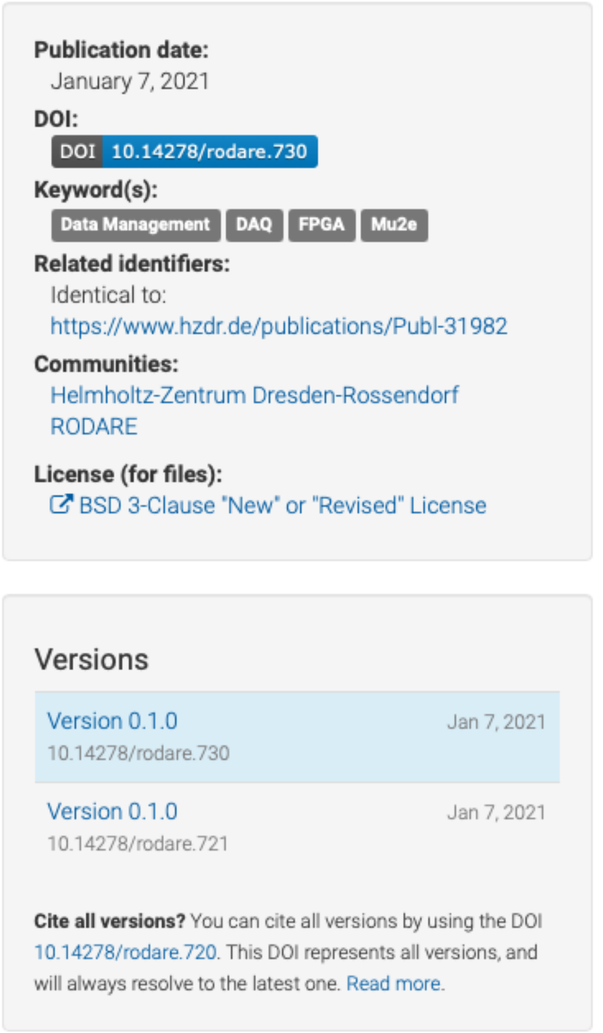
Afterwards RODARE registers a new DOI for every new version.



When do I create a new version?

— If you wish to **add**, **edit** or **update** files of your record after it has been published.

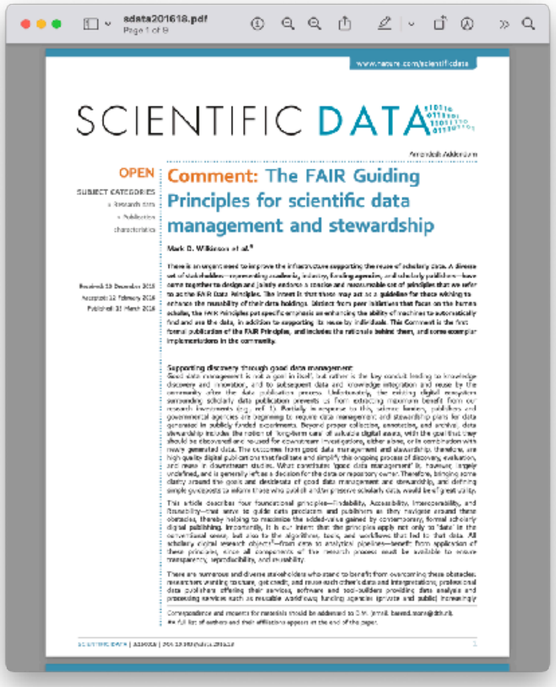
— Not necessary if you only change metadata of your upload.



The FAIR Principles and RODARE in Detail

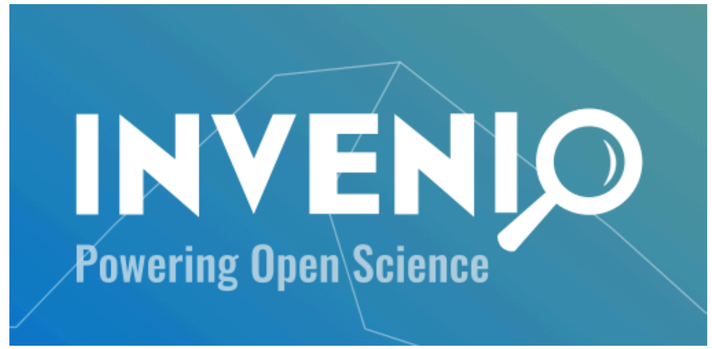
— How much of the FAIR principles we support is difficult to generalise …

— Certain things depend highly on the scientist and the community in the scientific field.

