



# InvenioRDM at HZDR: RO DARE

ROSSENDORF DATA REPOSITORY

DMA ST1 seminar, April 2024

Oliver Knodel // contact: [o.knodel@hzdr.de](mailto:o.knodel@hzdr.de)





# 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.

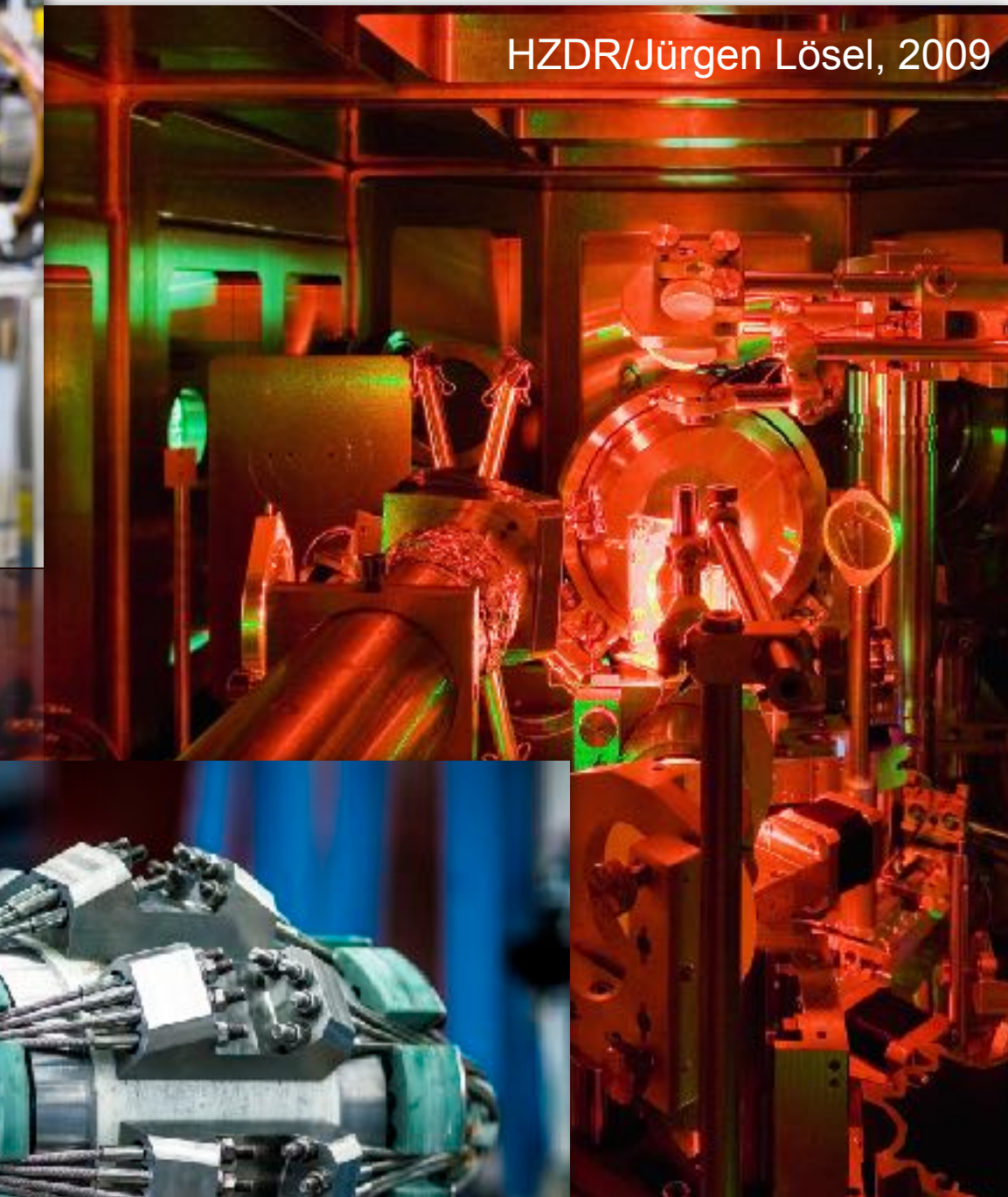
— Positrons, protons, neutrons as well as X-ray and gamma radiation.

## Dresden High Magnetic Field Laboratory (HLD)

— Europe's highest pulsed magnetic fields.

## Ion Beam Center (IBC)

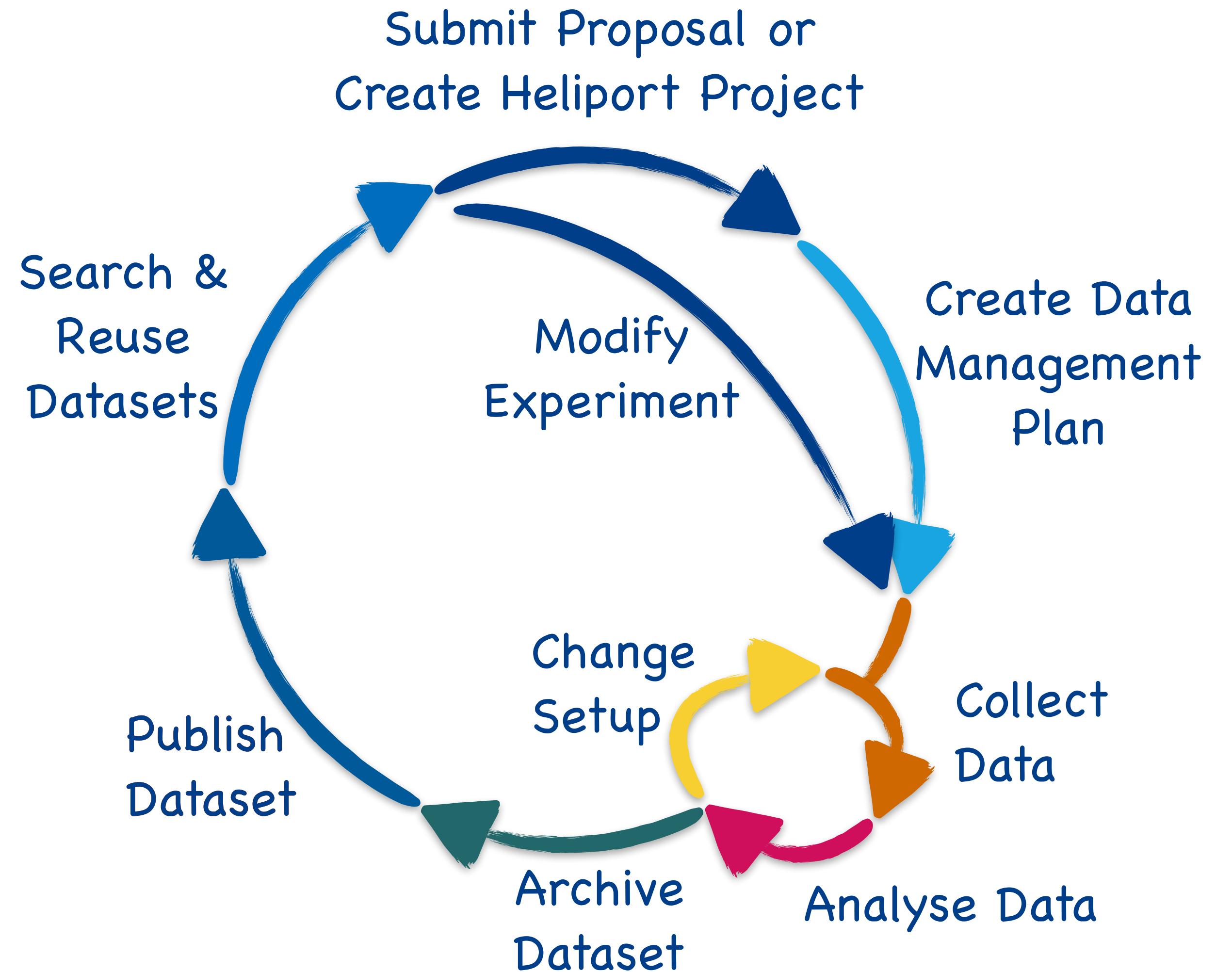
— Nanoscale surface analysis and modification.





# 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:



## The HZDR Data Policy



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

**Table of Contents**

	Page
Cover Sheet	1
Preamble	2
1 Definitions	2
2 General Principles	3
3 Research Data Management	3
4 Raw Data and associated Metadata	3
5 Result Data	4
6 Legal Requirements	4
7 Taking Effect	5

**List of Annexes**

Appendix	Content
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**

Abbreviation	Full Name
CC BY	Creative Commons Attribution 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	Rosendorf research Data Repository



# 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

The screenshot shows the title page of the HZDR Data Policy document. The title is "HZDR Data Policy" with the subtitle "Terms and Conditions for the Storage, Access and Curation of Research Data". The document is identified by DOI 10.14278/roda.re.2269. The table of contents lists sections from Cover Sheet to Taking Effect. The list of annexes includes a checklist for a Data Management Plan and a Data Cite Metadata Schema. The list of revisions shows a single revision on 01.05.2018. The list of abbreviations includes Creative Commons licenses, Data Management Plan, Digital Object Identifier, and the RODARE repository.

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

Page	Page
Cover Sheet.....	1
Preamble .....	2
1 Definitions .....	2
2 General Principles .....	3
3 Research Data Management.....	3
4 Raw Data and associated Metadata .....	3
5 Result Data.....	4
6 Legal Requirements .....	4
7 Taking Effect.....	5

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

Abbreviation	Full Name
	Creative Commons Attributive License
	Creative Commons Universal License
	Data Management Plan
	Digital Object Identifier
data	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

The screenshot shows the RODARE website interface. At the top, there is a navigation bar with the RODARE logo, a search bar, and buttons for 'Upload' and 'Communities'. The main content area is divided into several sections:

- Recent uploads:** A list of three recent uploads, each with a 'View' button.
  - March 10, 2021 (v1) Dataset Open Access:** Research data 'Fluorination of graphene leads to susceptibility for nanopore formation by highly charged ion impact'. Authors: Creutzburg, Sascha; Hübner, René; Facsko, Stefan. Description: The depository contains STM images, experimental data from charge exchange measurements and data from charge exchange simulations. Uploaded on March 10, 2021.
  - March 9, 2021 (v1) Software Open Access:** PIconGPU setup: LPWFA downramp injection. Authors: Pausch, Richard; Couperus Cabadag, Jurjen Pieter; Bastrakov, Sergei; Busmann, Michael; Irman, Arie; Kurz, Thomas; Schöbel, Susanne; Schramm, Ulrich; Steiniger, Klaus; Ufer, Patrick; Widera, René; Debus, Alexander. Description: PIconGPU source code and setup files used for the LPWFA downramp injection simulation study. Uploaded on March 9, 2021.
  - March 2, 2021 (v1) Dataset Open Access:** Data for: Experimental studies on bubble aspect ratio and corresponding correlations under bubble swarm condition. Authors: Liu, Liu; Zhang, Heyang; Yan, Hongjie; Ziegenhein, Thomas; Heßenkemper, Hendrik; Zhou, Ping; Lucas, Dirk. Description: Zip-file that contains the raw images on a study on bubble aspect ratio under swarm condition. Further information can be found in the respective paper. Uploaded on March 3, 2021.
- RODARE Docs:** A section with an information icon and text: 'Have a look at the restructured documentation and blog system of RODARE. We now can more easily notify about news and features. You also find tutorials there. Visit: <https://rodare.hzdr.de/about>.'
- RODARE now offers usage statistics!** A section with a bar chart icon and text: 'Thanks to the great folks @inveniosoftware we are able to provide usage statistics for record views and downloads. Read the blog post to get more information about the new feature.'
- RODARE ROSSENDORF DATA REPOSITORY:** A section with the RODARE logo and text: 'Welcome to Rodare! The new data publication platform at HZDR. Read more about Rodare on our overview page.'

