



Navigating the Data Patchwork: Strategies for Integrating Metadata Catalogs, Data Publications, and Archives

NOBUGS, Grenoble, September 2024

O. Knodel, D. Pape, M. Voigt, T. Gruber, S. Müller and G. Juckeland



Outline

RDM Ecosystem and Strategy

Policy and data publication
systems and services at HZDR

The HZDR

Research areas in the fields:
Energy, Health, Matter

(Meta)Data Patchwork

Approach to combine metadata
sources from various different
systems



Our Research Facility and our Large Scale Research Infrastructures

The Helmholtz-Zentrum Dresden - Rossendorf

- Employees approx. **1,470**. Thereof **670** scientists.
- Member of:

HELMHOLTZ
RESEARCH FOR GRAND CHALLENGES

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)

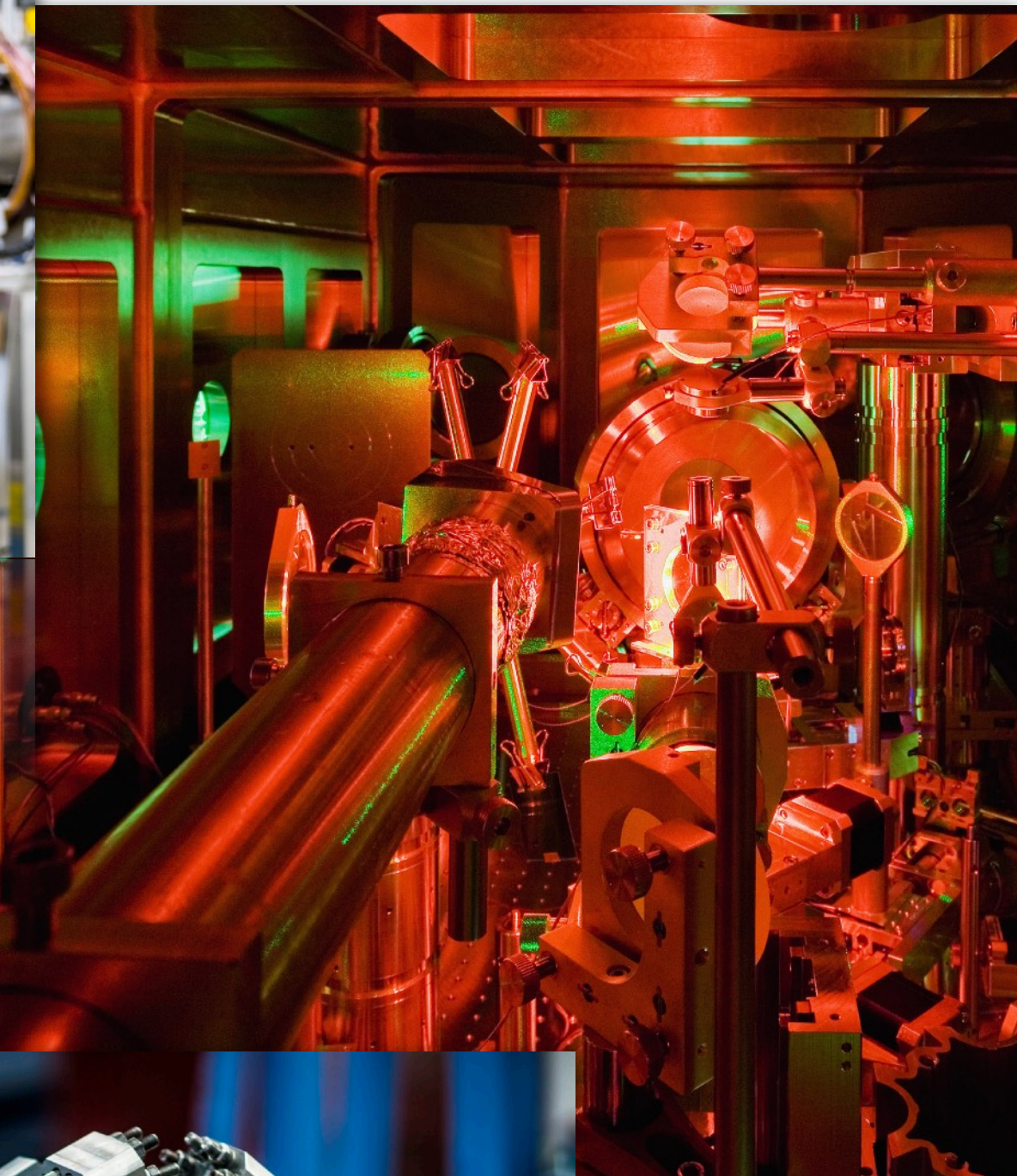
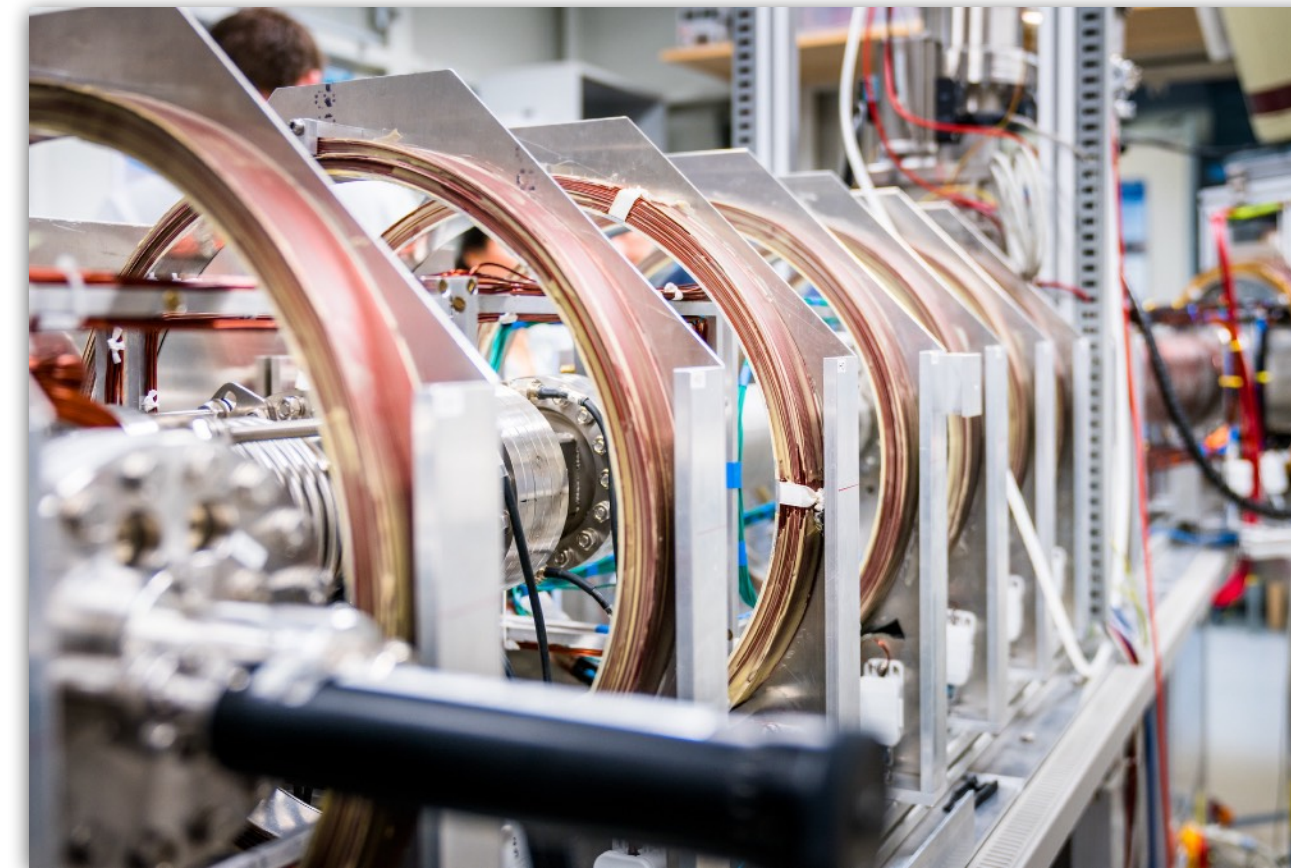
- Europe's highest pulsed magnetic fields.