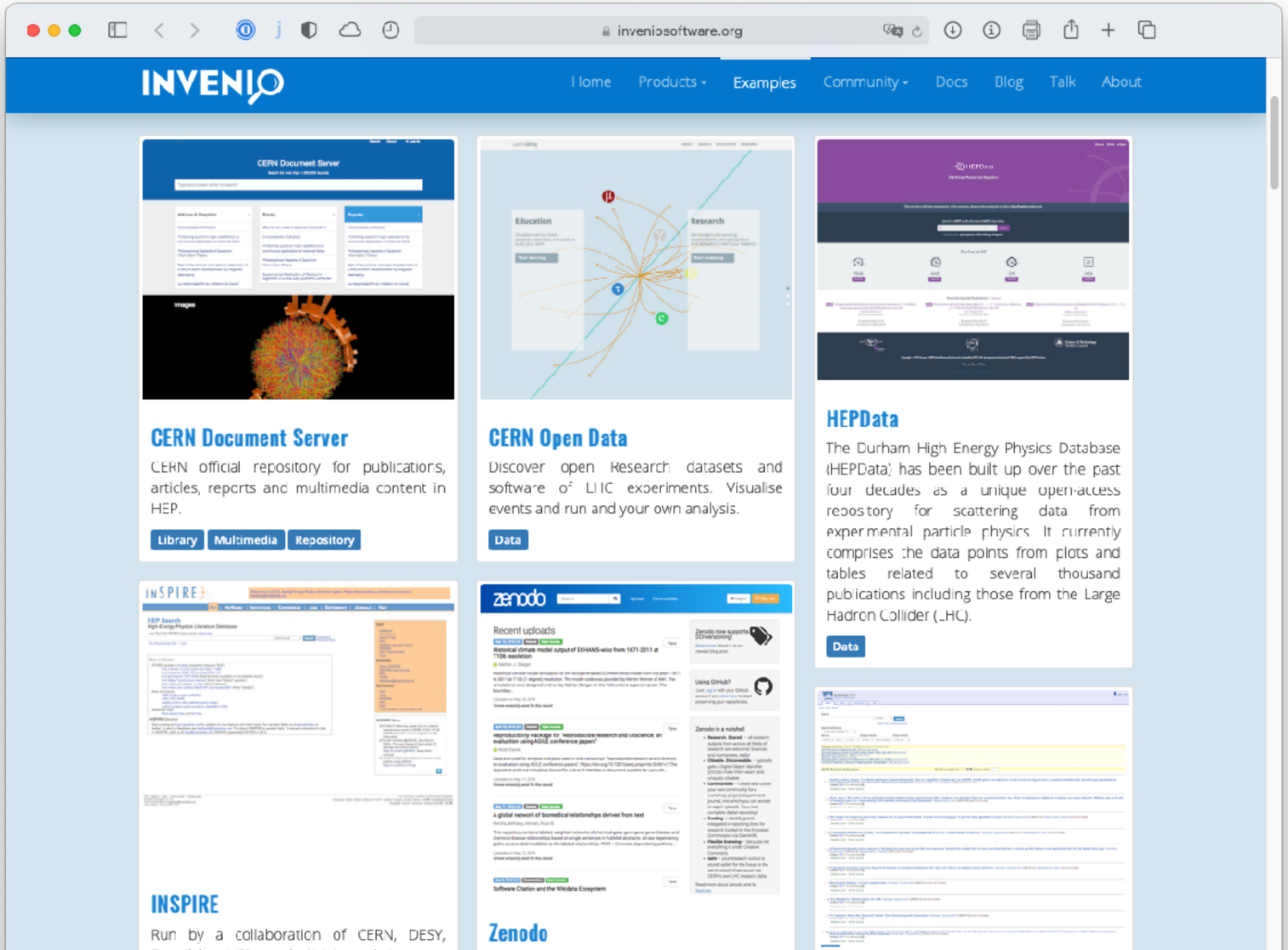


Background: Invenio → Zenodo → RODARE

- Rodare is build on top of **Zenodo**, an interdisciplinary open research data repository service
- Zenodo is built on the foundation of the Invenio digital library and adds a couple of features to the Invenio framework:
  - Persistant identifiers (DOIs),
  - Communities,
  - GitHub integration,
  - OpenAire support,
  - …
- Zenodo itself is built and operated by **CERN** and **OpenAIRE**.
- Many organizations adapted Zenodo and keeping the clones up to date is a challenge…



The InvenioRDM (Research Data Management) Projekt

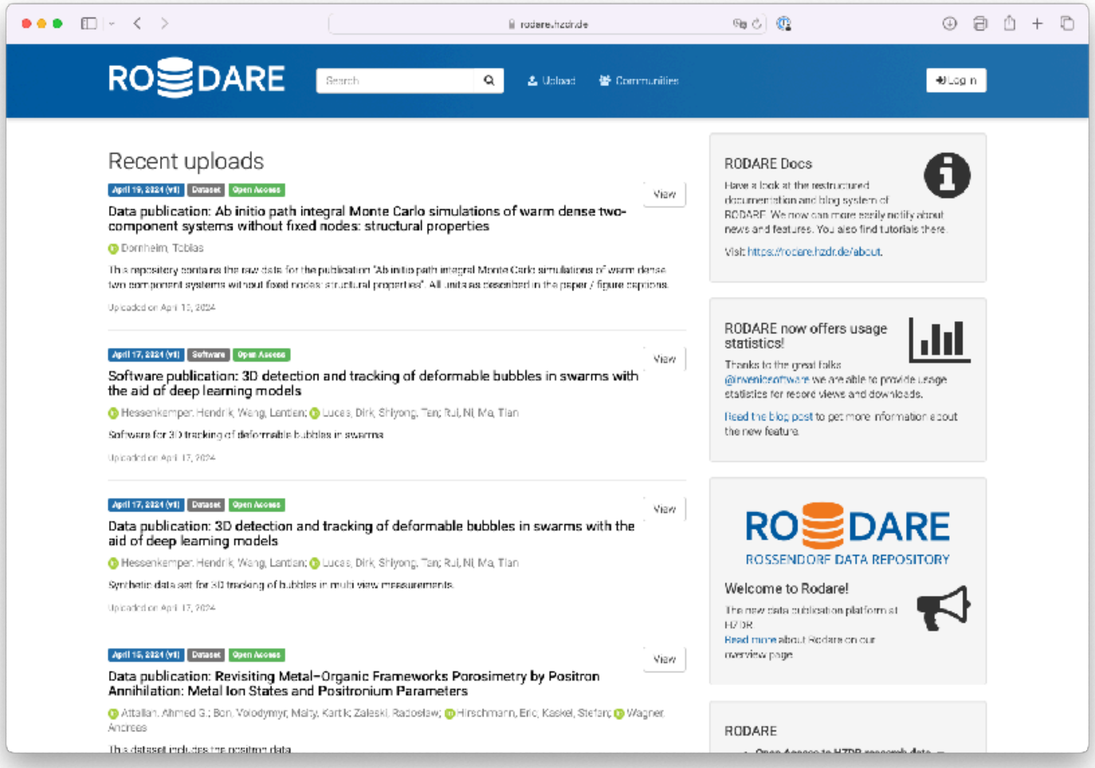


CERN
NII
NORTHWESTERN UNIVERSITY
TRECCANI LA CULTURA ITALIANA
Brookhaven National Laboratory
Caltech Library
TIND
HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF
WWU MÜNSTER
data futures
OpenAIRE
Universität Hamburg DER FORSCHUNG | DER LEHRE | DER BILDUNG
eKonnect
GEO GROUP ON EARTH OBSERVATIONS
INFN
TÜBİTAK ULAKBİM
DRESDEN concept
HZDR