<https://rodare.hzdr.de>

Powered by:



Metadata:



DataCite  
FIND, ACCESS, AND REUSE DATA

Registered in:

re3data.org  
REGISTRY OF RESEARCH DATA REPOSITORIES



<http://doi.org/10.17616/R3BR40>



# Different Upload Types and Previews



February 2, 2021 Software Open Access

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

Koller, Fabian; Foeschel, Franz; Gu, Junmin; Huebl, Axel

Other(s): Fortmann-Grote, Carsten; Stańczak, Dominik; Amundson, James; Donnelly, Ray; Widera, René; Zenker, Eric; Rastrakov, Sergei; Liche, Rémi; Amonim, Ifigia Diana; Rastrakova, Ksenia; Pausch, Richard; Ordyna, Paweł

openPMD is an open metadata format for open data workflows in open science. This library provides a common high-level API for openPMD writing and reading. It provides a common interface to I/O libraries and file formats such as HDF5 and ADIOS. Where supported, openPMD api implements both serial and MPI parallel I/O capabilities.

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 Particle Accelerators (CAMPx), 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-0.13.2.zip

- openPMD-openPMD-api-7a7468
  - .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 794 Bytes
      - install\_problem.md 794 Bytes
      - question.md 610 Bytes
    - ci
      - sanitizer
        - clang 374 Bytes
        - Laack.supp
      - spark
        - compilers.yaml 4.2 kB
        - config.yaml 24 Bytes
        - packages.yaml 6.9 kB

3,961 views 211 downloads [See more details...](#)

Publication date: February 2, 2021

DOI: [10.14278/rodare.798](https://doi.org/10.14278/rodare.798)

Keyword(s): openPMD, Open Science, Open Data, HDF5, ADIOS, data, MPI, HPC, research, file-format, file-handling

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

Related identifiers: Cited by: 10.5281/zenodo.1167843, 10.5281/zenodo.1069484, 10.5281/zenodo.83594

Communities: OpenAIR, RODARE

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

Versions

Version 0.13.2	Feb 2, 2021
Version 0.13.1	Jan 9, 2021
Version 0.13.0	Jan 3, 2021
Version 0.12.0-alpha	Sep 8, 2020

September 8, 2020 Photo Open Access

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

Podlipec, Rok

Contact person(s): Gregor Hlawacek; Nico Klingner

Work package leader(s): Rok Podlipec

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

78 views 51 downloads [See more details...](#)

Publication date: September 8, 2020

DOI: [10.14278/rodare.514](https://doi.org/10.14278/rodare.514)

Related identifiers: Identical to: <https://www.hzdr.de/publications/PubI-31505>, Referenced by: <https://www.hzdr.de/publications/PubI-31504>

Communities: RODARE

License (for files): Creative Commons Attribution 4.0 International

Versions

Version 1	Sep 8, 2020
10.14278/rodare.514	

Cite all versions? You can cite all versions by using the DOI 10.14278/rodare.514. This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

Share

Cite as

Podlipec, 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>

Start typing a citation style...

Files (8.6 MB)

Name	Size	Preview	Download
e08_s02_t02_F10_I_A4_CellMask_Tin2_Alexa_longTermExp_10X000.tif	655.7 kB		
md5:2ae0c087d13bec04a544b2534d09301			
e08_s02_t02_F10_I_A4_CellMask_Tin2-Alexa_longTermExp_50X000.tif	358.1 kB		
md5:73bef27205e3cfc232331b96921e9b4			



# Search and Filter on Metadata (DataCite)

The screenshot shows the RODARE search results page. The header includes the RODARE logo, a search bar, and navigation links for Upload, Communities, and Log In. The main content area displays a list of search results, each with a date, version, access type, title, author(s), and a 'View' button. The left sidebar contains filters for Access Right (Open, Restricted, Closed), File Type (Zip, Pdf, Xlsx, etc.), and Keywords (Flow, Tomography, Ray, etc.).

Found 708 results. Sort by: Most recent

Access Right

- Open (465)
- Restricted (148)
- Closed (95)

File Type

- Zip (242)
- Pdf (48)
- Xlsx (44)
- Txt (25)
- Csv (16)
- 7z (15)
- Tif (15)
- Png (14)
- Xz (14)
- Opj (12)

Keywords

- Flow (52)
- Tomography (49)
- Ray (44)
- Data (41)
- Electron (39)
- X (39)
- Phase (36)
- Computer (37)
- Beam (33)
- Two (30)

Results:

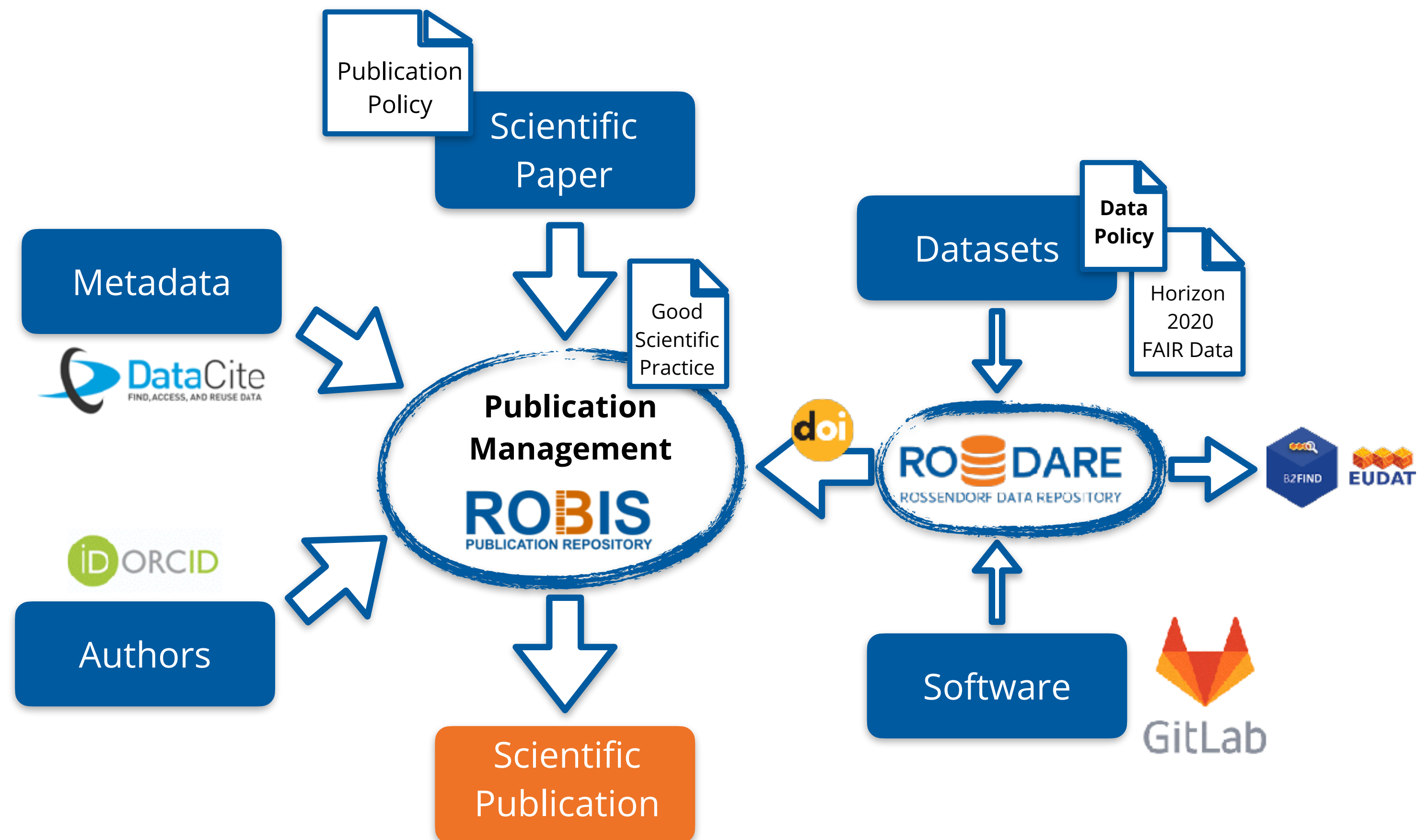
- April 19, 2024 (v1) Dataset Open Access  
**Data publication: Ab initio path integral Monte Carlo simulations of warm dense two-component systems without fixed nodes: structural properties**  
Dornheim, Tobias  
This repository contains the raw data for the publication "Ab initio path integral Monte Carlo simulations of warm dense two-component systems without fixed nodes: structural properties". All Units as described in the paper / figure captions.  
Uploaded on April 19, 2024
- April 17, 2024 (v1) Dataset Restricted Access  
**Electrical characterization of multi-gated WSe2/MoS2 van der Waals heterojunctions**  
Chava, Phanish, Kateel, Vaishnavi, Watanabe, Kenji, Tanihushi, Takashi, Helm, Manfred, Mikolajek, Thomas, Erbe, Artur  
Transistor measurements  
Uploaded on April 18, 2024
- April 17, 2024 (v1) Software Open Access  
**Software publication: 3D detection and tracking of deformable bubbles in swarms with the aid of deep learning models**  
Hessenkemper, Hendrik, Wang, Lantian, Lucas, Dirk, Shiyong, Tan, Rui, Ni, Ma, Tian  
Software for 3D tracking of deformable bubbles in swarms  
Uploaded on April 17, 2024
- April 17, 2024 (v1) Dataset Open Access  
**Data publication: 3D detection and tracking of deformable bubbles in swarms with the aid of deep learning models**  
Hessenkemper, Hendrik, Wang, Lantian, Lucas, Dirk, Shiyong, Tan, Rui, Ni, Ma, Tian  
Synthetic data set for 3D tracking of bubbles in multi view measurements.  
Uploaded on April 17, 2024
- April 15, 2024 (v1) Dataset Open Access  
**Data publication: Revisiting Metal–Organic Frameworks Porosimetry by Positron Annihilation: Metal Ion States and Positronium Parameters**  
Attallah, Ahmed S., Bon, Volodymyr, Marly, Kartik, Zaleski, Radoslaw, Hirschmann, Eric, Kaskel, Stefan, Wagner, Andreas  
This dataset includes the experimental data...

```
{
  "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...

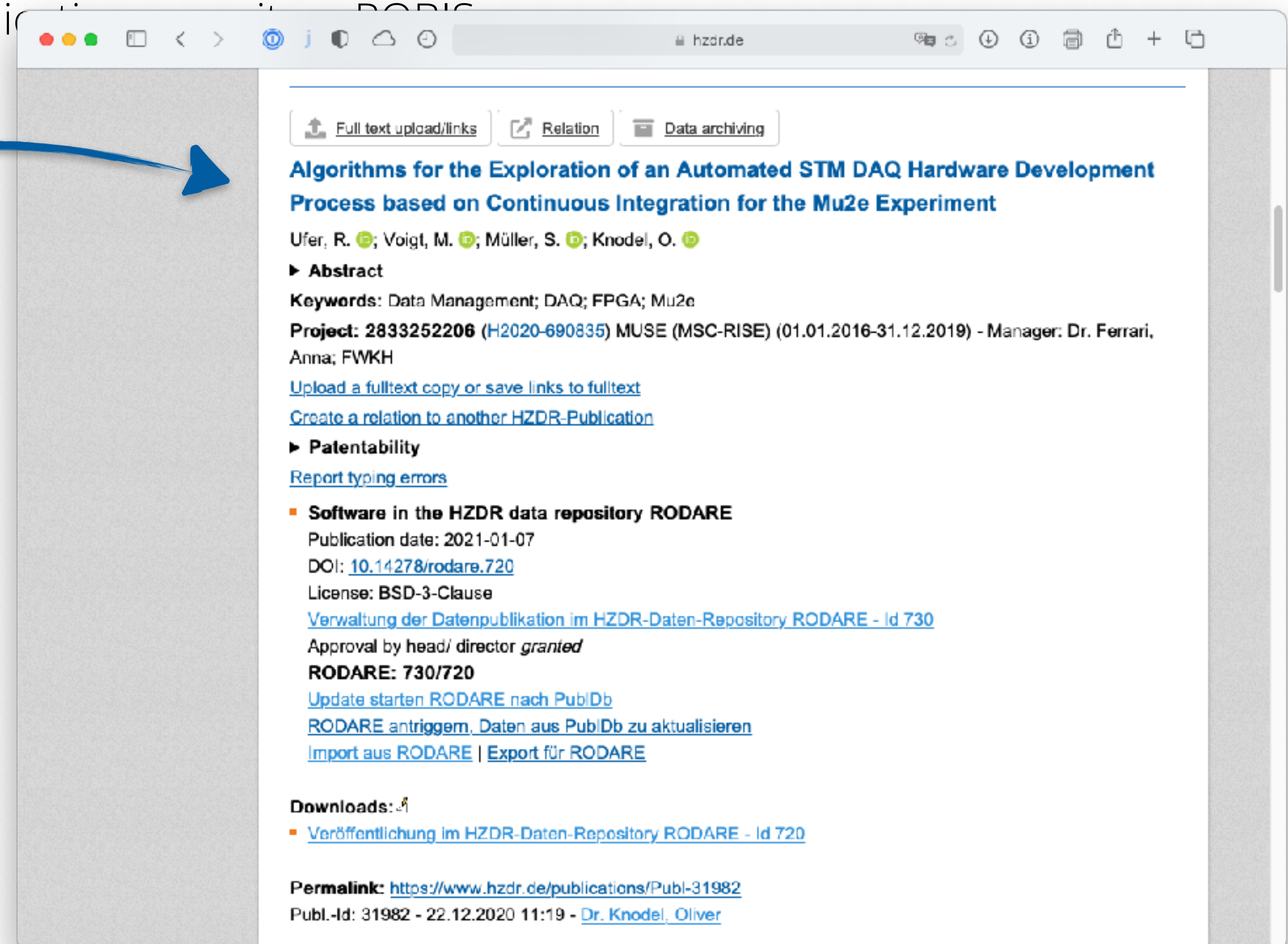




# Integrated in the HZDR Publication Repository





- HZDR publications must be registered in our overall publication repository ROBIS.
- Before publishing you will be redirected to ROBIS.

 Register in HZDR publication repository



Full text upload/links Relation Data archiving

**Algorithms for the Exploration of an Automated STM DAQ Hardware Development Process based on Continuous Integration for the Mu2e Experiment**

Ufer, R. ; Voigt, M. ; Müller, S. ; Knodel, O. 

► **Abstract**