Ion Beam Center (IBC)

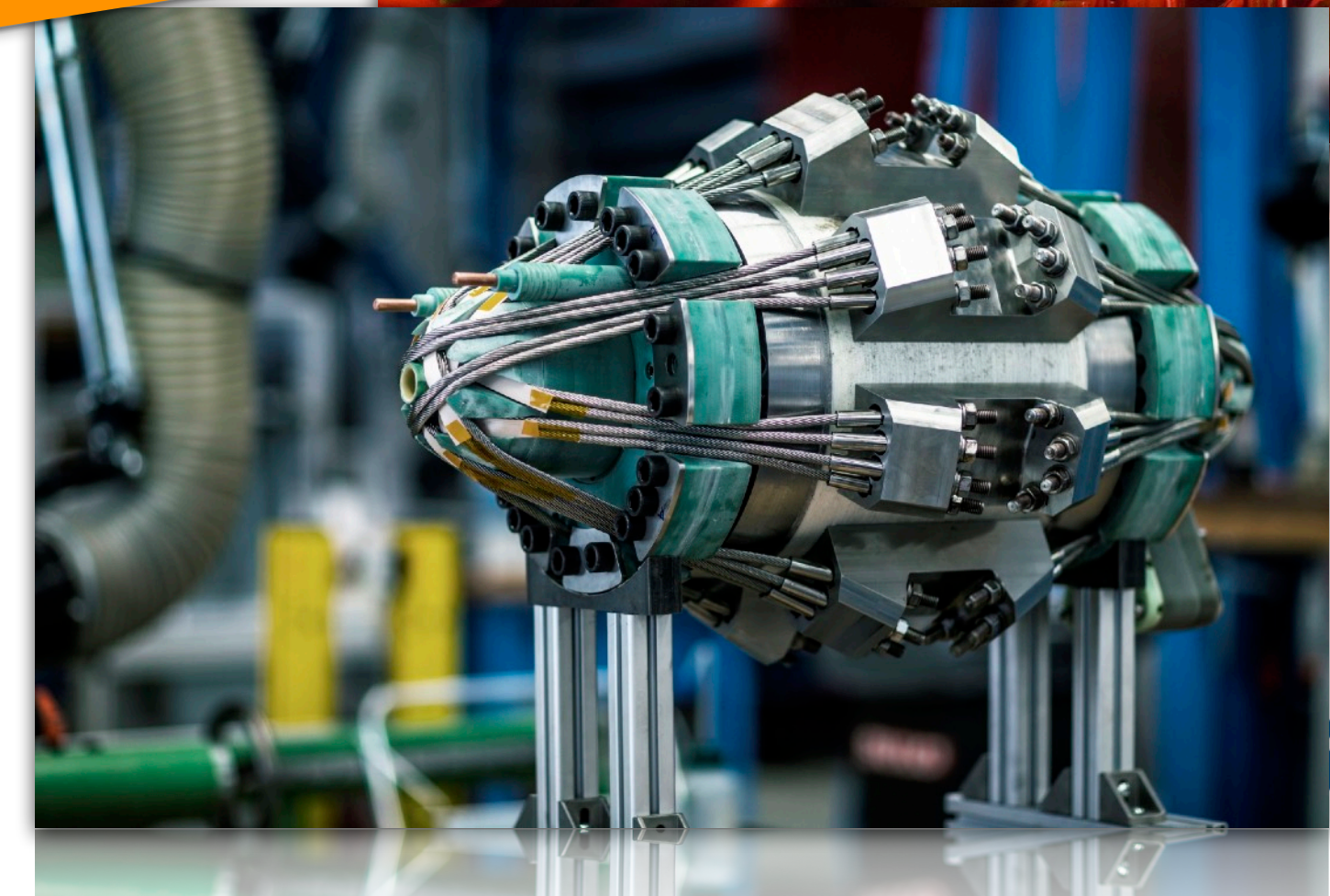
- Nanoscale surface analysis and modification.

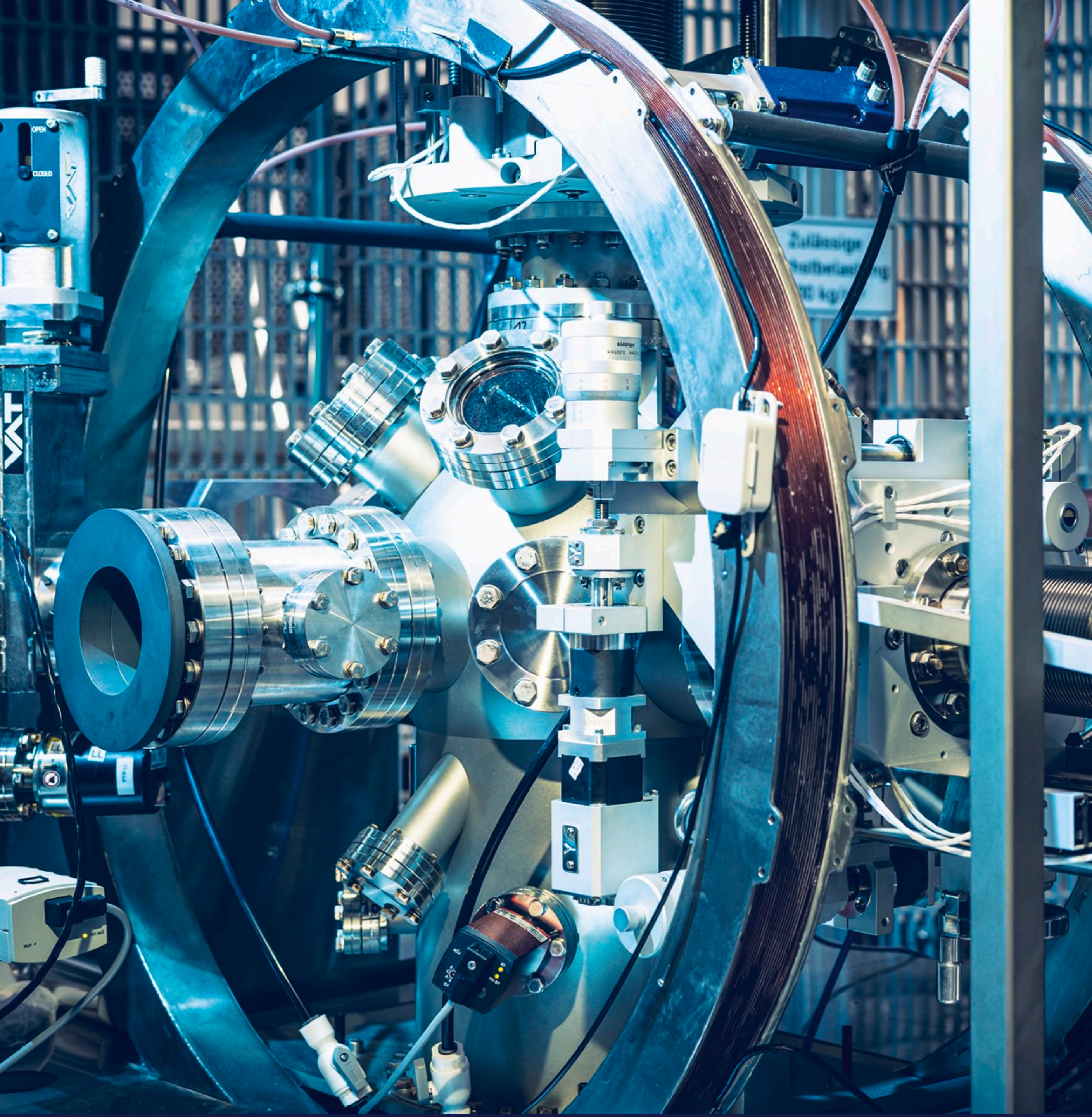
DREsden Sodium facility for DYNamo (DRES-DYN)

- Large scale liquid sodium experiments with geo- and astrophysical background.



The Rossendorf Beamline at ESRF (ROBL)