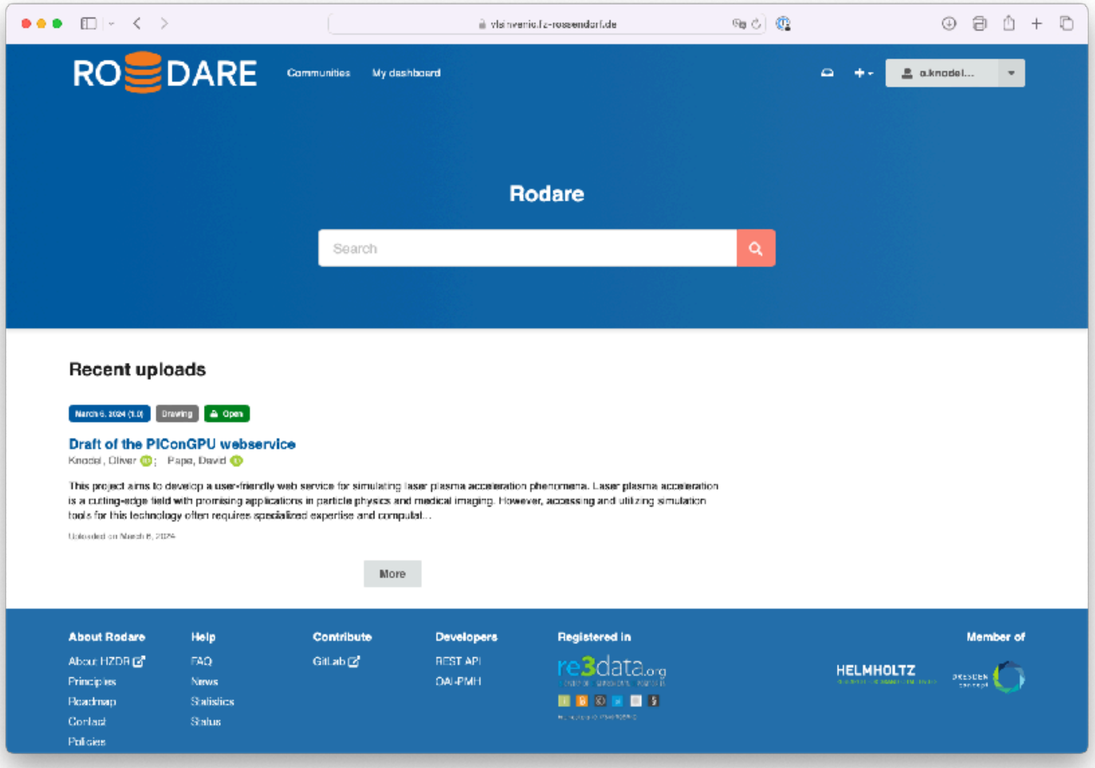
Transfer RODARE to RODARE(RDM)

- We started the customisation of InvenioRDM to create a new RODARE (RDM)
- At this point, we decided to change the design… but that will take a while…

INVENIO



INVENIO RDM



DRESDEN concept HZDR

First Steps with InvenioRDM

1. Install CLI tool

— Install the InvenioRDM CLI tool, e.g. via pip:

```
pip install invenio-cli
```

2. Check system requirements

— You can check if the proper requirements are installed via invenio-cli:

```
invenio-cli check-requirements
```

3. Scaffold project

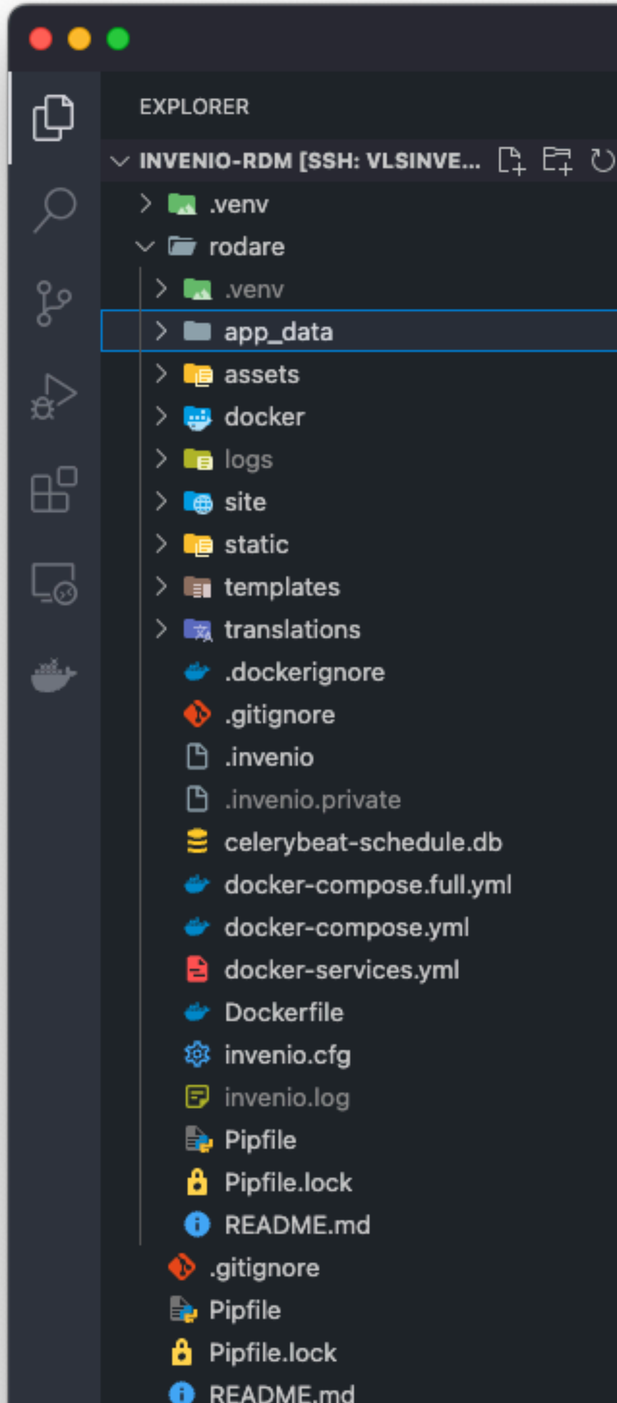
— Scaffold your InvenioRDM instance. Replace <version> with the version you want:

- LTS release (for production systems): v9.1
- STS release (for feature previews): v11.0

```
invenio-cli init rdm -c <version>
```

```
# e.g:
invenio-cli init rdm -c v9.1
```

You will be asked several questions. If in doubt, choose the default.