**Keywords:** Data Management; DAQ; FPGA; Mu2e

**Project:** 2833252206 (H2020-690835) MUSE (MSC-RISE) (01.01.2016-31.12.2019) - Manager: Dr. Ferrari, Anna; FWKH

[Upload a fulltext copy or save links to fulltext](#)

[Create a relation to another HZDR-Publication](#)

► **Patentability**

[Report typing errors](#)

■ **Software in the HZDR data repository RODARE**

Publication date: 2021-01-07

DOI: [10.14278/rodare.720](https://doi.org/10.14278/rodare.720)

License: BSD-3-Clause

[Verwaltung der Datenpublikation im HZDR-Daten-Repository RODARE - Id 730](#)


Approval by head/ director *granted*

**RODARE: 730/720**

[Update starten RODARE nach PubIDb](#)

[RODARE antriggern, Daten aus PubIDb zu aktualisieren](#)

[Import aus RODARE](#) | [Export für RODARE](#)

**Downloads:** 

■ [Veröffentlichung im HZDR-Daten-Repository RODARE - Id 720](#)

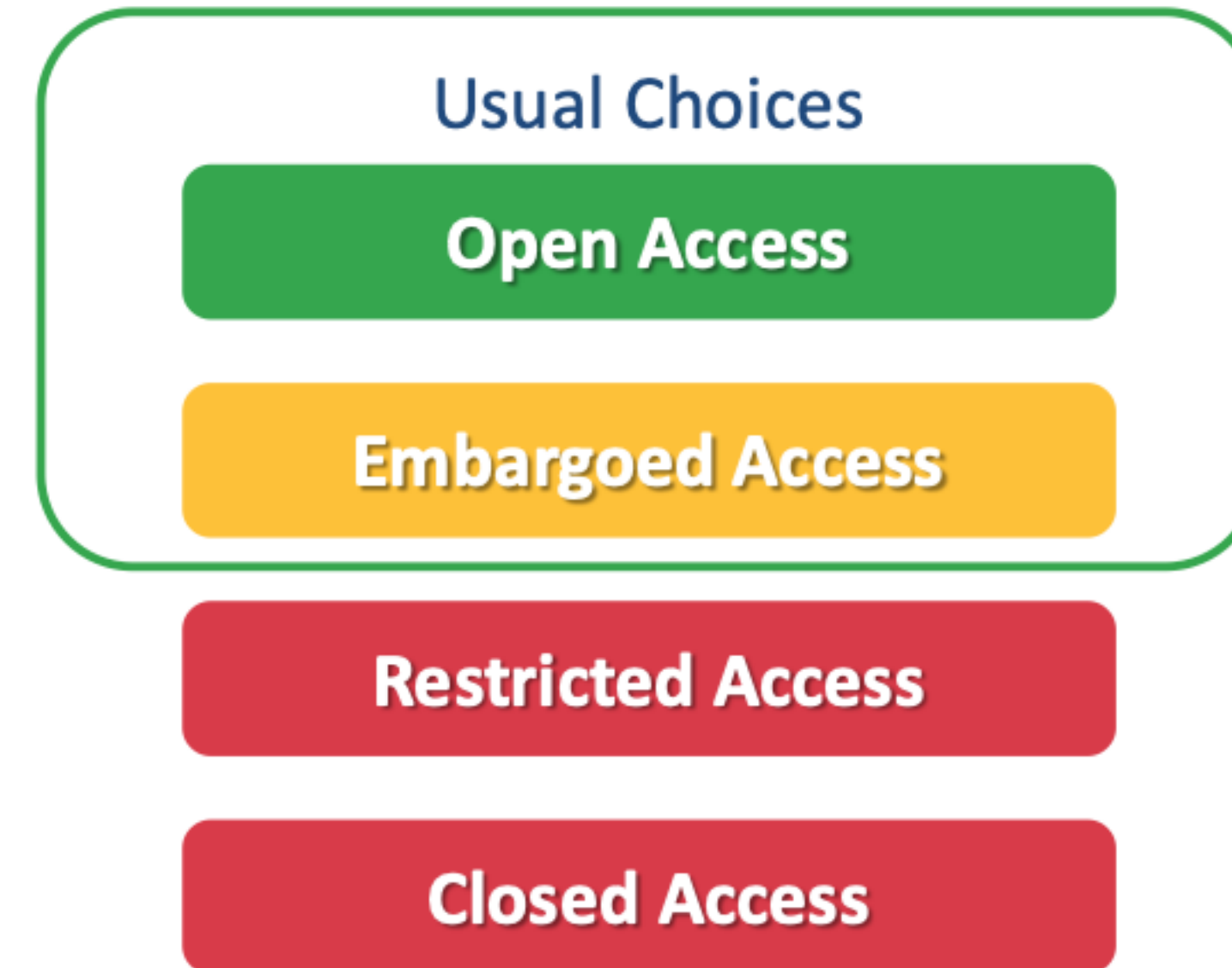
**Permalink:** <https://www.hzdr.de/publications/Publ-31982>

Publ.-Id: 31982 - 22.12.2020 11:19 - [Dr. Knodel, Oliver](#)



# 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.

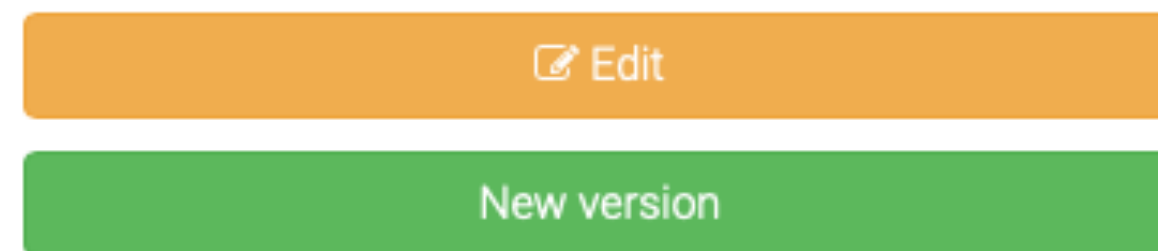


# DOI Versioning

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.

**Publication date:**  
January 7, 2021

**DOI:**  
DOI 10.14278/rodare.730

**Keyword(s):**  
Data Management DAQ FPGA Mu2e

**Related identifiers:**  
Identical to:  
<https://www.hzdr.de/publications/Publ-31982>

**Communities:**  
Helmholtz-Zentrum Dresden-Rossendorf  
RODARE

**License (for files):**  
[BSD 3-Clause "New" or "Revised" License](#)

**Versions**

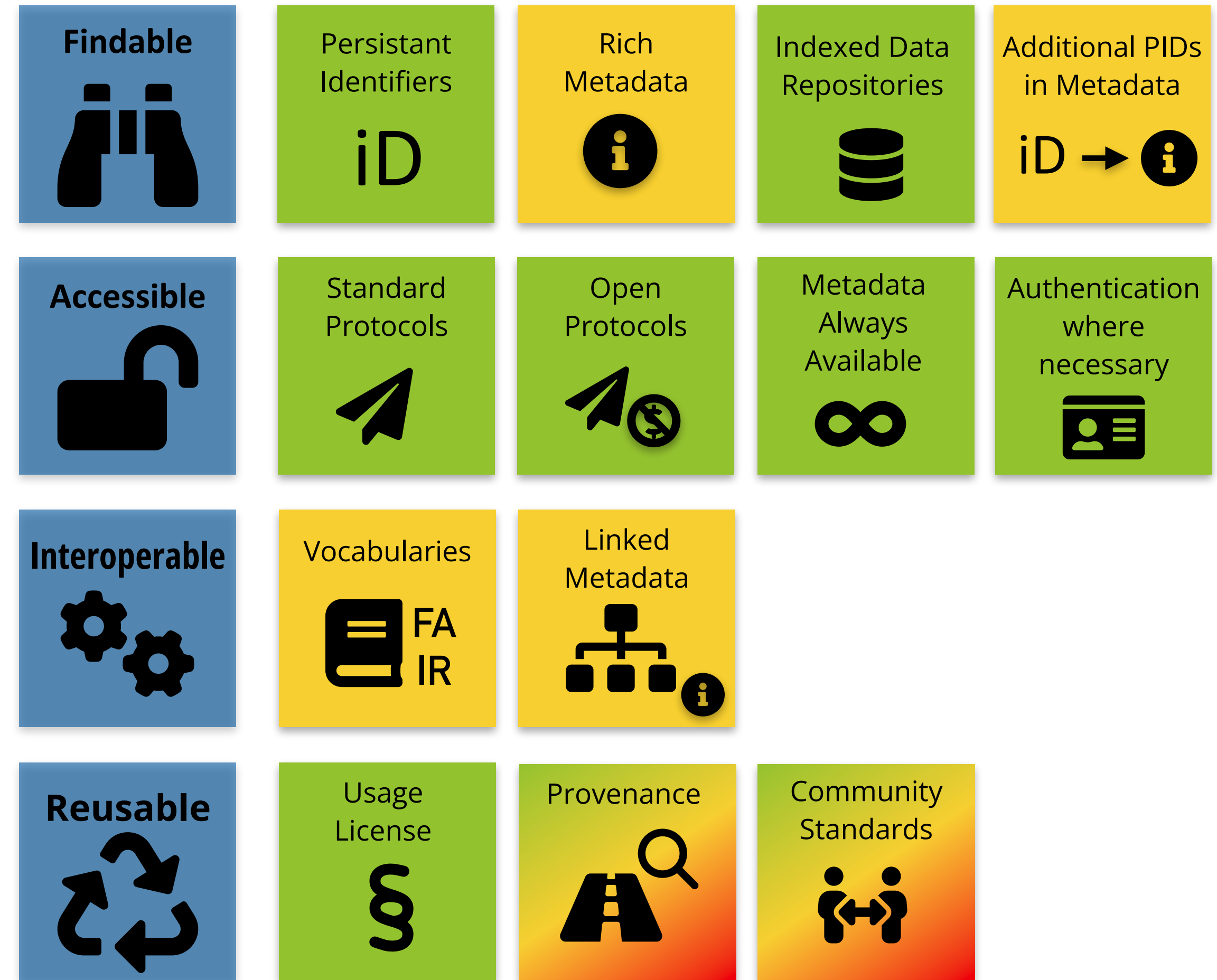
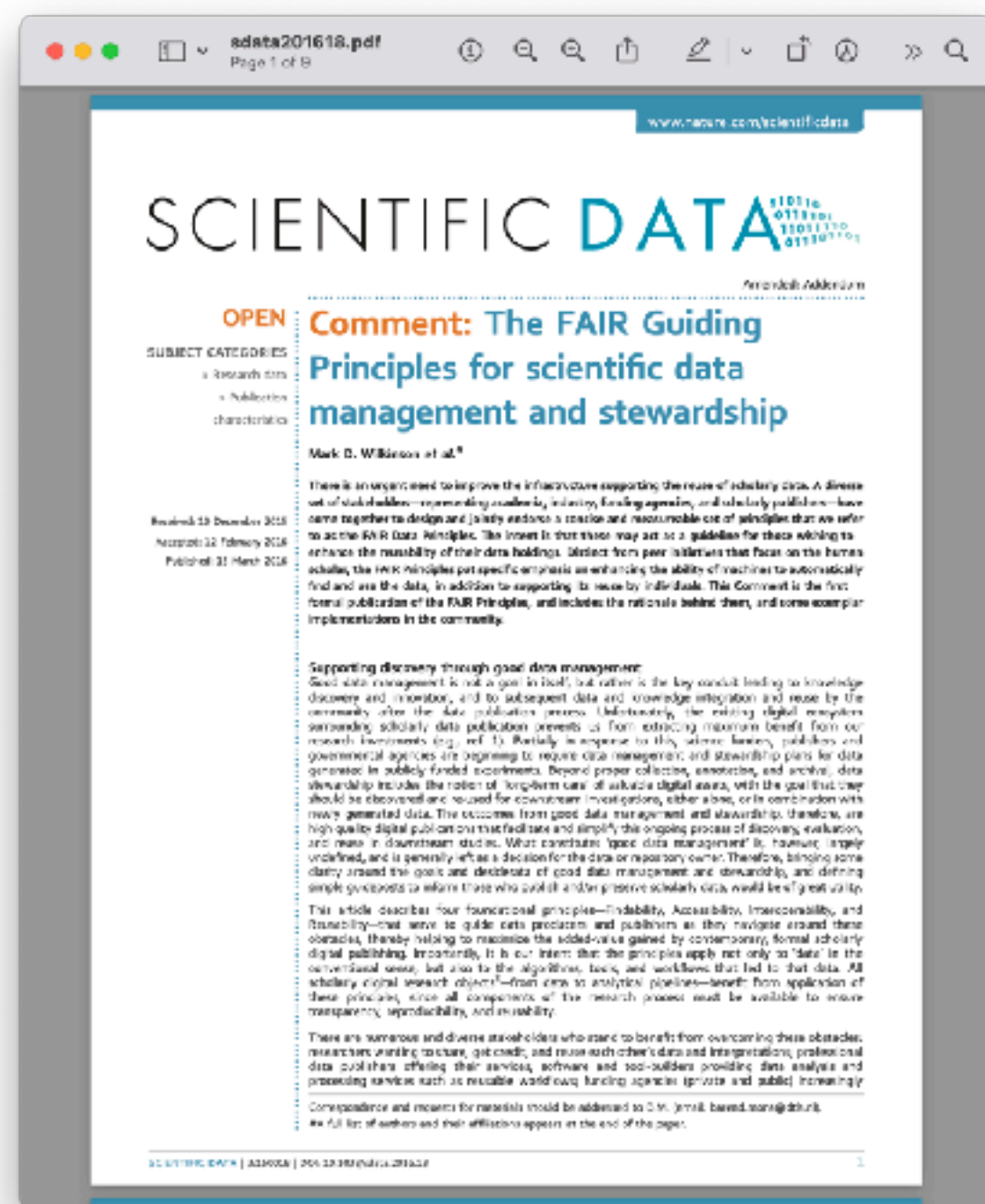
Version 0.1.0 10.14278/rodare.730	Jan 7, 2021
Version 0.1.0 10.14278/rodare.721	Jan 7, 2021

**Cite all versions?** You can cite all versions by using the DOI [10.14278/rodare.720](#). This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)



# 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.



■ Fully supported
 ■ Difficult
 ■ Depending on user/community



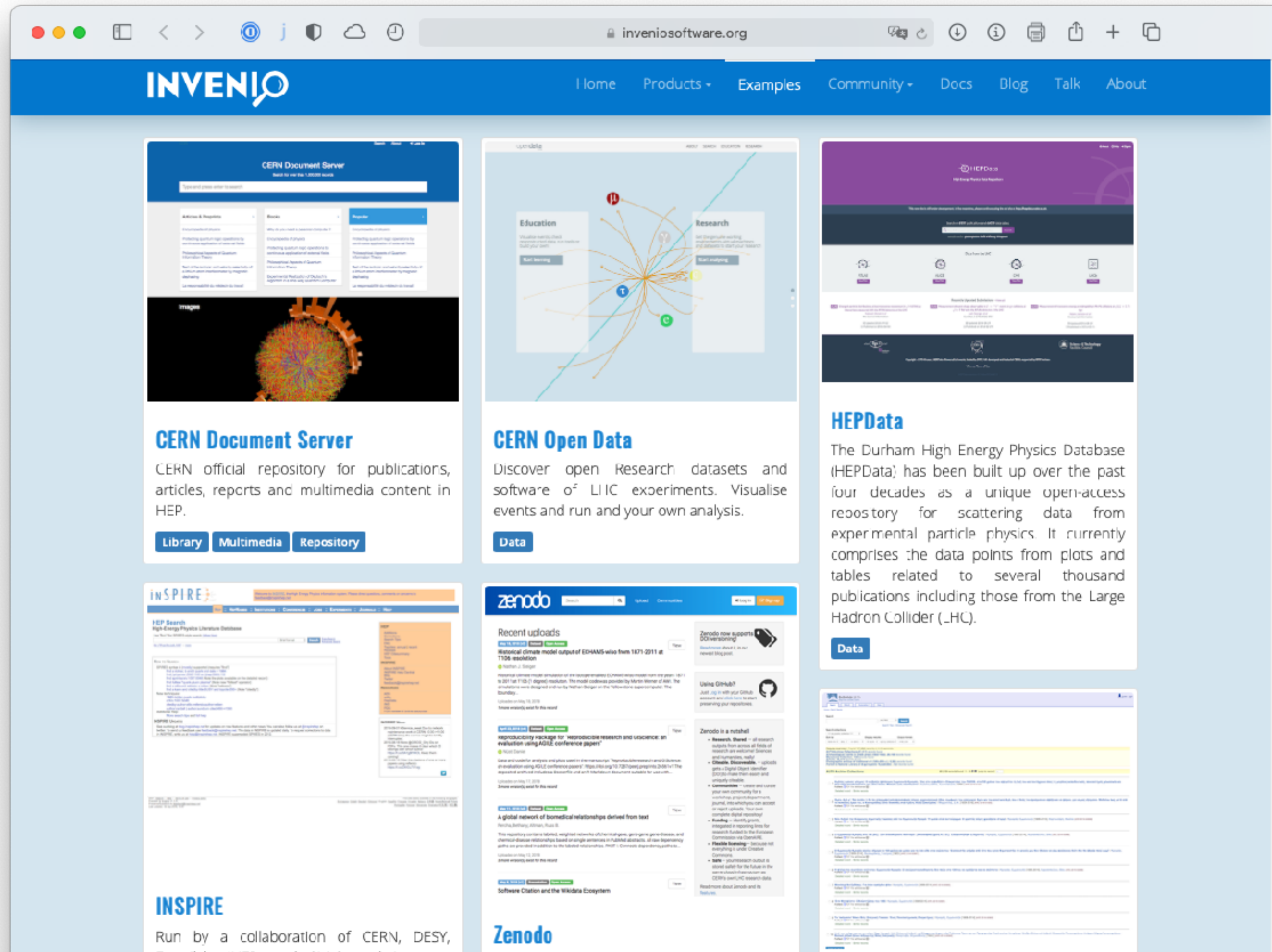
# Background: Invenio → Zenodo → RODARE

- 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



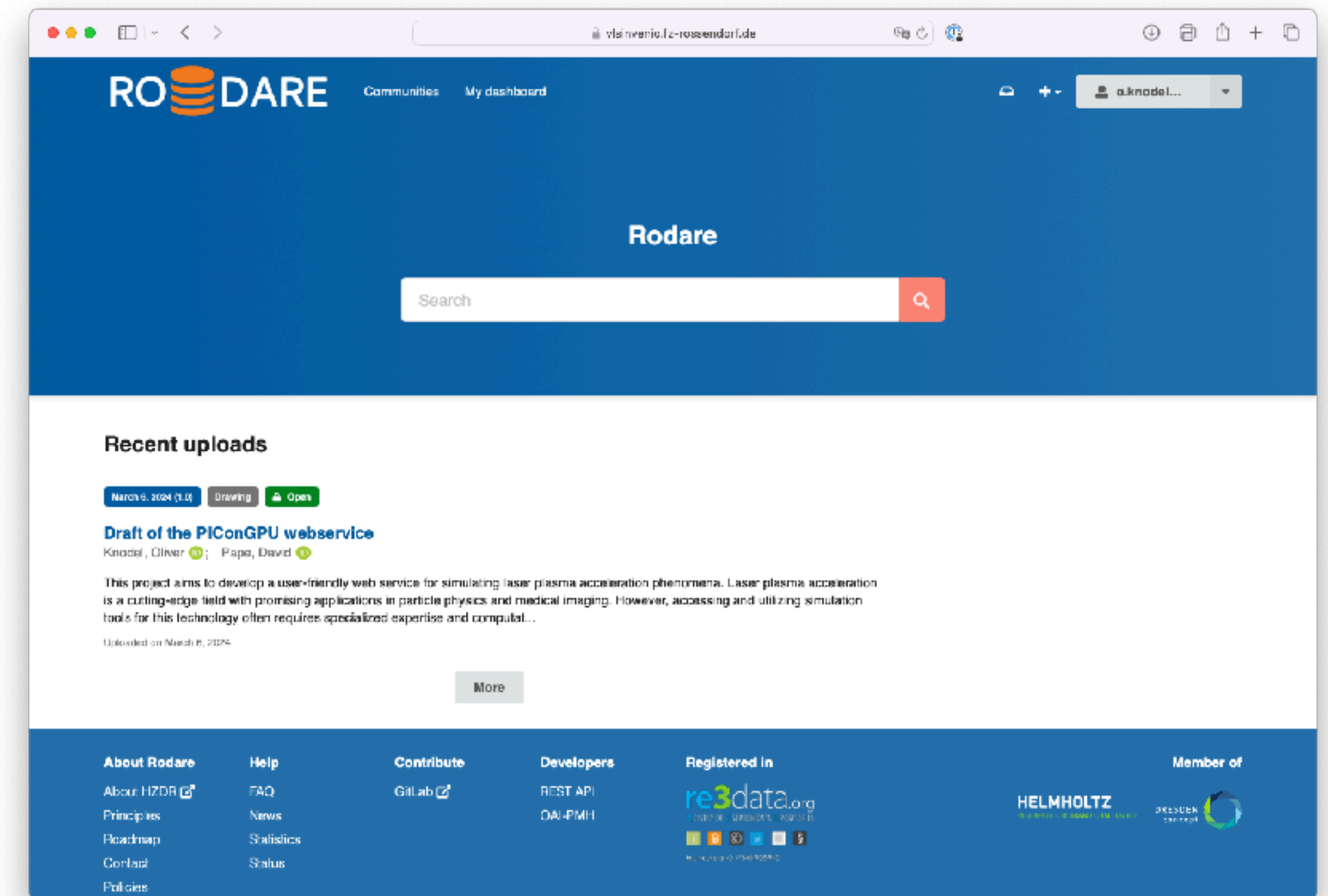
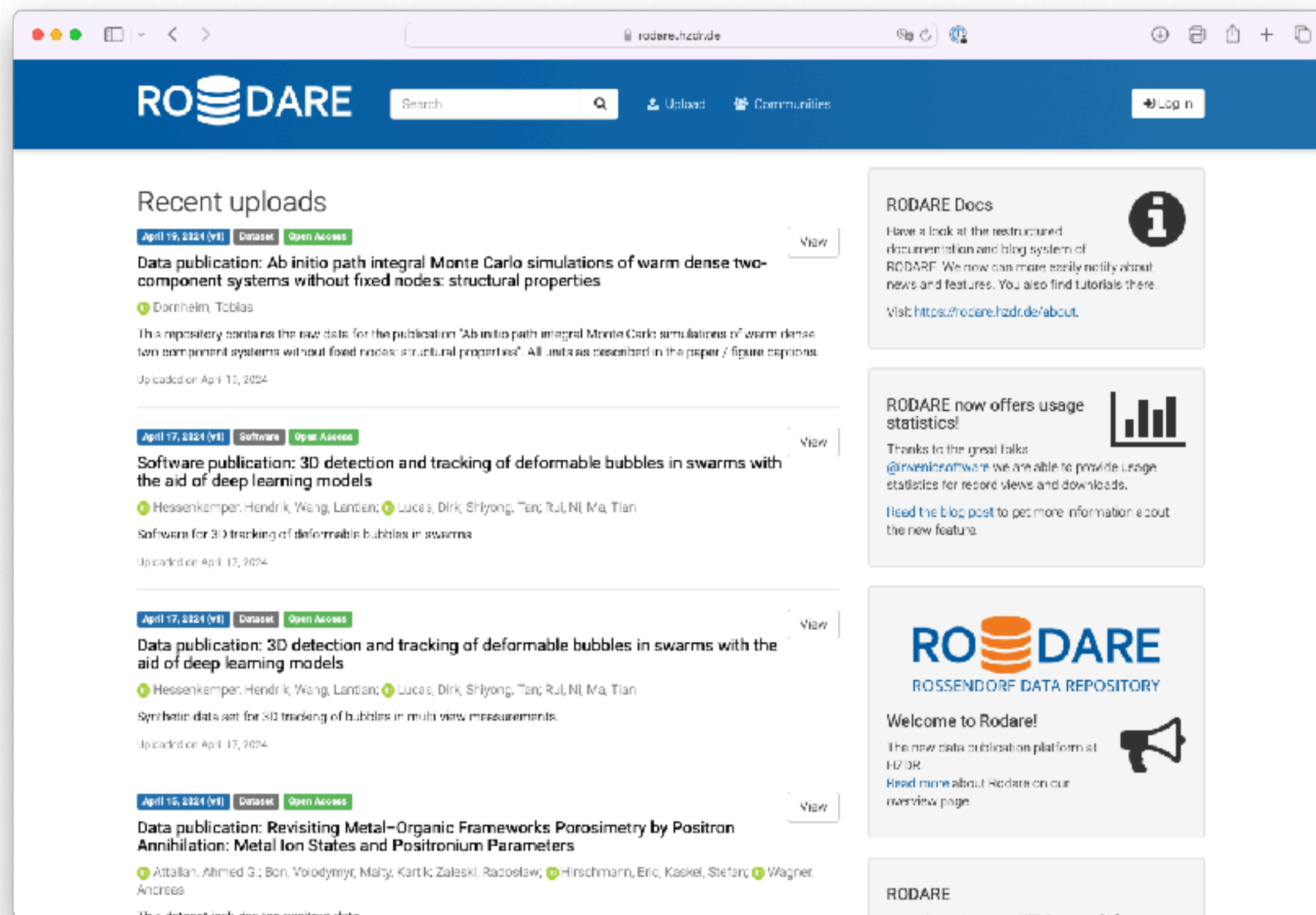
NORTHWESTERN UNIVERSITY





# 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...



# 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:

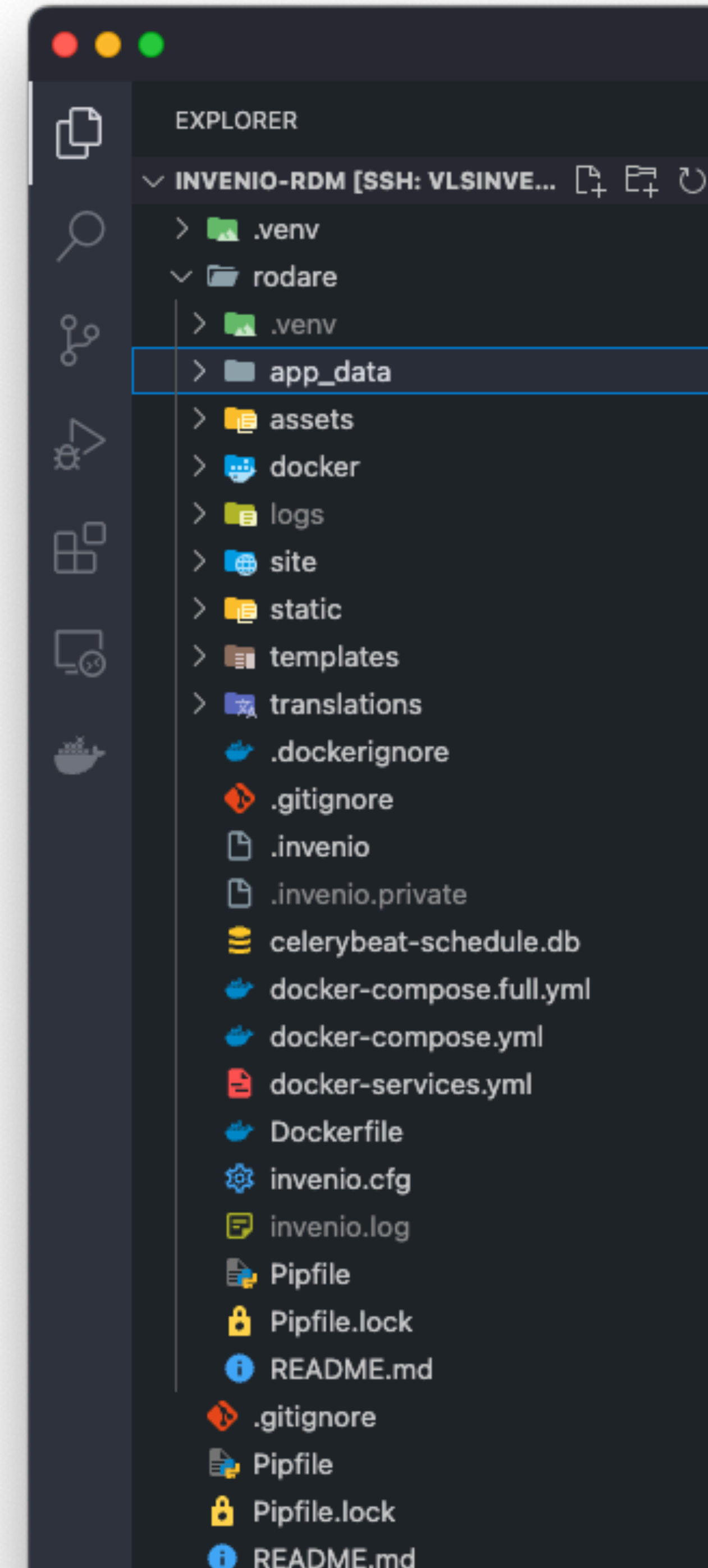
- LTS release (for production systems): v9.1
- STS release (for feature previews): v11.0