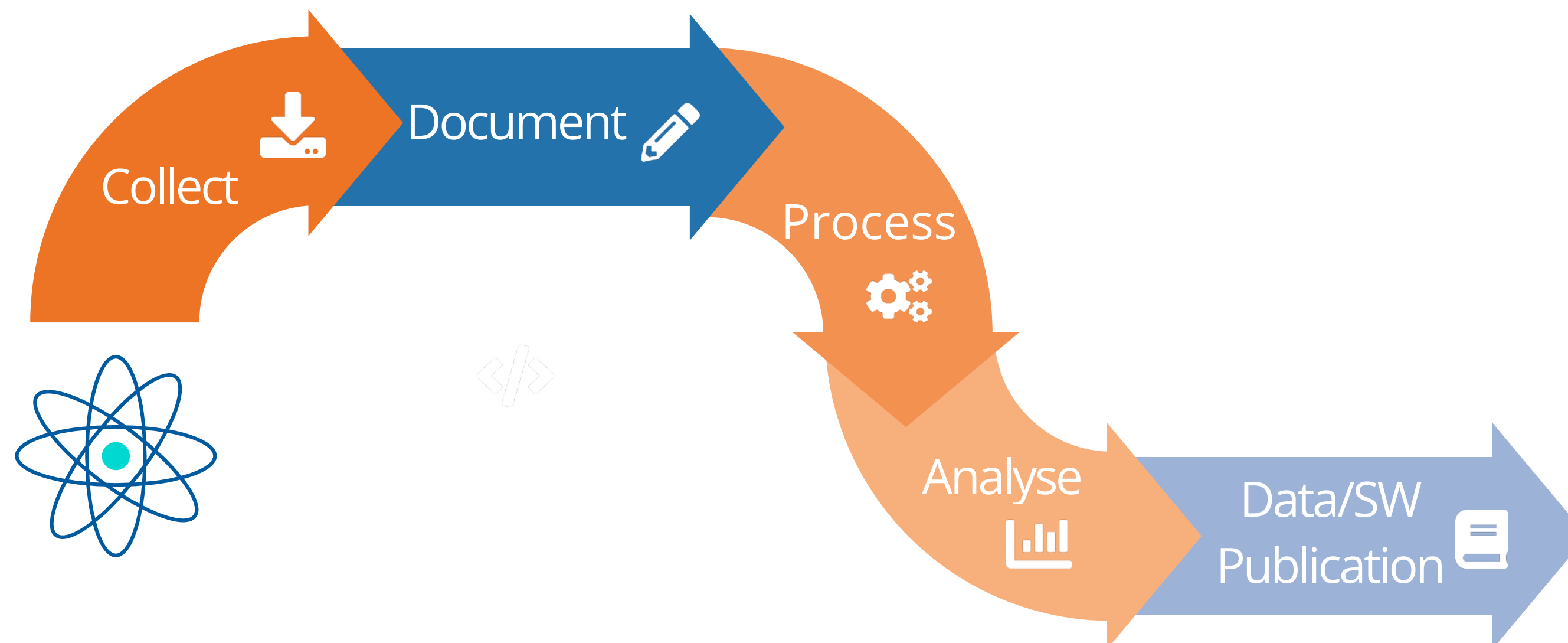


Our RDM Ecosystem and Strategy



The Mission for the Data Management and HPC Group

- Supporting our scientists and research experiment with tools:
 - electronic lab books,
 - interactive analysis,
 - publication repositories for data and software,
 - scientific workflow management,
 - PID (handle) generation and management.
- Providing support in connecting experiments, detectors and diagnostics to our infrastructure.
- Establishment of analysis workflows and HPC applications.



Development of the HZDR

Data Management Ecosystem

Scientific Computing Department at HZDR — 2017

Establishment of the Scientific Computing division within the IT as a bridge to the science

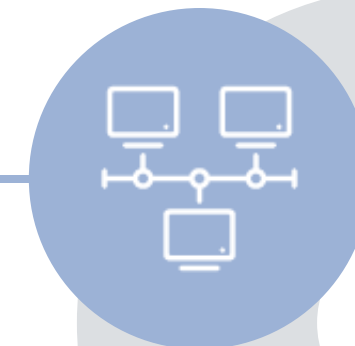
Data Policy — 2018

Legal framework for the institutional data management at HZDR.



RDM Landscape and Data Management Group — 2019

Analysis of data/service landscape at HZDR and development of an uniform data lifecycle.



HZDR RDM Strategy — 2025

Identification of the need to develop a data management strategy together with our scientists.



Data Repository RODARE — 2018

Provision of the Rosendorf Data Repository for data and software publications.



HZDR Metadata Catalogue — 2024

Metadata catalogue for additional experiment-specific metadata as extension of RODARE.

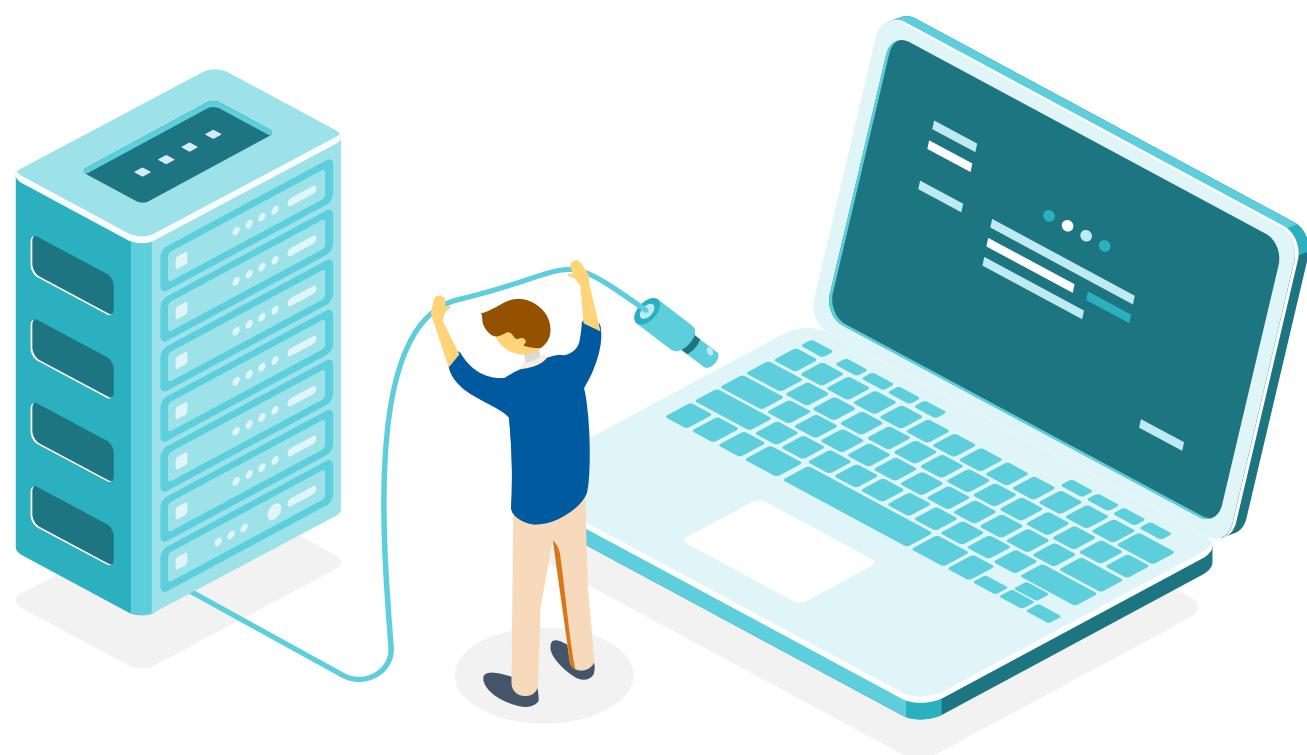


Our Motivation for Developing a Full Digital Data Lifecycle:

- The HZDR has a data policy since May 2018.
- Reasons for the development:
 - Establishment and legitimisation of publication, handling and deletion of data generated or taken at the HZDR,
 - Legal framework for data management and...




... to further develop data management services together with our scientists!



The HZDR Data Policy

DOI10.14278/rodare.2269

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

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

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 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

Our Contribution to Support FAIR Research: RODARE



[Upload](#) [Communities](#) [Log in](#)

Recent uploads

September 24, 2024 (v1) Dataset Open Access

Data publication: Cavity-mediated thermal control of metal-to-insulator transition in 1T-TaS2

Jarc, Giacomo; Mathengattil, Shahla Yasmin; Montanaro, Angela; Giusti, Francesca; Rigoni, Enrico Maria; Sergo, Rudi; Fassioli, Francesca; Winnerl, Stephan; Zilio, Simone Dal; Mihailovic, Dragan; Prelovšek, Peter; Eckstein, Martin; Fausti, Daniele

Original datasets corresponding to the publication.

Uploaded on September 24, 2024

View

September 19, 2024 (v1) Dataset Open Access

Heat flow data from the fungus Schizophyllum commune

Fahmy, Karim; Günther, Alix; Bertheau, Rahel; Pape, David

The data set contains three typical heat flow curves recorded from the fungus Schizophyllum commune and exemplifies the evaluation of such data by the software tool metabolator (<https://rodare.hzdr.de/record/3150>).

Uploaded on September 19, 2024

View

September 19, 2024 (v0.2.0) Software Open Access

METABOLATOR: Analysis of Microcalorimetric Metabolic Data Using Monod's Equation

Pape, David; Lokamani, Mani; Seal, Ayush; Kelling, Jeffrey; Knodel, Oliver; Fahmy, Karim; Juckeland, Guido

Curve fitting automation for metabolic load of bacteria in solutions.

Uploaded on September 19, 2024

2 more version(s) exist for this record

View

September 18, 2024 (v1) Dataset Open Access

Data publication: Formation of martensitic microstructure in epitaxial Ni-Mn-Ga films after fast cooling

View

RODARE Docs

Have a look at the restructured documentation of RODARE. We now can more easily notify about news and features, such as the search options in Rodare.

Visit us <https://rodare.hzdr.de/about>.



Featured HZDR Large-Scale Research Facilities

In order to make our large-scale HZDR facilities more visible, Rodare now offers specific communities and an overview on our front page!