Deployment Options: Container or Local

4. Build, setup and run

— Now that the scaffolding is complete, it is time to check the development requirements:

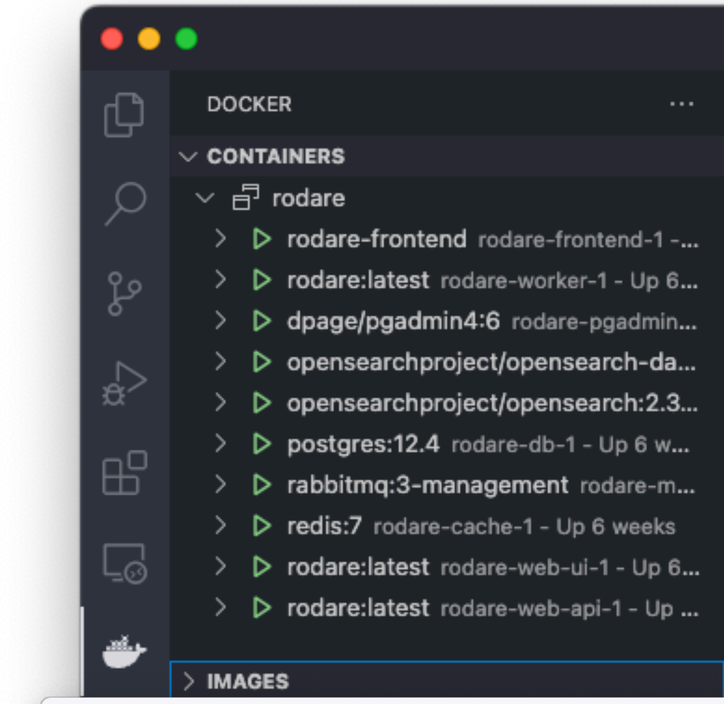
```
cd my-site/
invenio-cli check-requirements –development
```

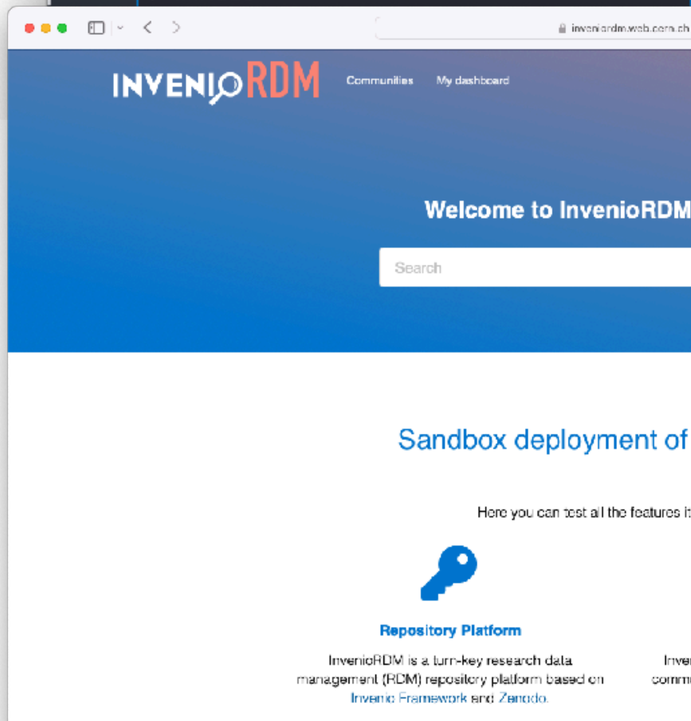
— YouYou can run the main InvenioRDM application in two modes (choose one):

- Containerized application and services (good for a quick preview).
  `invenio-cli containers start --lock --build –setup`
- Local application with containerized services (good for developers or if you want to customize InvenioRDM).

  ```
  invenio-cli install
  invenio-cli services setup
  invenio-cli run
  ```