```
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:

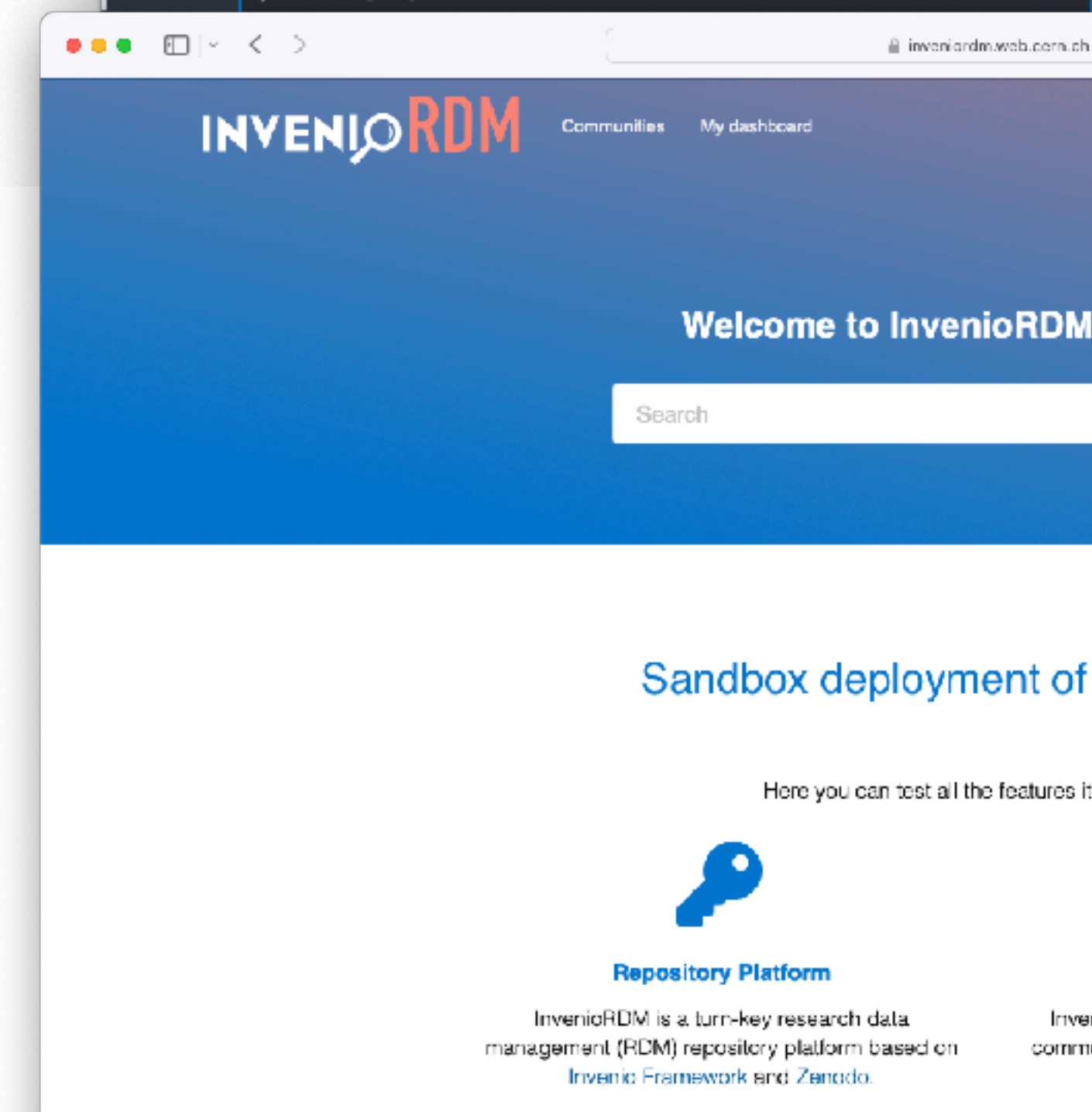
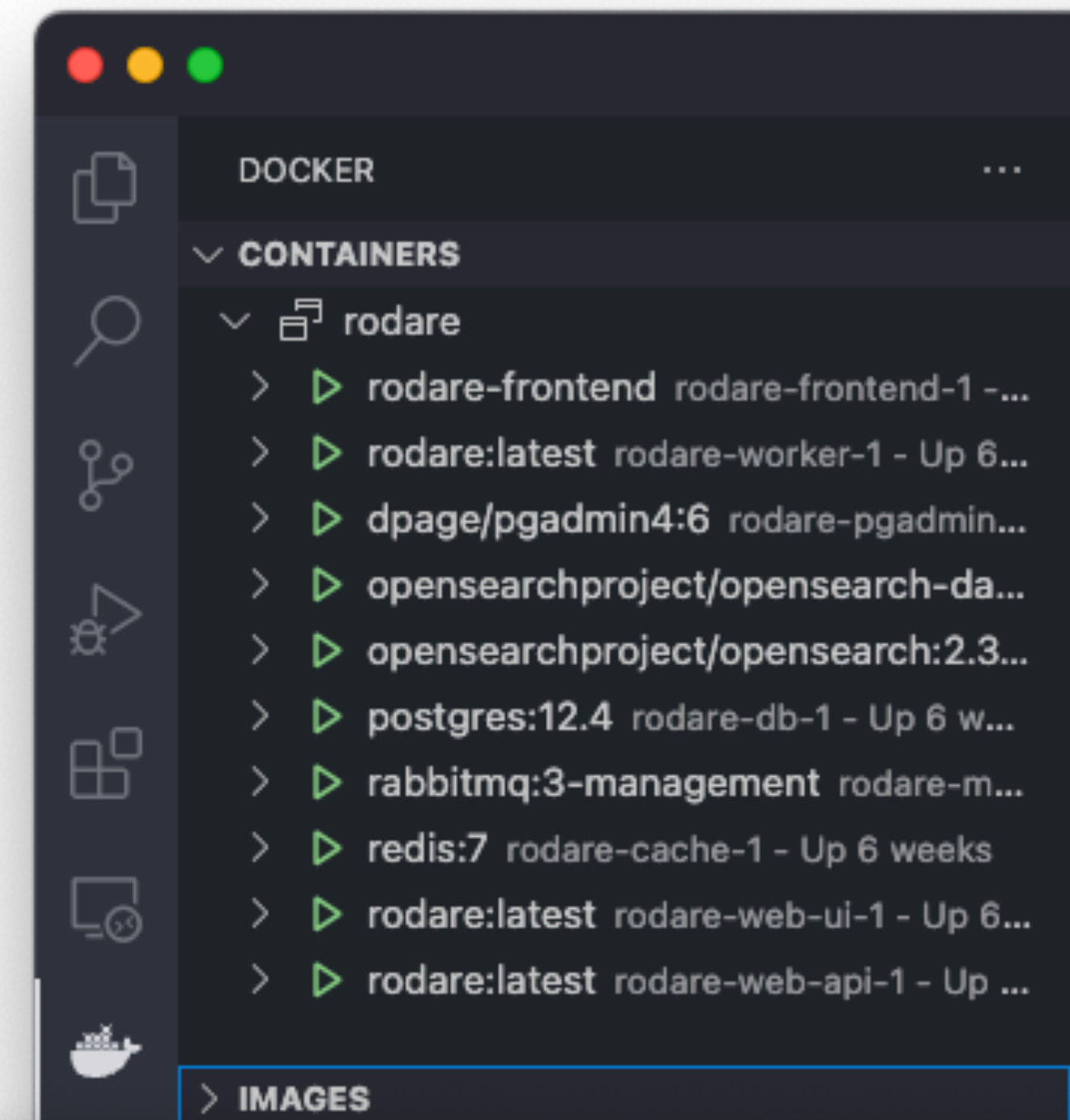
```
cd my-site/  
invenio-cli check-requirements --development
```