Overview of published data sets of selected large-scale research facilities of the HZDR:

- **ATHENA** — Accelerator Technology Helmholtz iNfrAstructure
- **CARBOSOLA**
- **Dresden High Magnetic Field Laboratory**
- **DRESDYN** — DREsden Sodium facility for DYNamo and thermohydraulic studies
- **ELBE** — Electron Linac for beams with high Brilliance and low Emittance
 - ▶ ELBE Beamlines
- **Felsenkeller** — The underground ion accelerator lab
- **HECToR** — High-power ultrafast computed tomography
- **HIBEF** — Helmholtz International Beamline for Extreme Fields


Powered by:




Harvested via OAI-PMH by:



Registered in:





REGISTRY OF RESEARCH DATA REPOSITORIES



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

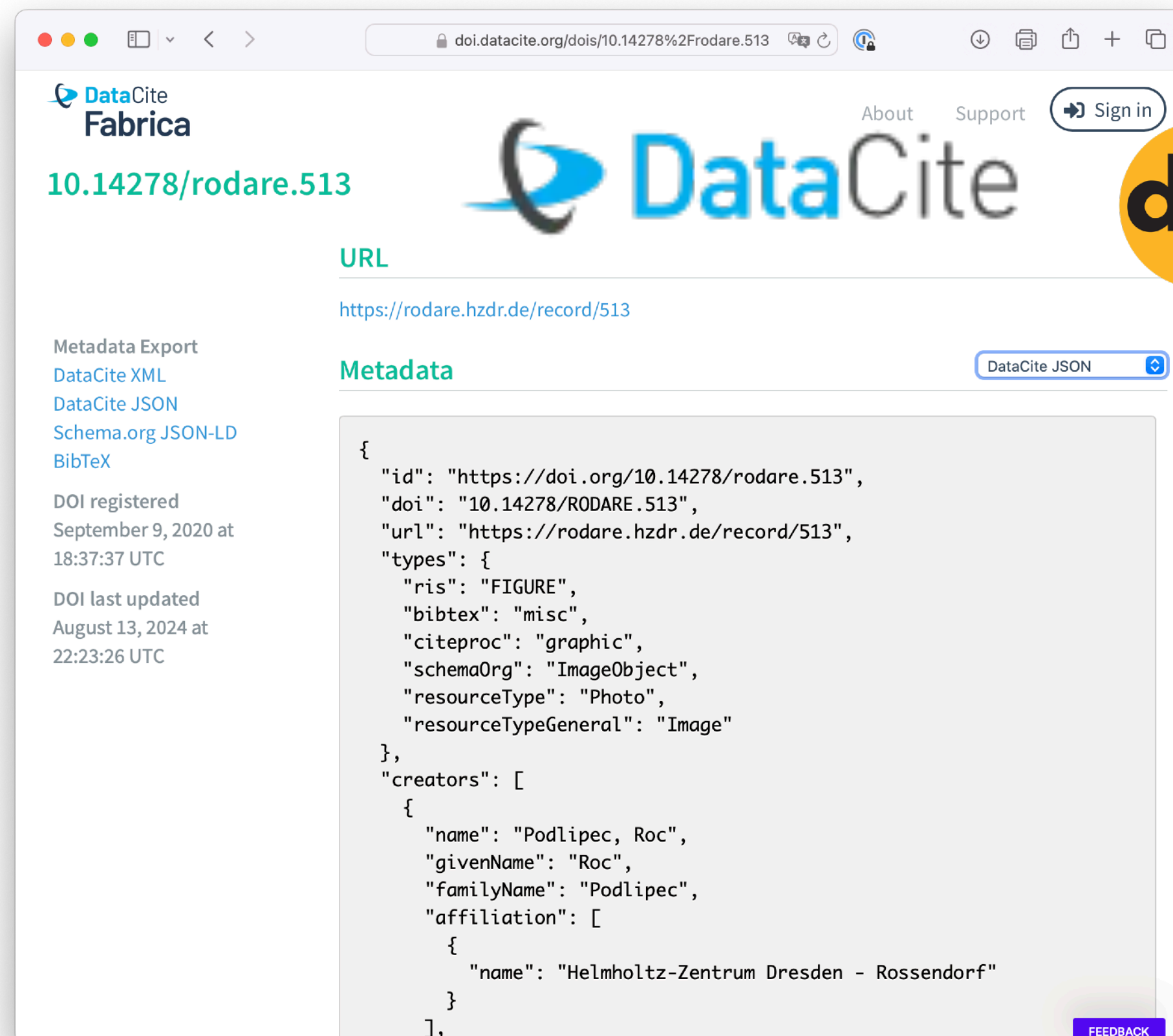
DRESDEN
concept



8

Metadata and Data in RODARE

- In our data publication system the datasets are described via **DataCite** metadata to be FAIR.
- The DataCite metadata is attached to the DOI and harvested via portals, such as **B2Find**.



DOI: 10.14278/rodare.513

URL: <https://rodare.hzdr.de/record/513>

Metadata Export

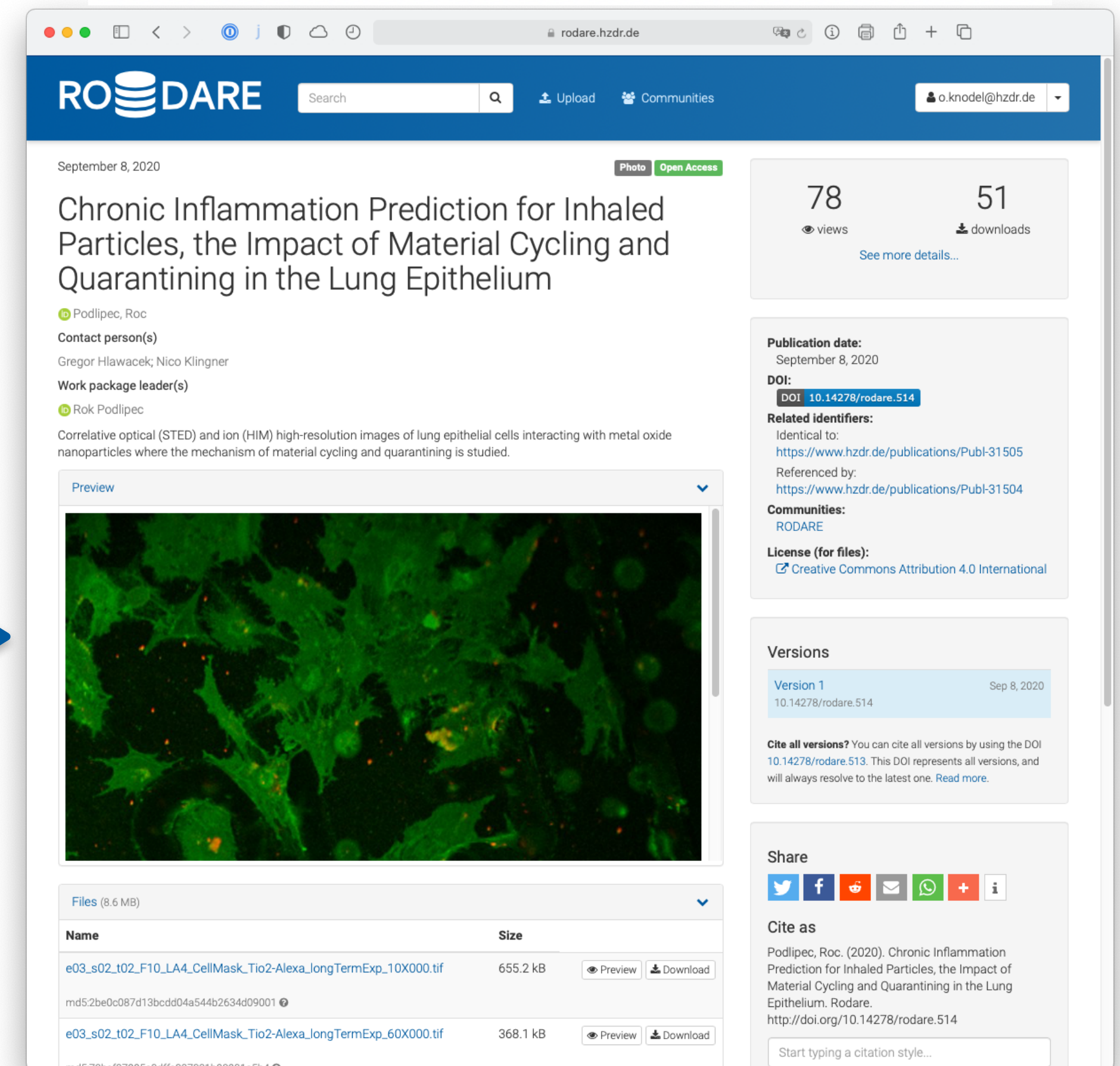
- DataCite XML
- DataCite JSON
- Schema.org JSON-LD
- BibTeX

DOI registered: September 9, 2020 at 18:37:37 UTC

DOI last updated: August 13, 2024 at 22:23:26 UTC

Metadata (DataCite JSON)

```
{
  "id": "https://doi.org/10.14278/rodare.513",
  "doi": "10.14278/RODARE.513",
  "url": "https://rodare.hzdr.de/record/513",
  "types": {
    "ris": "FIGURE",
    "bibtext": "misc",
    "citeproc": "graphic",
    "schemaOrg": "ImageObject",
    "resourceType": "Photo",
    "resourceTypeGeneral": "Image"
  },
  "creators": [
    {
      "name": "Podlipec, Roc",
      "givenName": "Roc",
      "familyName": "Podlipec",
      "affiliation": [
        {
          "name": "Helmholtz-Zentrum Dresden - Rossendorf"
        }
      ]
    }
  ]
}
```



September 8, 2020

Photo Open Access

78 views 51 downloads

See more details...