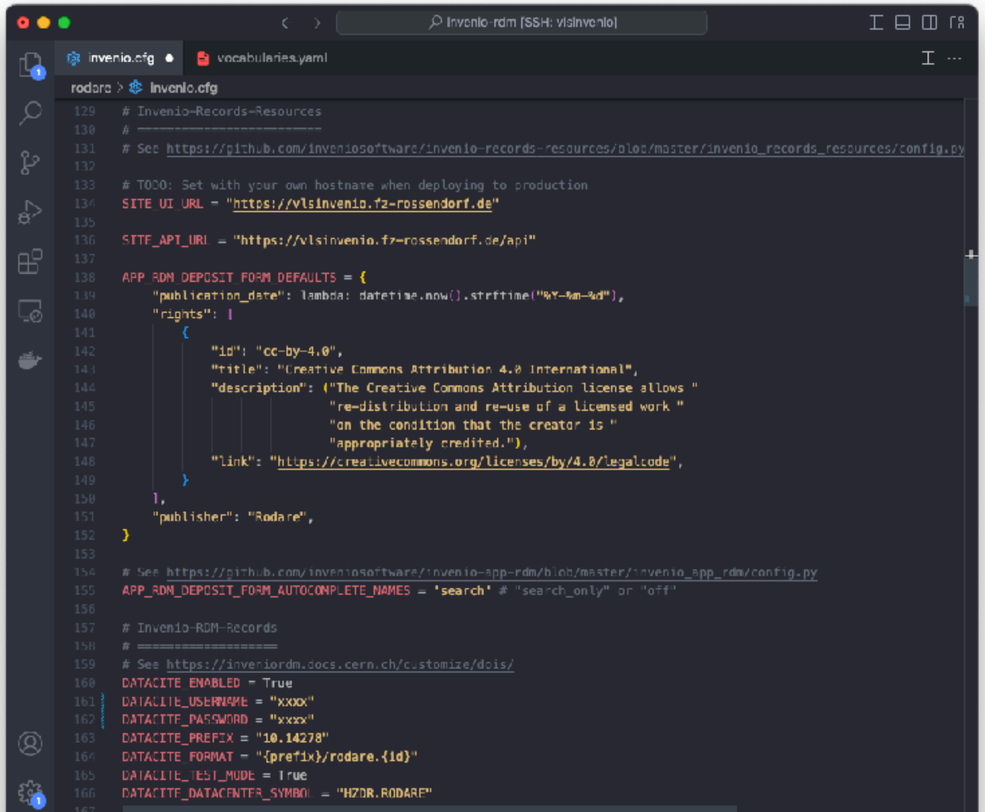
5. Explore InvenioRDM¶

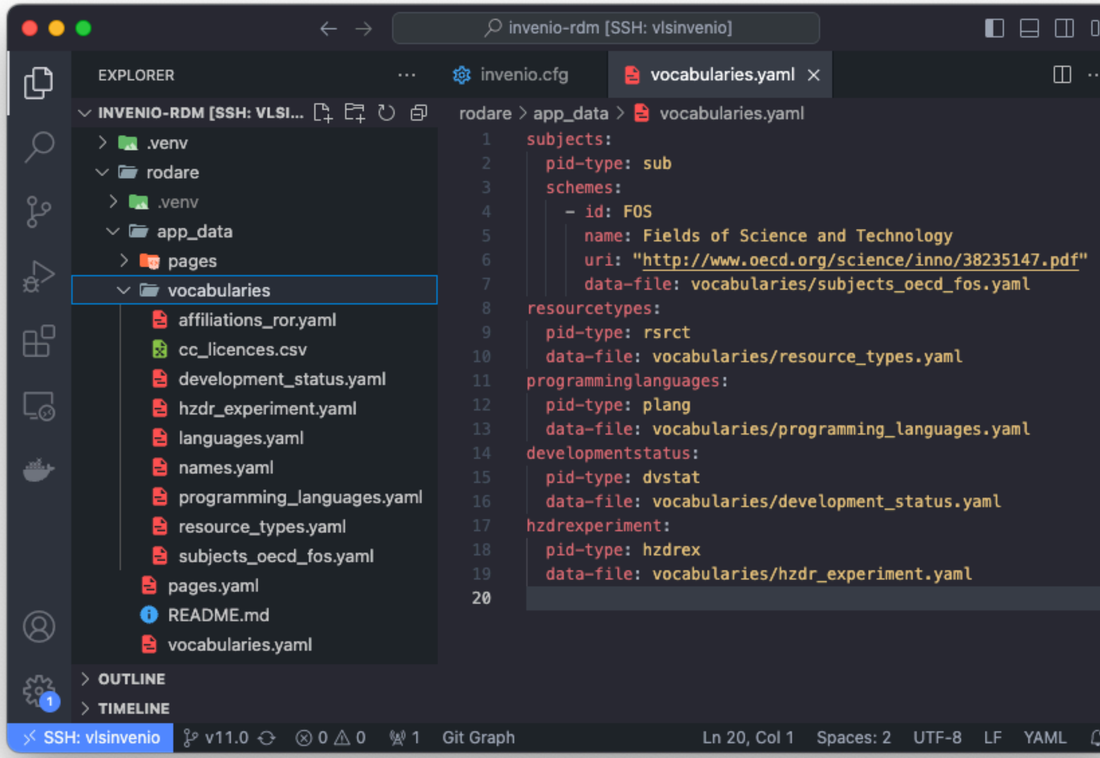




Configuration and Customisation

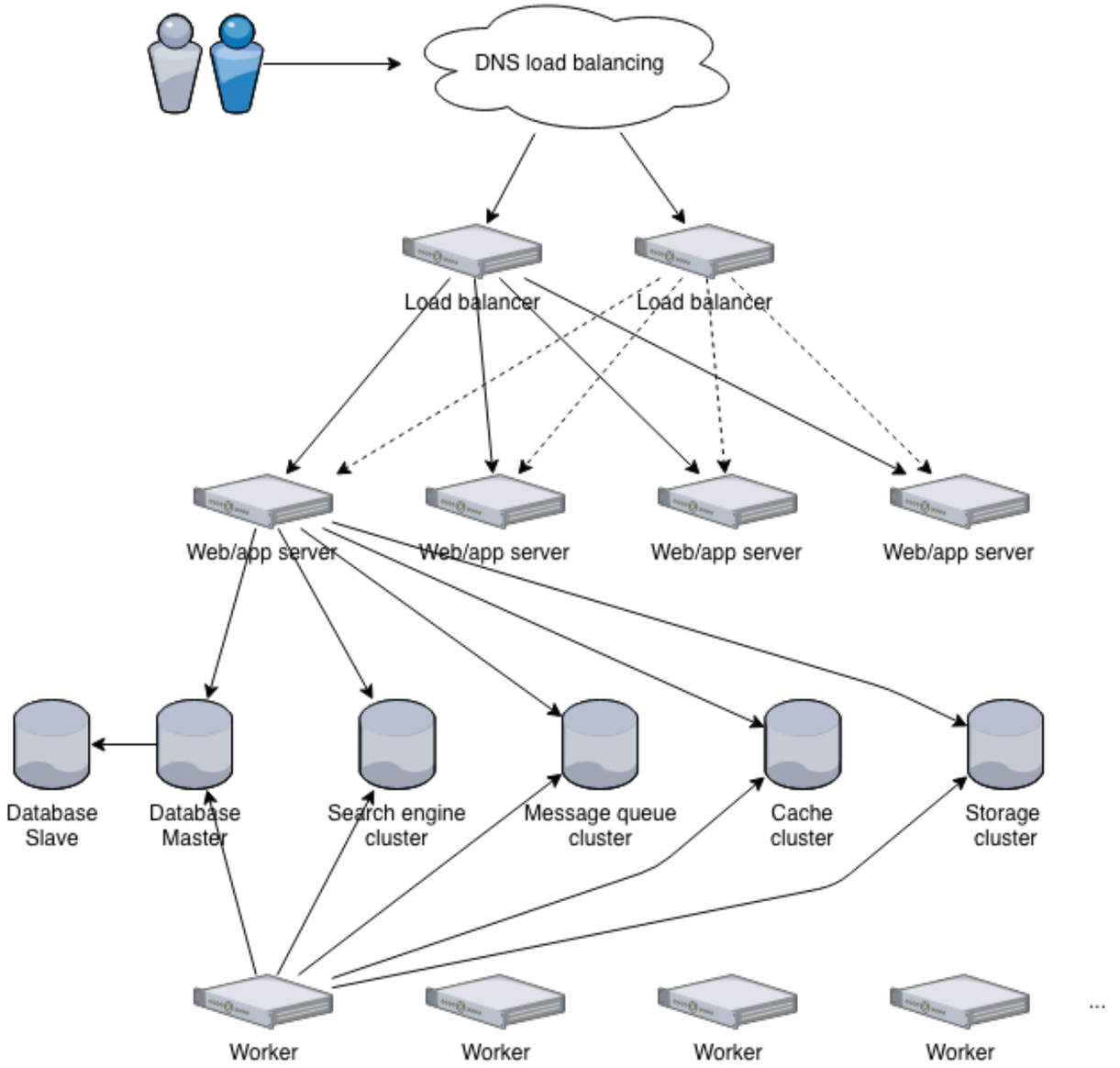
- Configuration via **`invenio.cfg`**: e.g. DataCite, OpenID, Shibboleth, OAI-PMH, storage, custom metadata, communities, …
- Customisation of record landing pages and vocabularies
- Overall design and Look-and-Feel



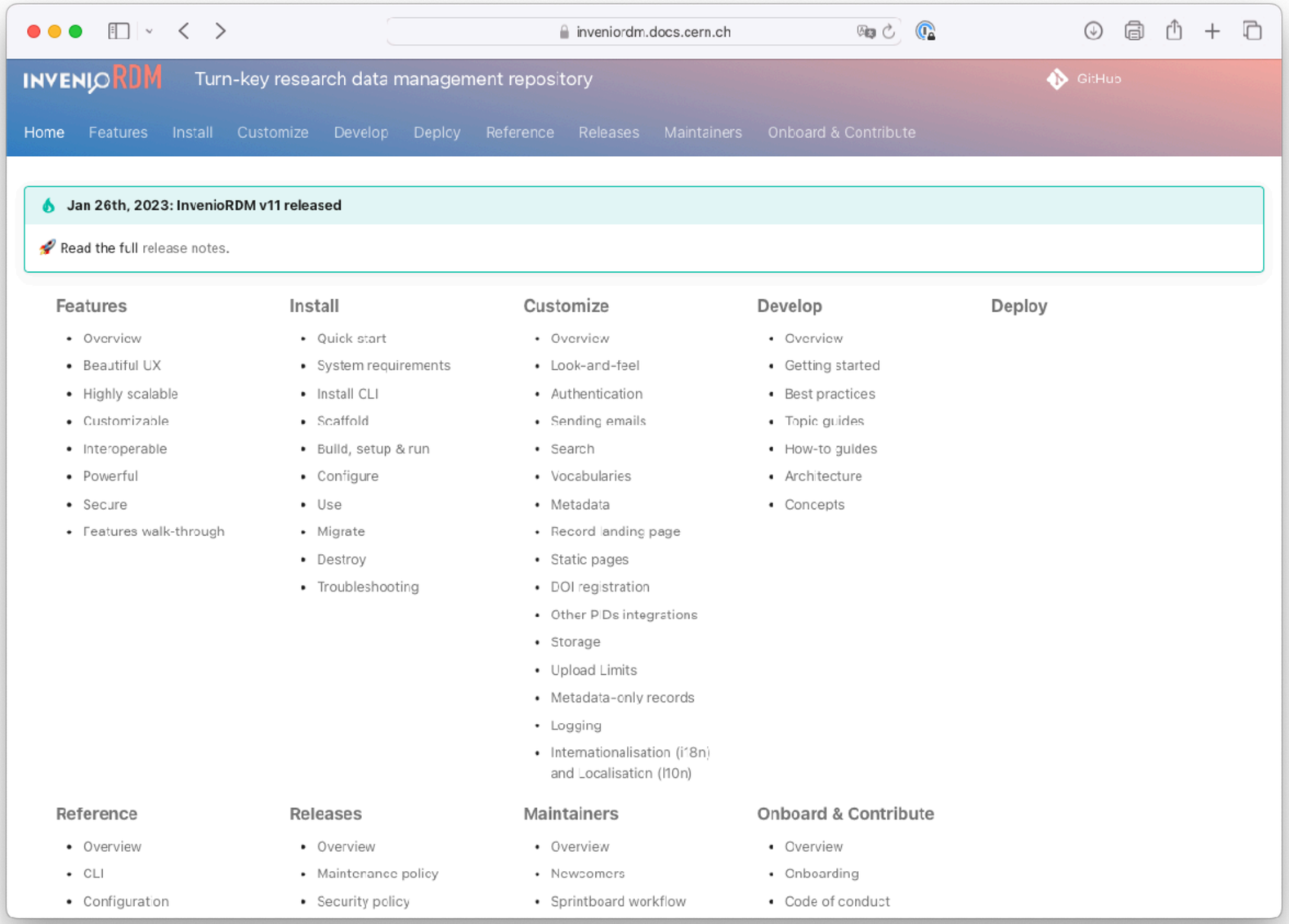


Infrastructure

- **Web servers**: The load balancer proxies traffic to one of several web servers
- **Application servers**: The web server proxies traffic usually (but not necessarily) to a single application server running on the same machine. Invenio is a Python application, and thus make use of the WSGI standard
- **Storing records**: Invenio store records as JSON documents in an SQL database
- **Search and indexing**: Invenio uses OpenSearch as its underlying search engine since OpenSearch is fully JSON-based
- **Background processing**: Invenio relies on an application called Celery for distributed background processing
- **Caching and temporary storage**: Invenio uses an in-memory cache like Redis or Memcache
- **Storing files**: Invenio comes with a default object storage REST API to expose files. Underneath the hood, Invenio can however store files in multiple different storage systems due to a simple storage abstraction layer.