- You 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).

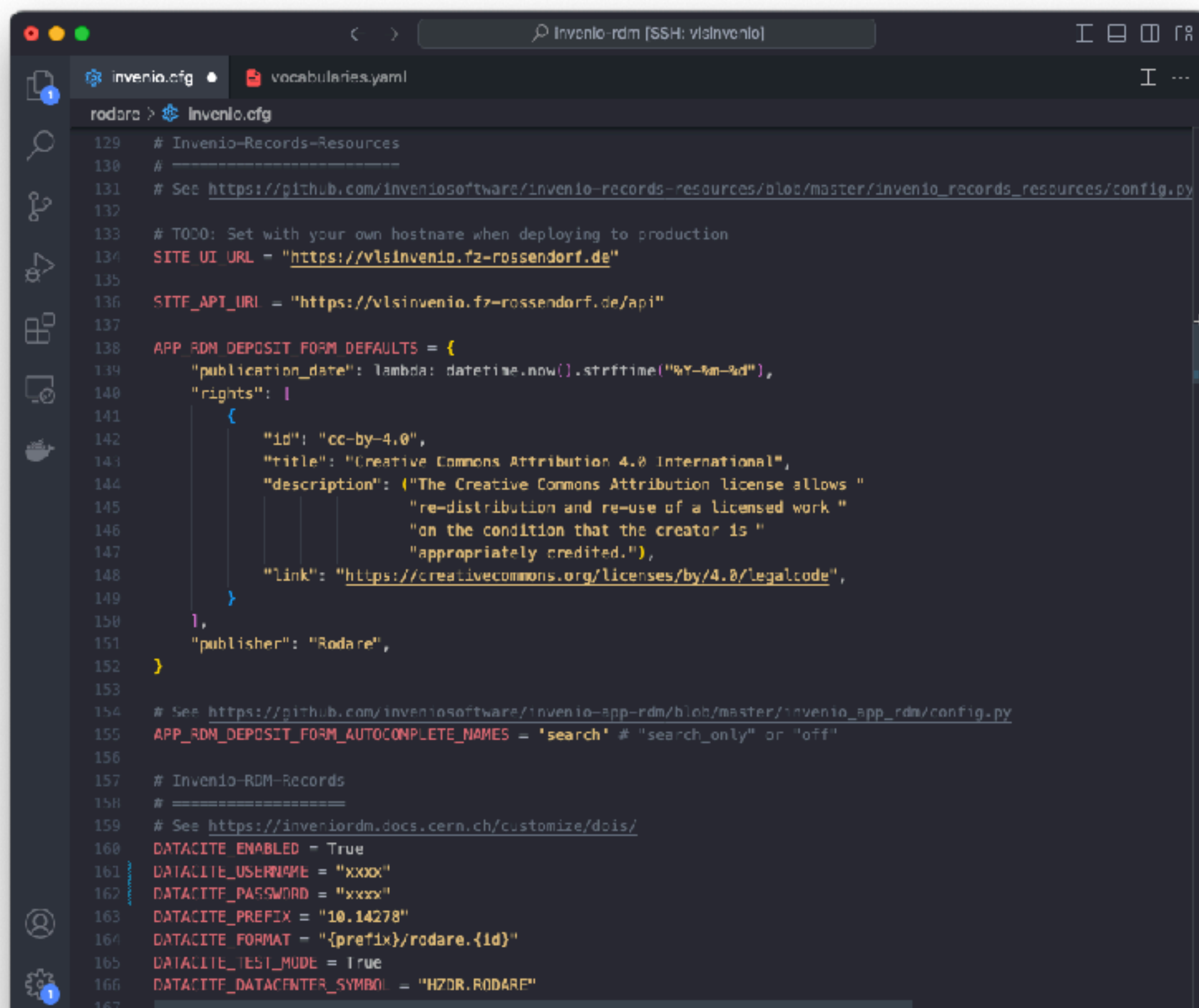
```
invenio-cli install  
invenio-cli services setup  
invenio-cli run
```