Publication date: September 8, 2020

DOI: 10.14278/rodare.514

Related identifiers: Identical to: <https://www.hzdr.de/publications/Publ-31505> Referenced by: <https://www.hzdr.de/publications/Publ-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.513. This DOI represents all versions, and will always resolve to the latest one. Read more.

Share

Cite as

Podlipec, Roc. (2020). Chronic Inflammation Prediction for Inhaled Particles, the Impact of Material Cycling and Quarantining in the Lung Epithelium. Rodare. <http://doi.org/10.14278/rodare.514>

Start typing a citation style...

Files (8.6 MB)

Name	Size	Preview	Download
e03_s02_t02_F10_LA4_CellMask_Tio2-Alexa_longTermExp_10X000.tif	655.2 kB		
md5:2be0c087d13bcd04a544b2634d09001			
e03_s02_t02_F10_LA4_CellMask_Tio2-Alexa_longTermExp_60X000.tif	368.1 kB		
md5:78be127005a3dfc237331b98921c5b4			

File Access:
Rest API



Filesystem

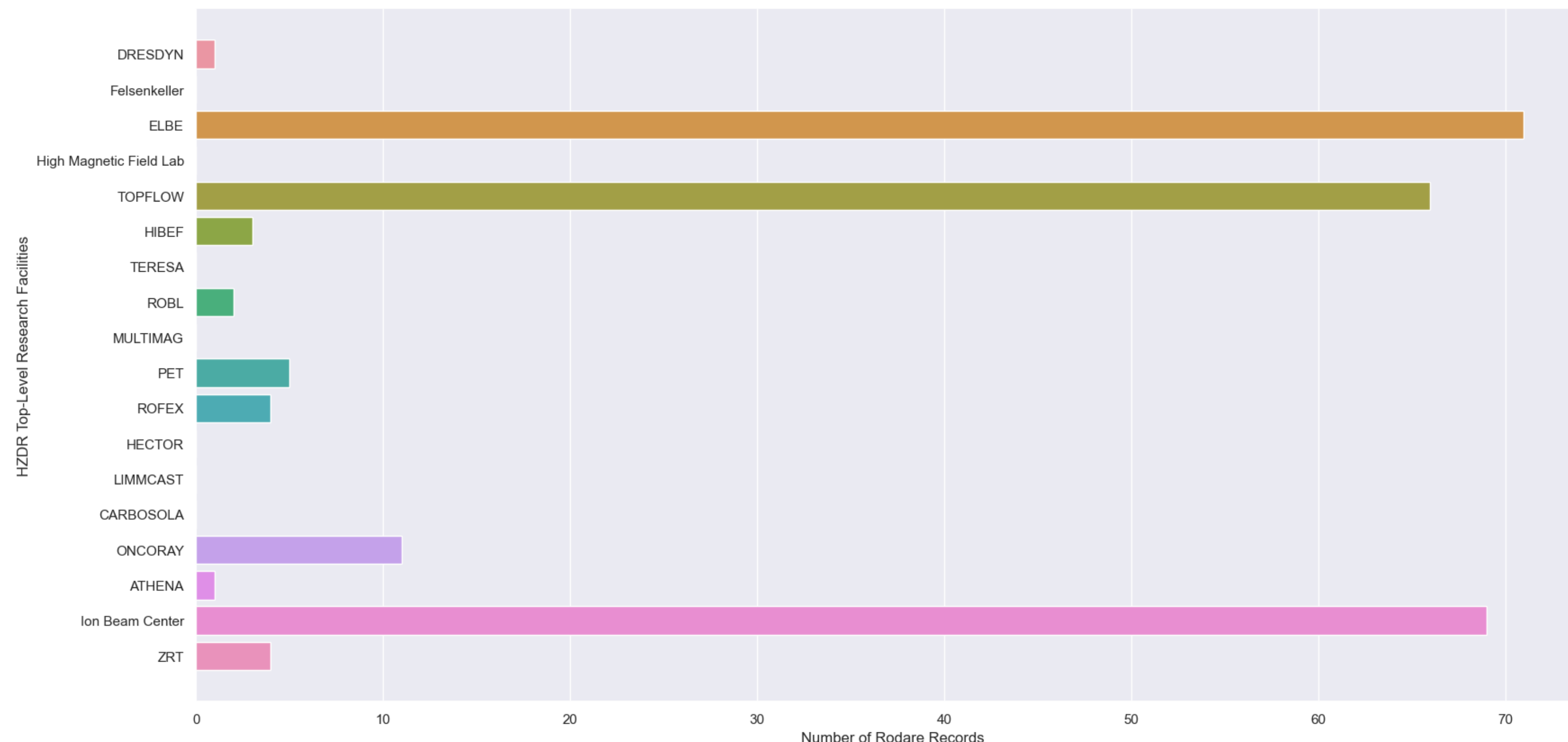
Metadata API:
Rest API, OAI-PMH



HZDR

Additional Metadata Through Communities

- Experiment-specific metadata is difficult because the RODARE metadata is the same for every entry.
- We integrated the facility information throughout Communities for our large-scale facilities:
 - **18** facilities from different research areas...
 - ...and one facility (ELBE radiation source) with **9** beamlines.



Featured HZDR Large-Scale Research Facilities

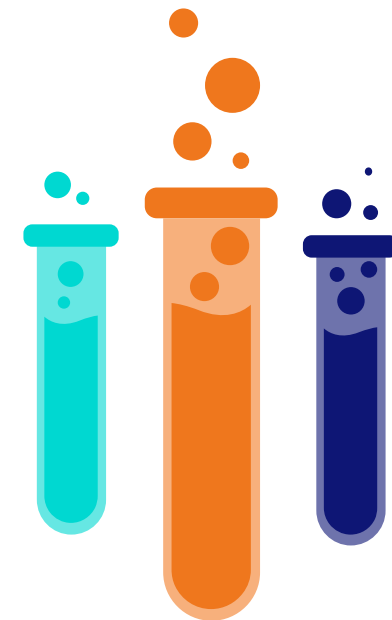


In order to make our large-scale HZDR facilities more visible, Rodare now offers specific communities and an overview on our front page!

Overview of published data sets of selected large-scale research facilities of the HZDR:

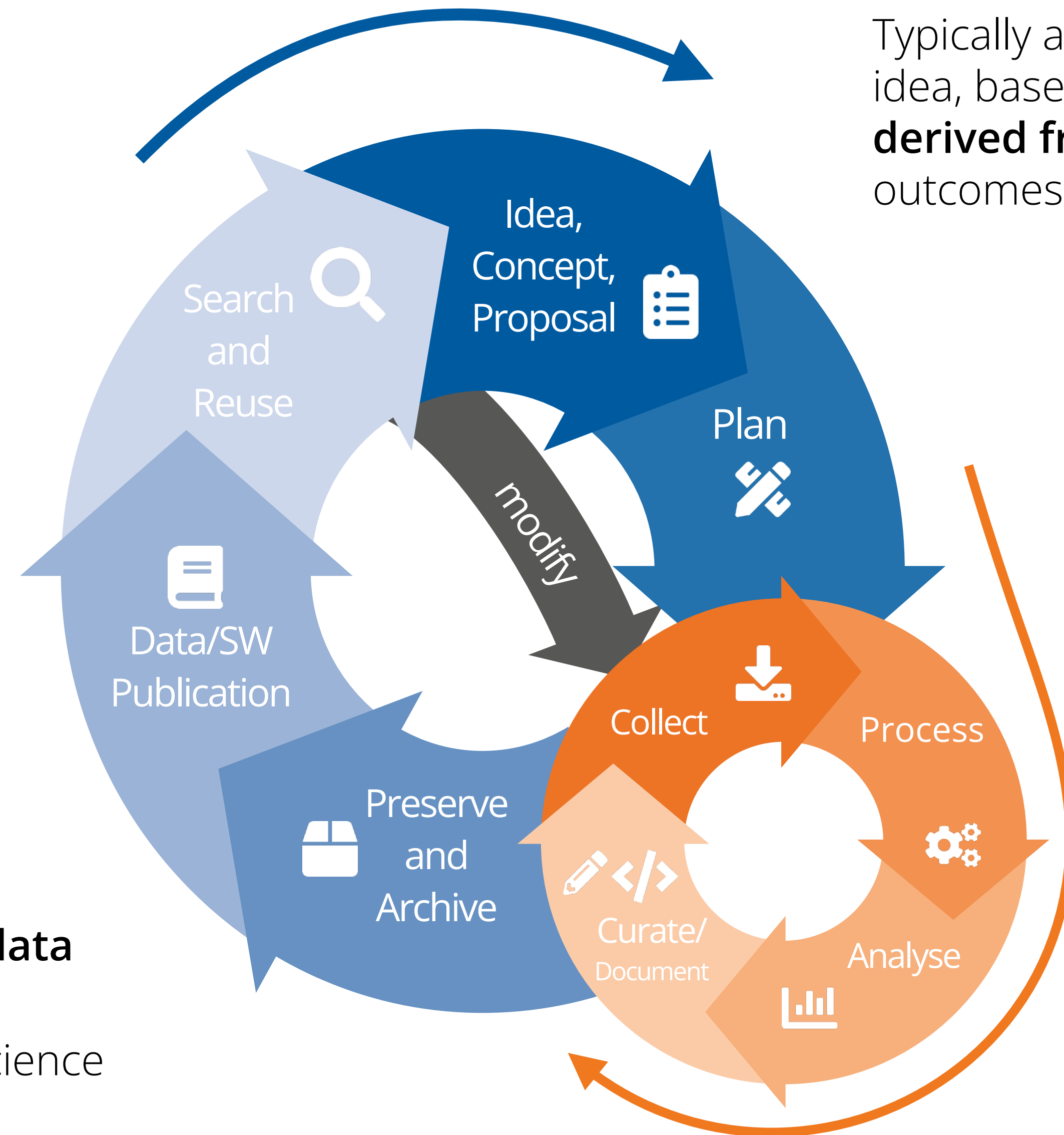
- **ATHENA** — Accelerator Technology
HEImholtz iNfrAstructure
- **CARBOSOLA**
- **Dresden High Magnetic Field Laboratory**
- **DRESDYN** — DREsden Sodium facility for DYNamo and thermohydraulic studies
- **ELBE** — Electron Linac for beams with high Brilliance and low Emittance
 - ▶ **⊕** ELBE Beamlines
- **HECToR** — High-power ultrafast computed tomography
- **HIBEF** — Helmholtz International Beamline for Extreme Fields