Community Website with Documentation



inveniordm.docs.cern.ch

Additional Community-Specific Metadata

- With RODARE we have one data repository for all scientific fields at HZDR.
- RODARE provides only bibliographic metadata to describe and search within the data records.
- More detailed scientific metadata can help scientists in their daily research.
- Our goal:

Ingest

Automate data importing

Embedding the detector and experiment infrastructure in automated metadata management pipelines

Annotate

Enrich your data

Add rich fully structured, hierarchical metadata to your data to give them more meaning

Find

Browse and search your data

Organise metadata and make the data findable and searchable via rich metadata

Publish

Add DOI and share

Publish data and make it citable via a digital object identifier, thus linking metadata and data to publication

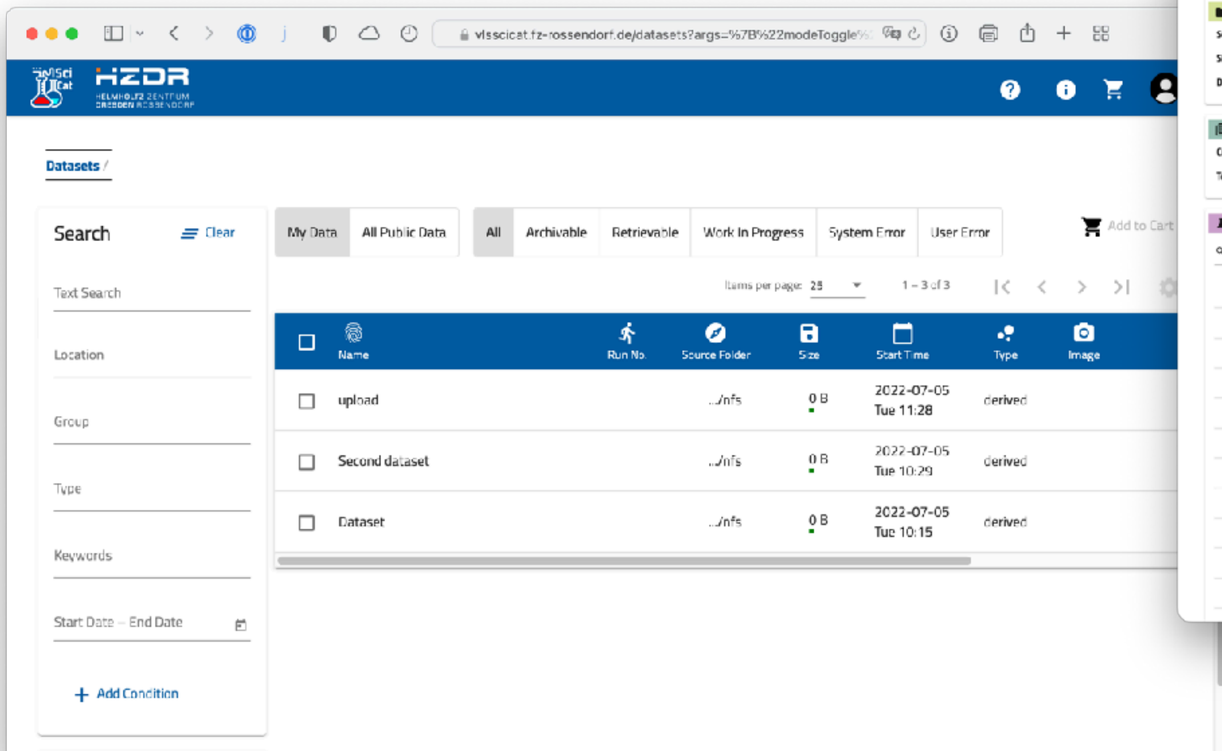
DRESDEN concept HZDR

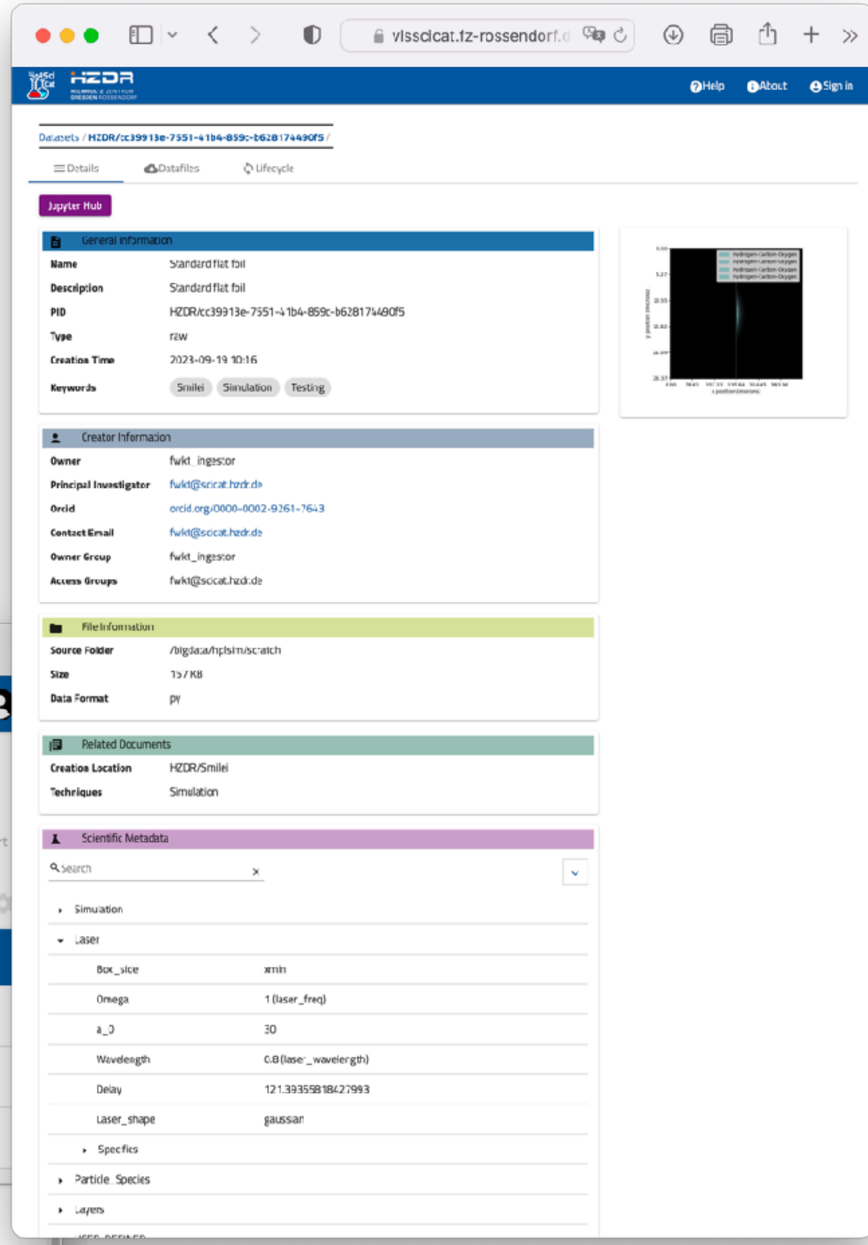
Metadata Catalogue SciCat

— An additional system (SciCat) for community-specific metadata is required.

— Every SciCat entry points to a RODARE entry (metadata only records are possible in one of the next releases).

— Nothing changes for the publication management system ROBIS.

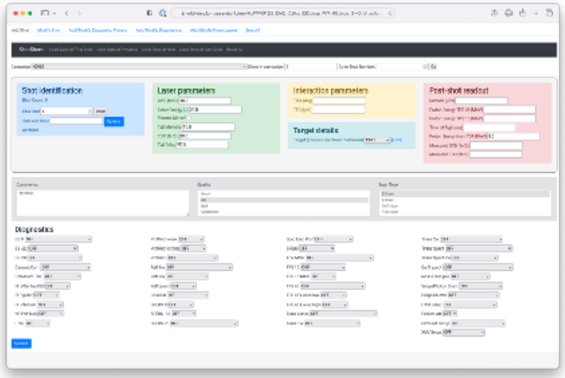




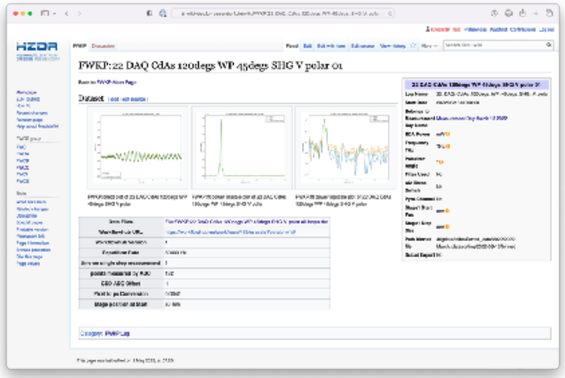
Metadata Catalogue SciCat and Data Repository RODARE (Draft)

Curated Metadata Source

ExperimentLogging app (ExL)



E-Logbook

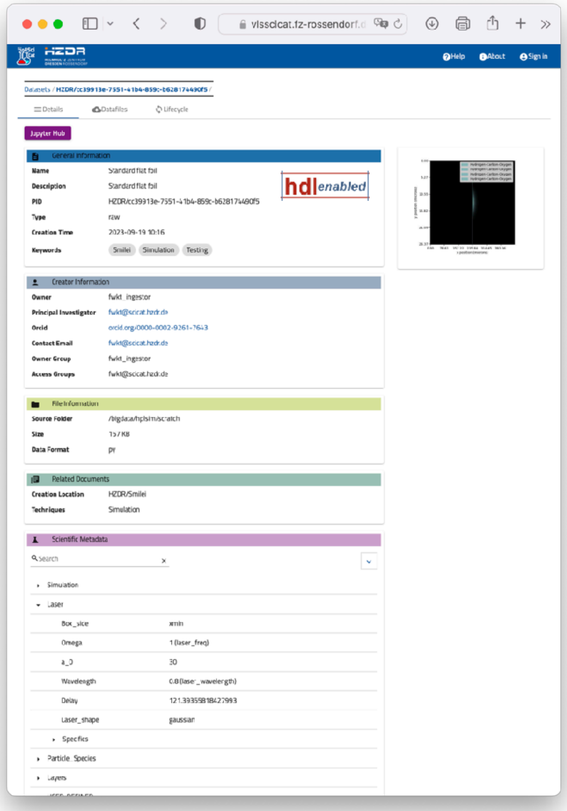


Metadata from Experiment/Simulation

Fully Automated Process

Public Metadata Catalogue

SciCat



Data Access

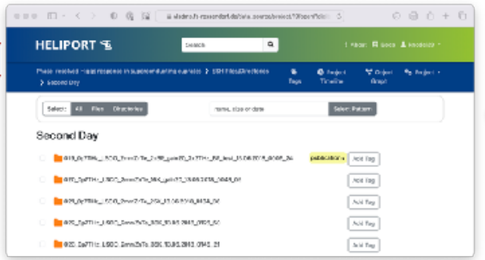
RODARE

doi

Dataset

HELIPORT

hdlenabled



Filesystem

Tape Archive