## 5. Explore InvenioRDM

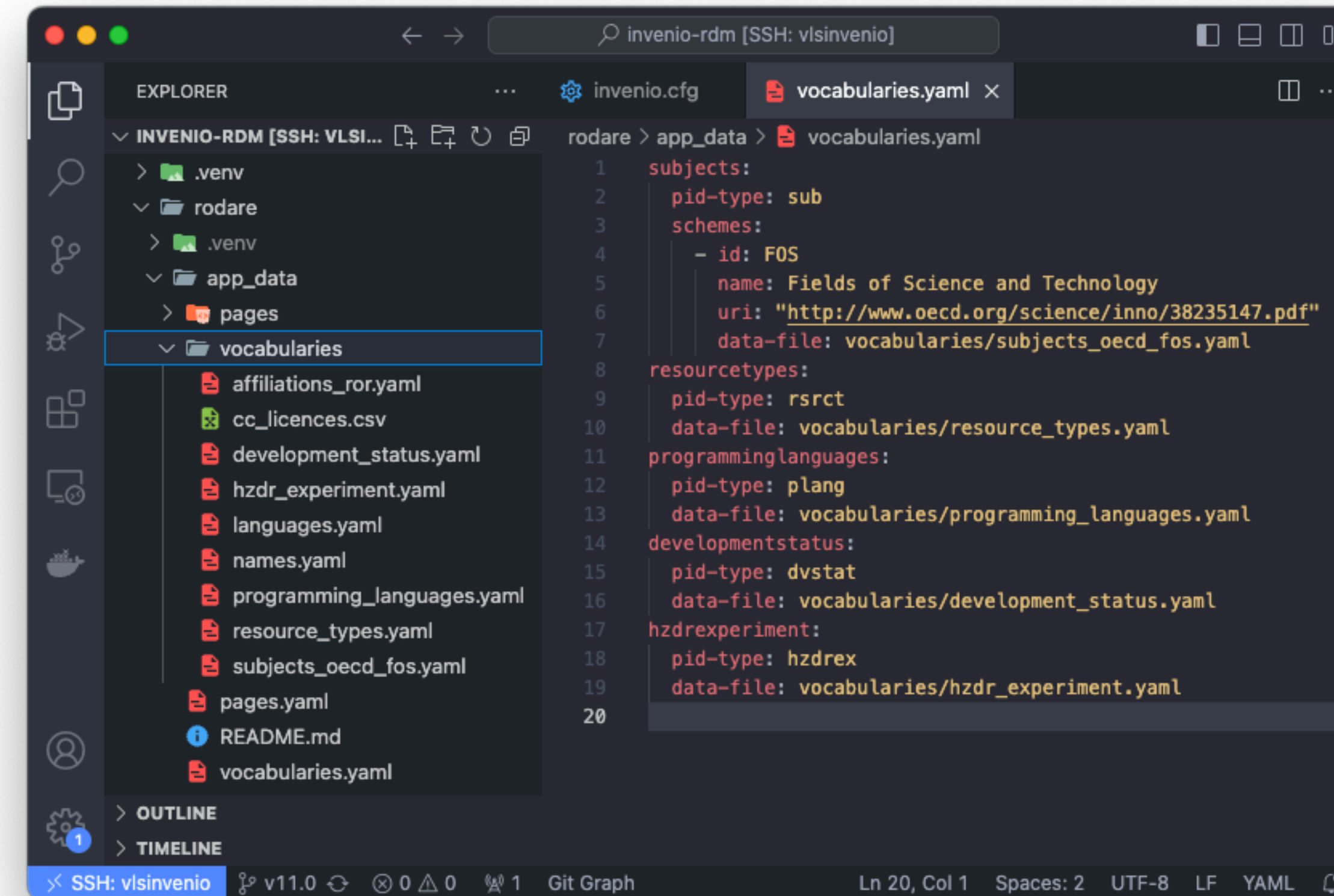


# Configuration and Customisation

- 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



```
rodare > invenio.cfg
129 # Invenio-Records-Resources
130 # =====
131 # See https://github.com/inveniosoftware/invenio-records-resources/blob/master/invenio_records_resources/config.py
132
133 # TODO: Set with your own hostname when deploying to production
134 SITE_UI_URL = "https://vlsinvenio.fz-rossendorf.de"
135
136 SITE_API_URL = "https://vlsinvenio.fz-rossendorf.de/api"
137
138 APP_RDM_DEPOSIT_FORM_DEFAULTS = {
139     "publication_date": lambda: datetime.now().strftime("%Y-%m-%d"),
140     "rights": [
141         {
142             "id": "cc-by-4.0",
143             "title": "Creative Commons Attribution 4.0 International",
144             "description": ("The Creative Commons Attribution license allows "
145                 "re-distribution and re-use of a licensed work "
146                 "on the condition that the creator is "
147                 "appropriately credited."),
148             "link": "https://creativecommons.org/licenses/by/4.0/legalcode",
149         }
150     ],
151     "publisher": "Rodare",
152 }
153
154 # See https://github.com/inveniosoftware/invenio-app-rdm/blob/master/invenio_app_rdm/config.py
155 APP_RDM_DEPOSIT_FORM_AUTOCOMPLETE_NAMES = 'search' # "search_only" or "off"
156
157 # Invenio-RDM-Records
158 # =====
159 # See https://inveniordm.docs.cern.ch/customize/dpis/
160 DATACITE_ENABLED = True
161 DATACITE_USERNAME = "xxxx"
162 DATACITE_PASSWORD = "xxxx"
163 DATACITE_PREFIX = "10.14278"
164 DATACITE_FORMAT = "{prefix}/rodare.{id}"
165 DATACITE_TEST_MODE = True
166 DATACITE_DATACENTER_SYMBOL = "HZDR.RODARE"
```

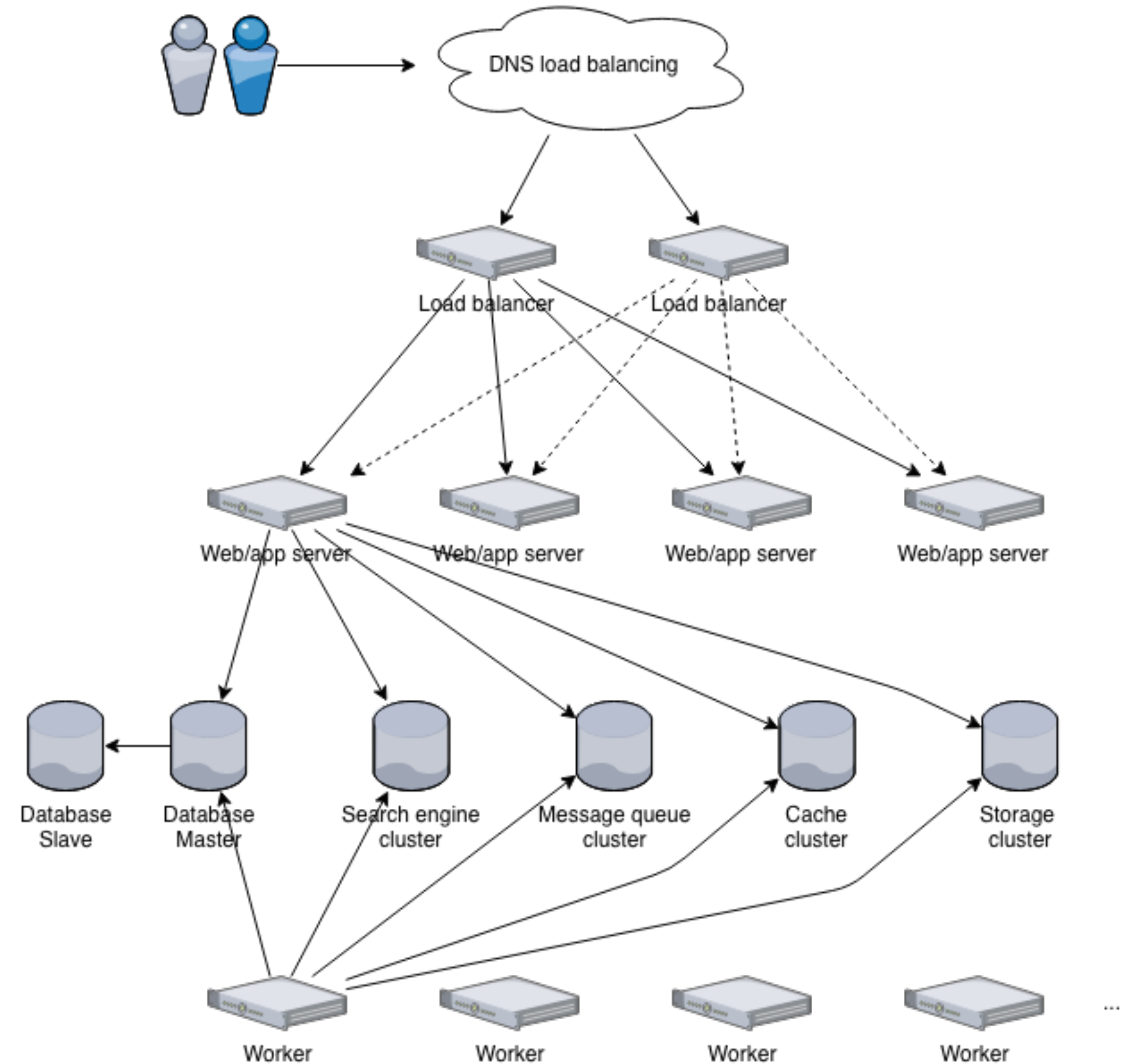


```
rodare > app_data > vocabularies.yaml
1 subjects:
2   pid-type: sub
3   schemes:
4     - id: FOS
5       name: Fields of Science and Technology
6       uri: "http://www.oecd.org/science/inno/38235147.pdf"
7       data-file: vocabularies/subjects_oecd_fos.yaml
8   resourcetypes:
9     pid-type: rsrct
10    data-file: vocabularies/resource_types.yaml
11   programminglanguages:
12     pid-type: plang
13     data-file: vocabularies/programming_languages.yaml
14   developmentstatus:
15     pid-type: dvstat
16     data-file: vocabularies/development_status.yaml
17   hzdrexperiment:
18     pid-type: hzdrex
19     data-file: vocabularies/hzdr_experiment.yaml
20
```

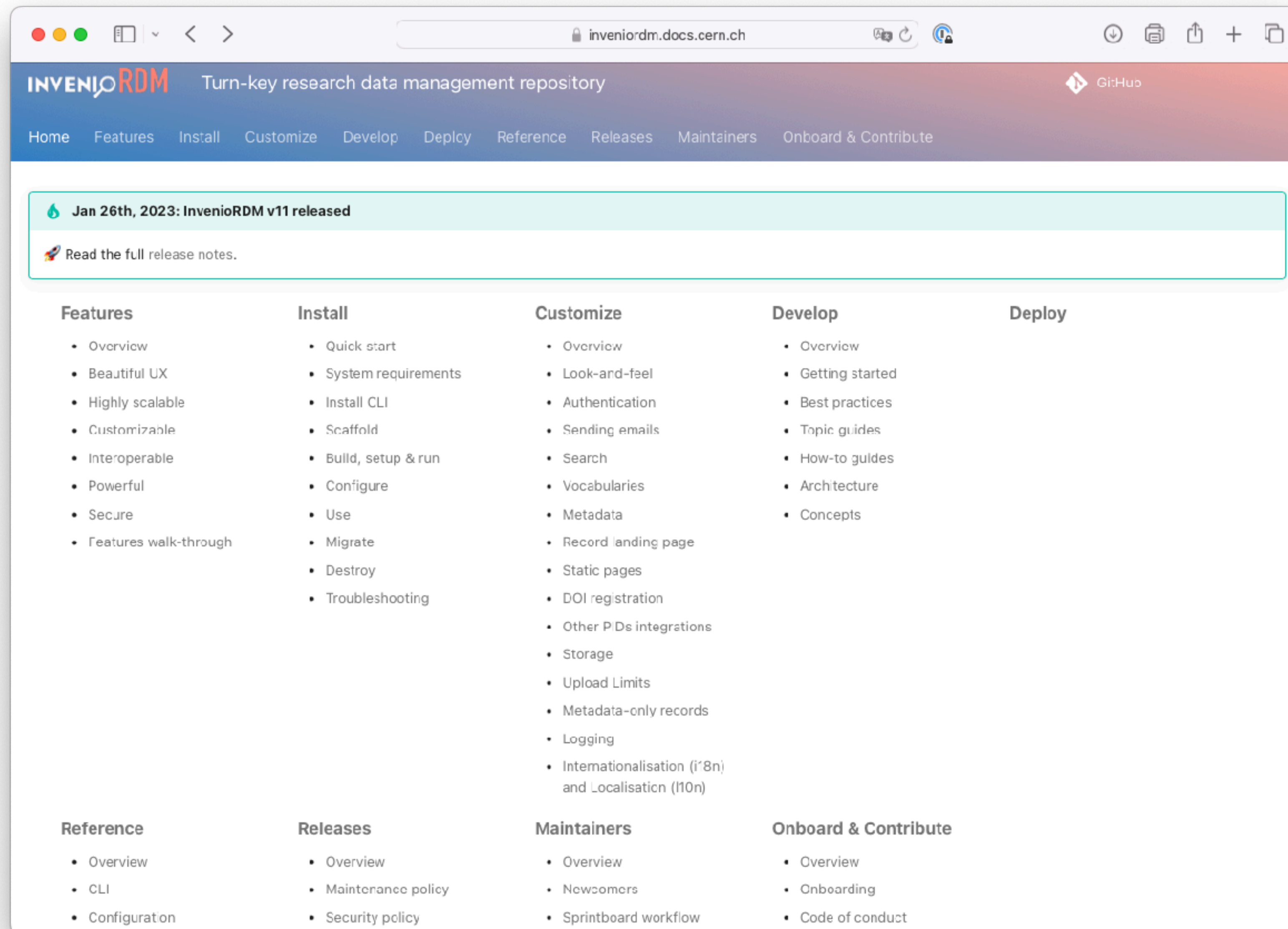


# Infrastructure

- **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](https://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



# 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.



Name	Run No.	Source Folder	Size	Start Time	Type	Image
upload		...nfs	0 B	2022-07-05 Tue 11:28	derived	
Second dataset		...nfs	0 B	2022-07-05 Tue 10:29	derived	
Dataset		...nfs	0 B	2022-07-05 Tue 10:15	derived	

**General Information**

Name	Standard flat foil
Description	Standard flat foil
PID	HZDR/cc39913e-7551-41b4-859c-b628174490f5
Type	r2w
Creation Time	2023-09-19 10:16
Keywords	Smilei Simulation Testing

**Creator Information**

Owner	fwkt_ingestor
Principal Investigator	fwkt@scicat.hzdr.de
Orcid	orcid.org/0000-0002-9261-7643
Contact Email	fwkt@scicat.hzdr.de
Owner Group	fwkt_ingestor
Access Groups	fwkt@scicat.hzdr.de