Data Management Lifecycle



Publication of **data** and **metadata** is at the end of a typical research lifecycle and can be the starting point for new iteration cycles...

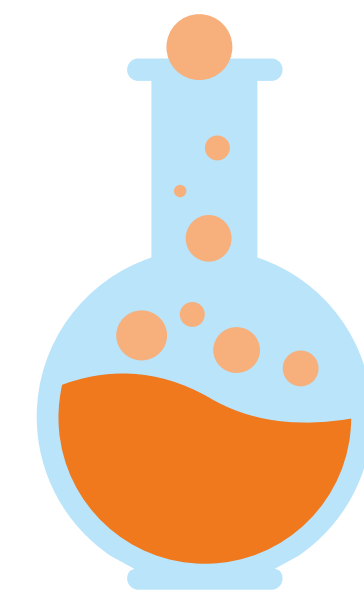
Archive or (additional) **data publication** to provide sustainable and open science



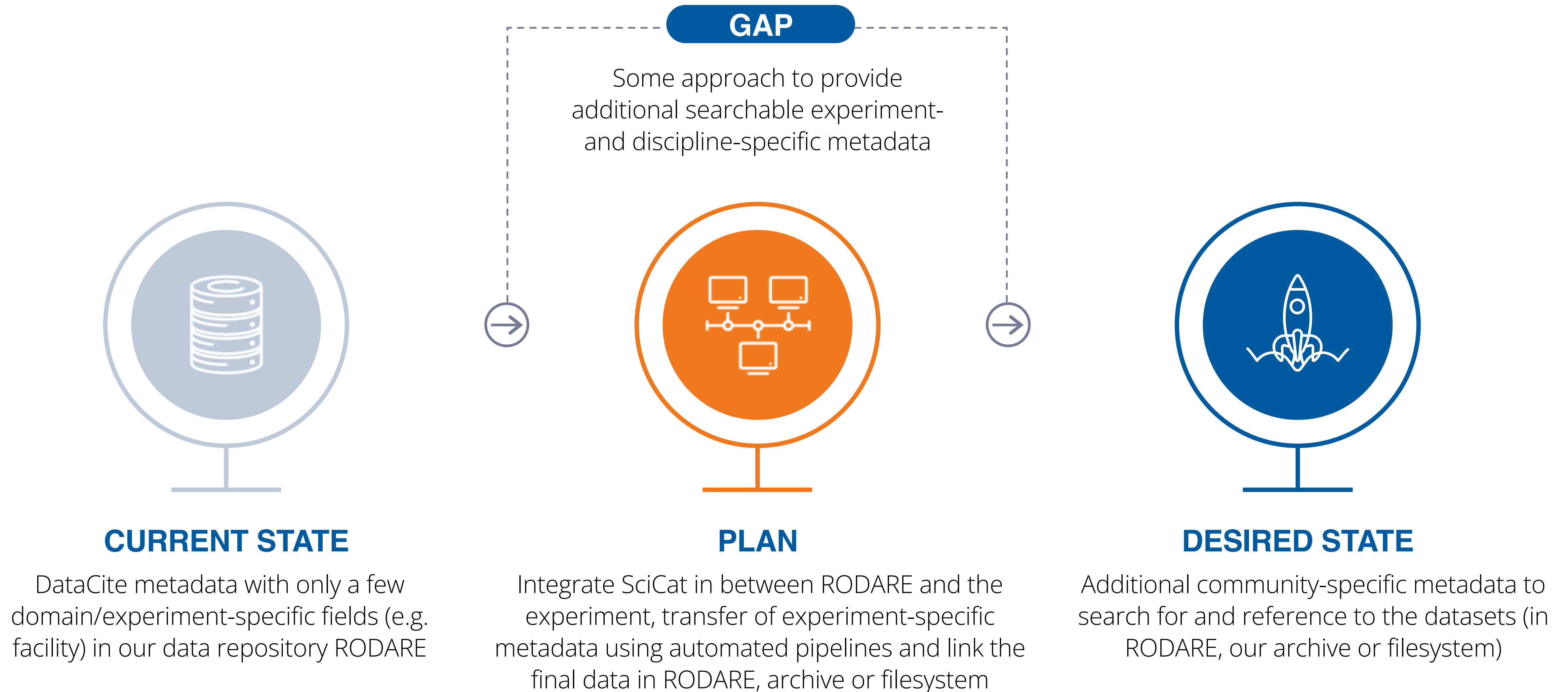
Typically an experiment starts with an idea, based on a research question, **derived from available research** outcomes

→ Valuable metadata in our proposal systems

During experiment **documentation** and **metadata** curation are essential for a **comprehensible** experiment