**File Information**

Source Folder	/bigdata/hplsm/scratch
Size	13 / KB
Data Format	py

**Related Documents**

Creation Location	HZDR/Smilei
Techniques	Simulation

**Scientific Metadata**

Simulation	
Laser	
Box_size	wnin
Omega	1 (laser_freq)
a_0	30
Wavelength	0.8 (laser_wavelength)
Delay	121.39355818427993
Laser_shape	gaussian
Specifics	
Particle_Species	
Layers	



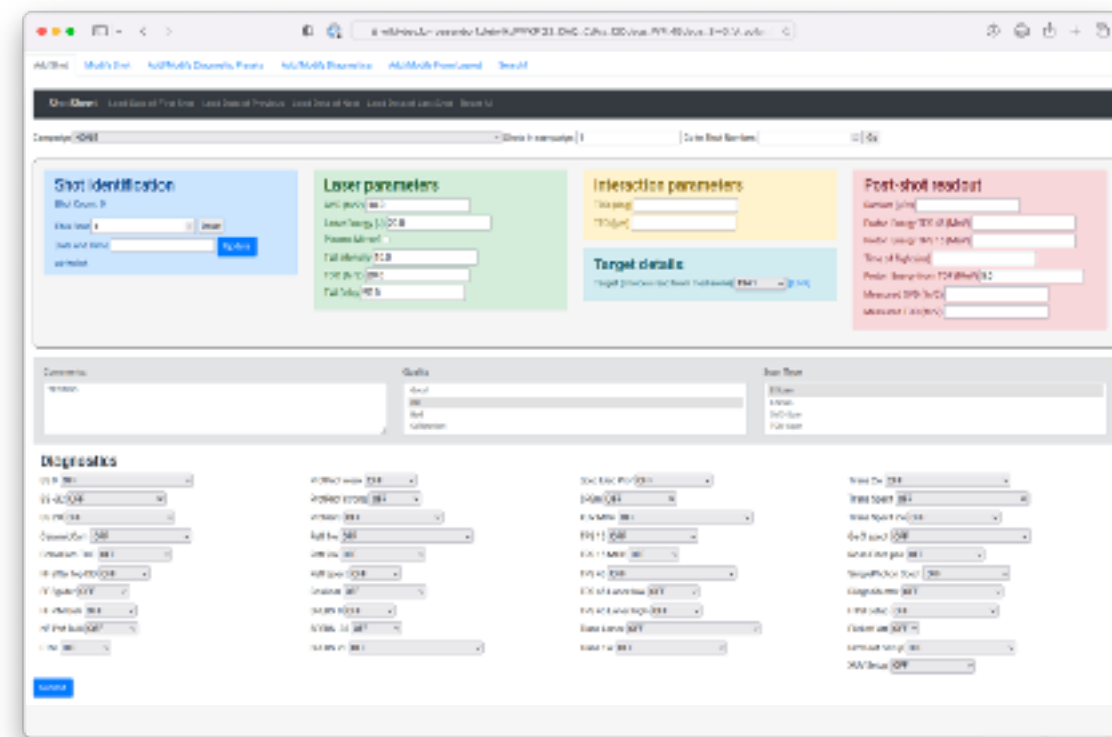
# Metadata Catalogue SciCat and Data Repository RODARE (Draft)

## Curated Metadata Source

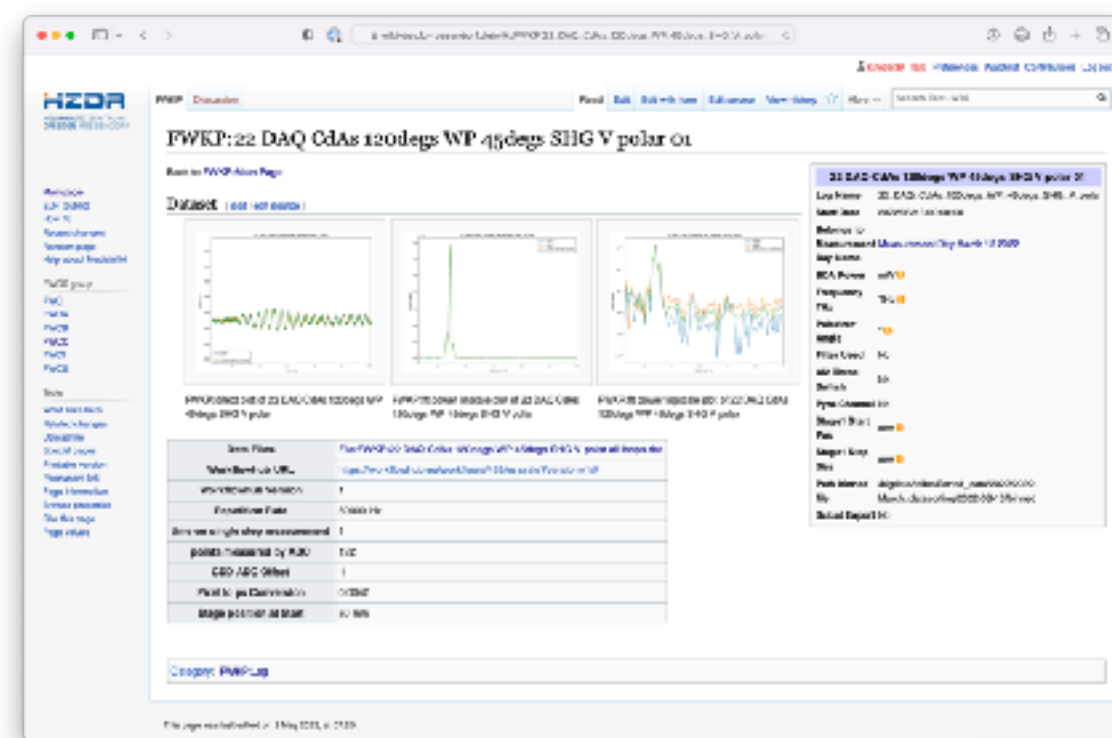
## Public Metadata Catalogue

## Data Access

ExperimentLogging app (ExL)

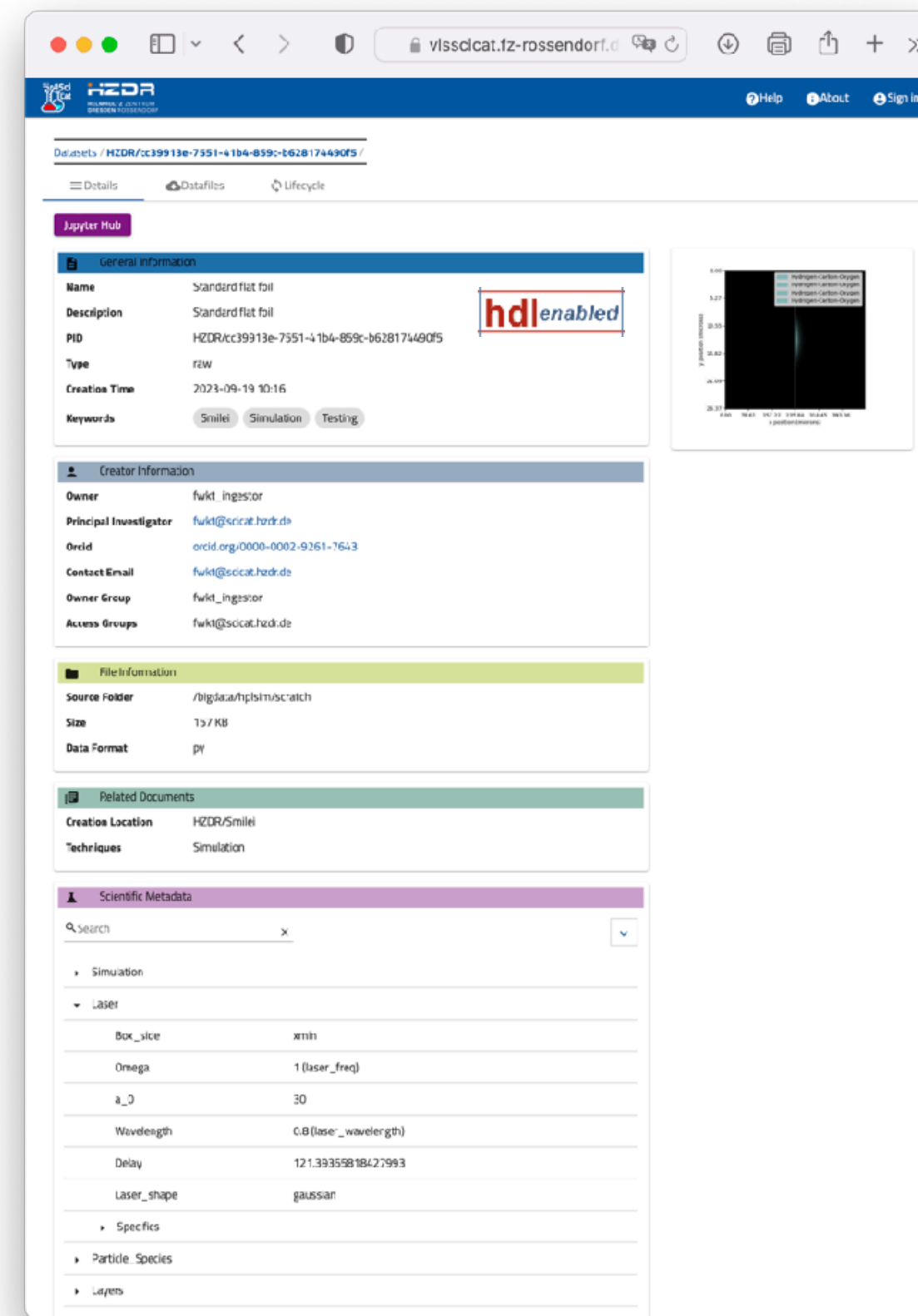


E-Logbook



Fully Automated Process

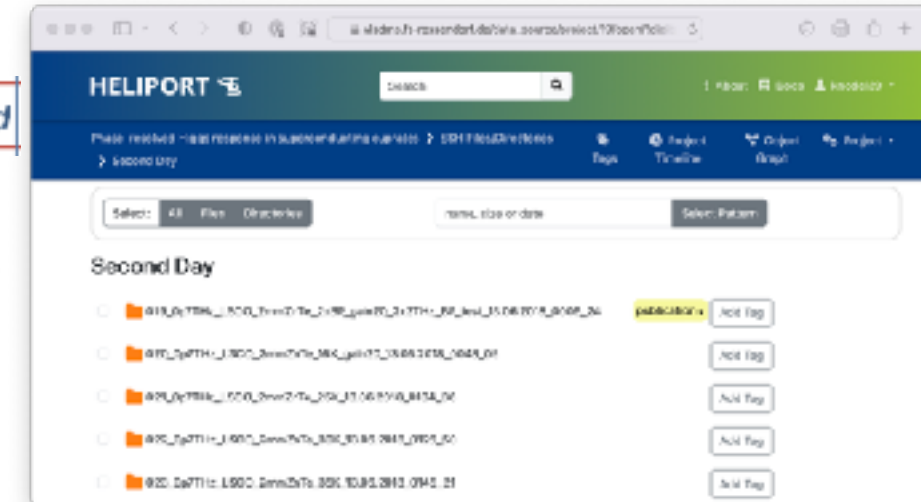
SciCat



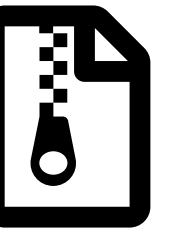
RODARE



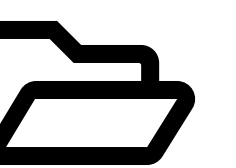
HELIPORT



Dataset



Filesystem



Tape Archive



Metadata from Experiment/Simulation