Gap Analysis of our Metadata Ecosystem

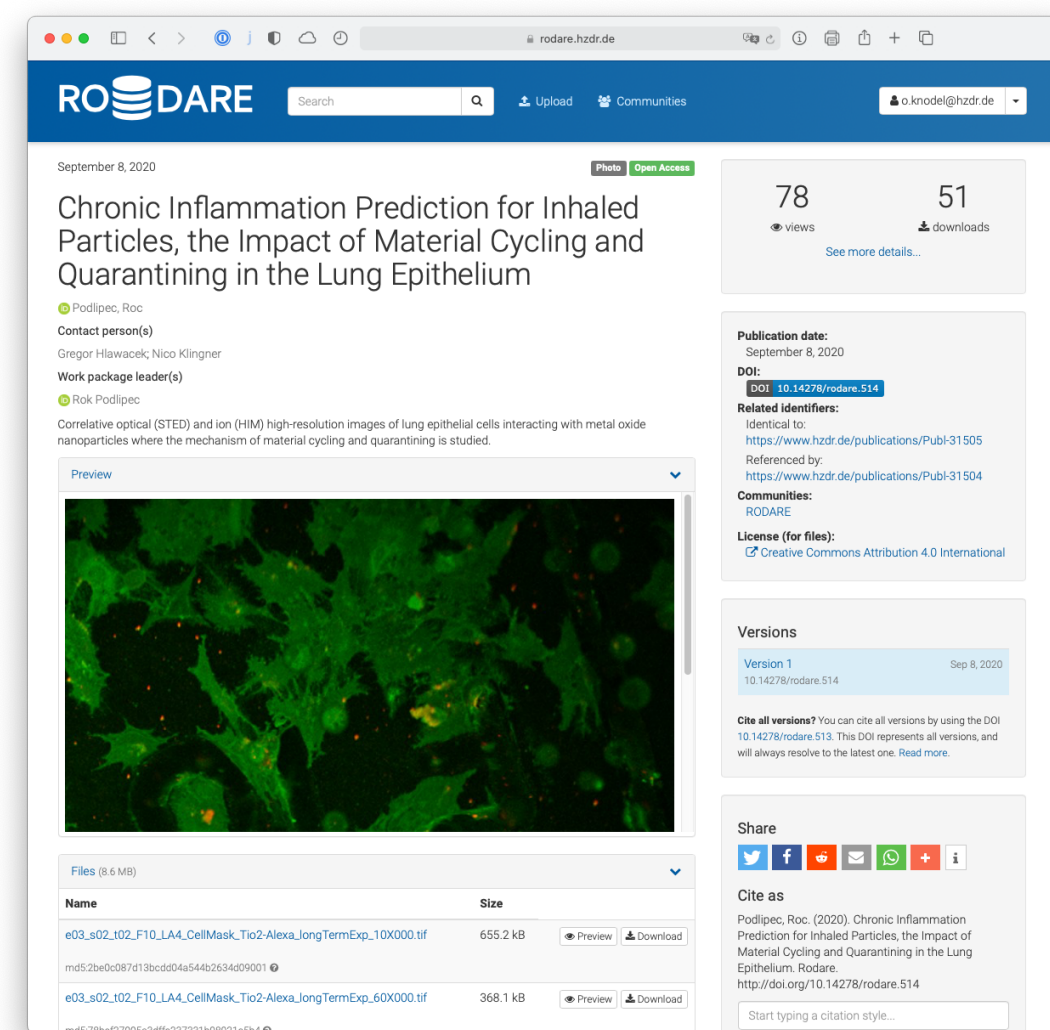
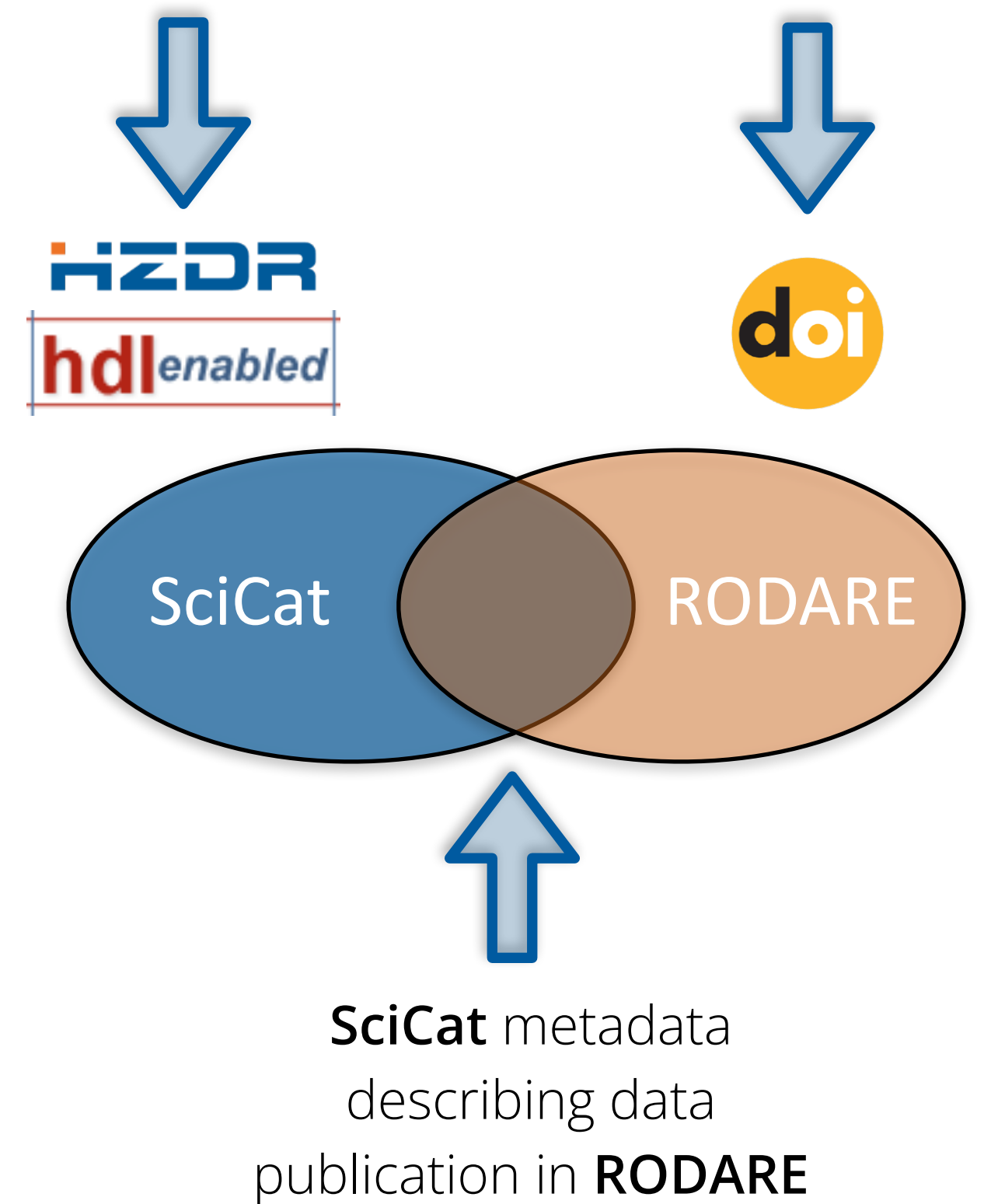


Data, Metadata and Publications at HZDR

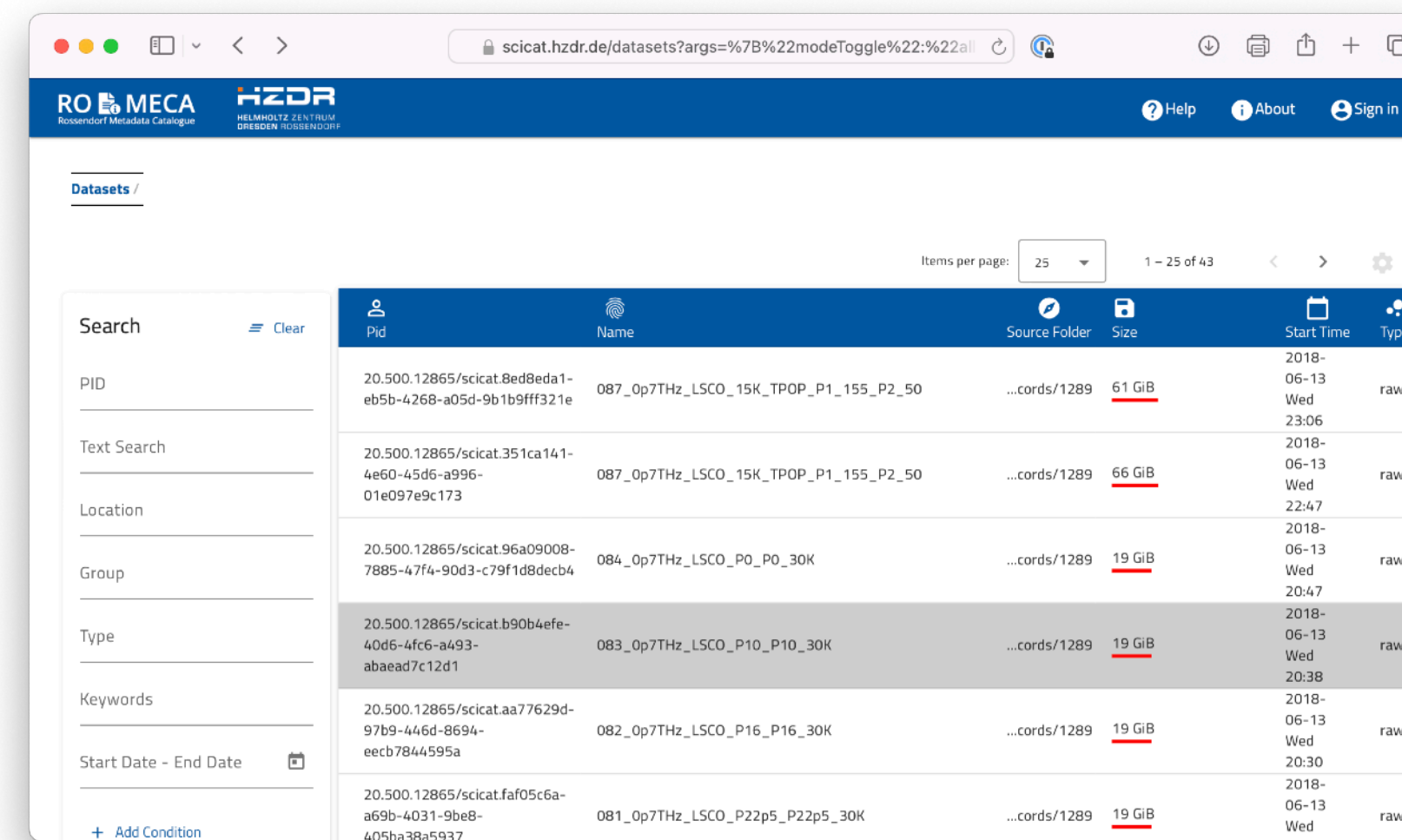
- Data Repository RODARE for citable **data/software publications** with **DOIs**
- **SciCat** for public findable metadata describing data and it's different locations with “HZDR Handles” as **PID**:
 - RODARE, if the data is published and used in scientific publications
 - Anywhere on our internal filesystems or cloud storage
 - In our internal archive for long-term storage

SciCat metadata for describing unpublished RAW or derived data located in our filesystems or archives
(external access can be granted)

Data and software publications in **RODARE**
(Open, closed, restricted or embargoed access)



Data publication system RODARE
rodare.hzdr.de



SciCat metadata catalogue
scicat.hzdr.de

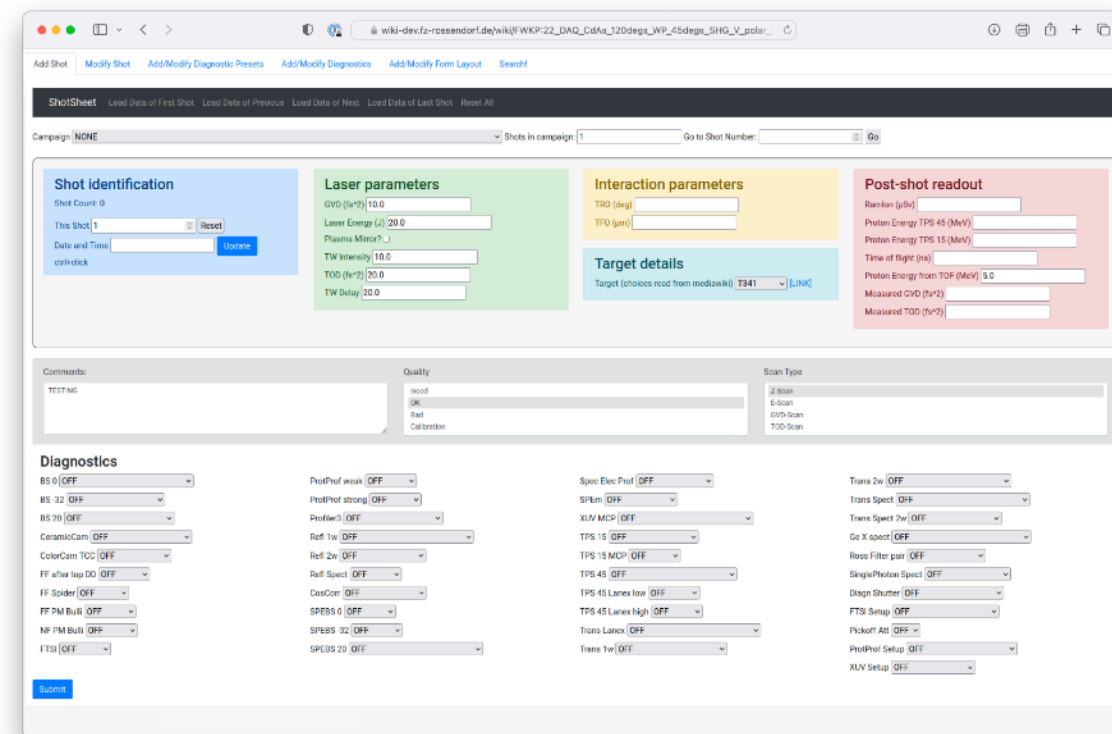
From the Experiment over the Metadata to the Publication of Data

Curated Metadata Sources

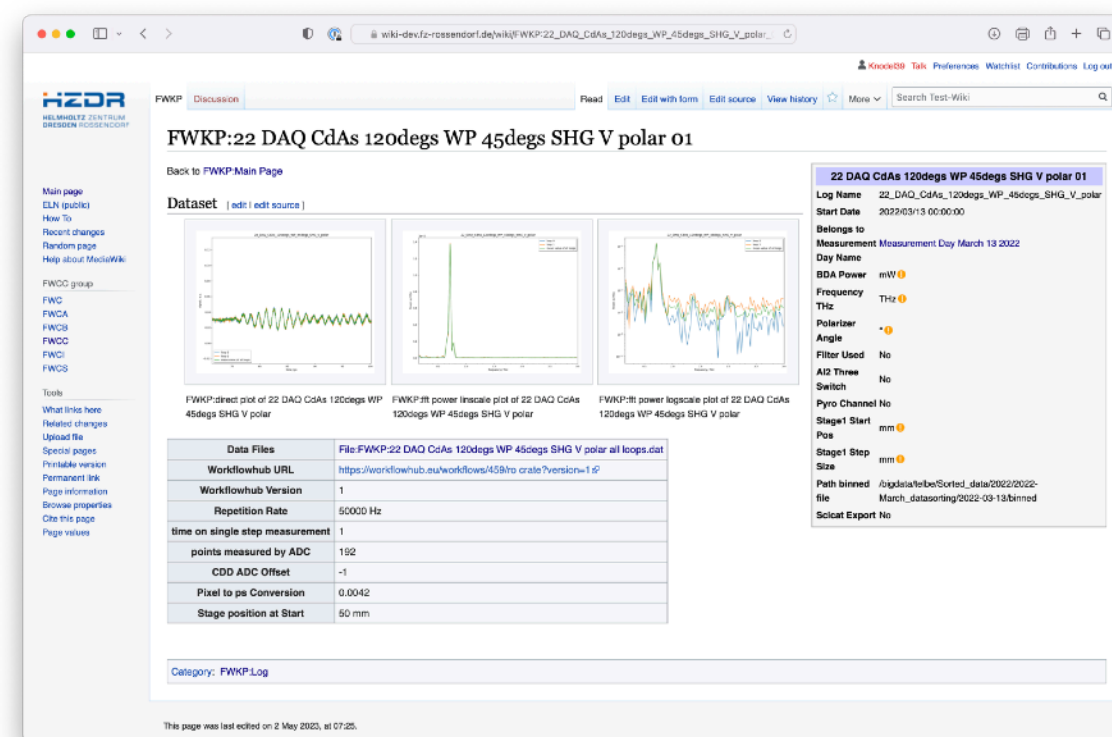
Public Metadata Catalogue

Subsequent Access to Data

ExperimentLogging app (ExL)



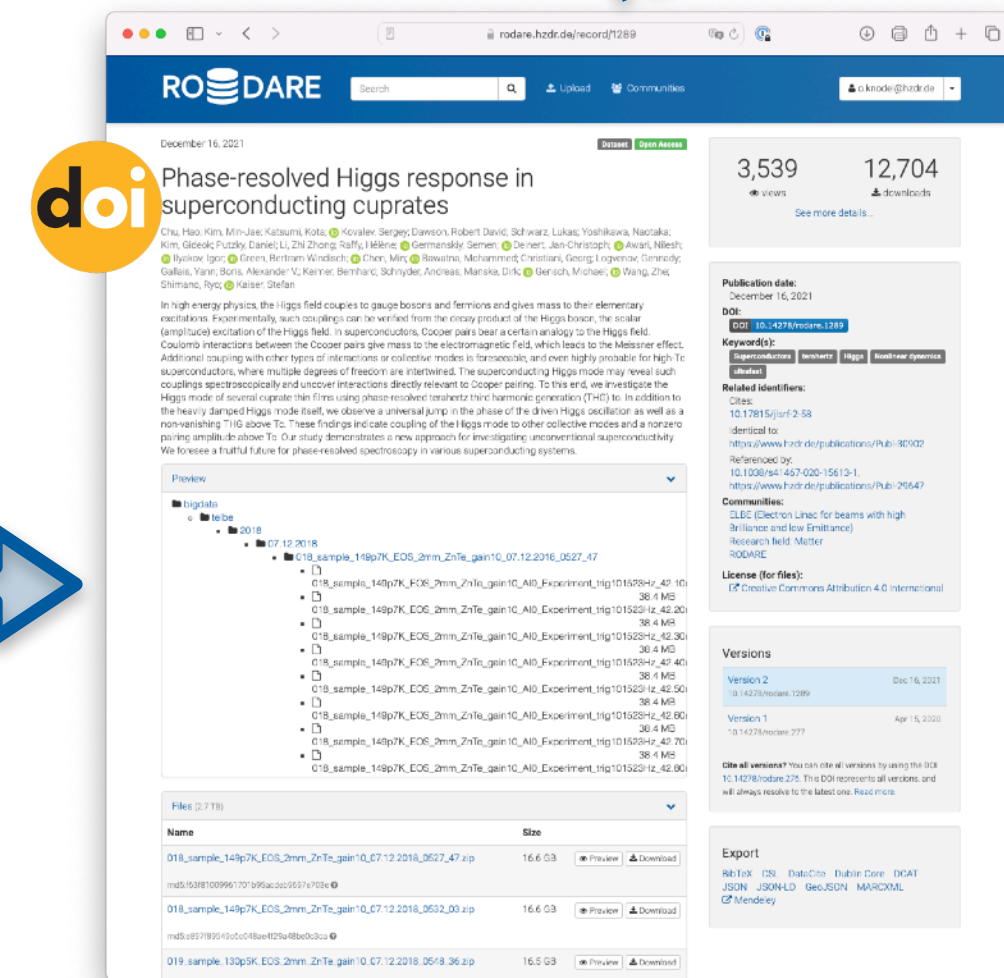
E-Logbook



SciCat



RODARE (Invenio)



Dataset



Filesystem



Tape Archive



Fully Automated Process for DRACO

Access: Private/Internal

Access: Public

Conclusions

- With RODARE, we have a data and software repository for all scientific research areas at the HZDR.
- RODARE only provides bibliographic metadata for description and search in the data sets.
- More detailed, experiment-specific, searchable metadata in SciCat can help scientists in their daily research.



Collect

Automated data
importing



Annotate

Enrich your data



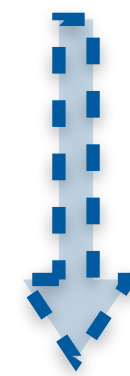
Publish

Add DOI and share



Find

Browse and search
your data



panosc
data portal

RODARE
ROSSENDORF DATA REPOSITORY



B2FIND
EUDAT

Current step: Provision of a PaNOSC search
API endpoint at the HZDR!

HZDR