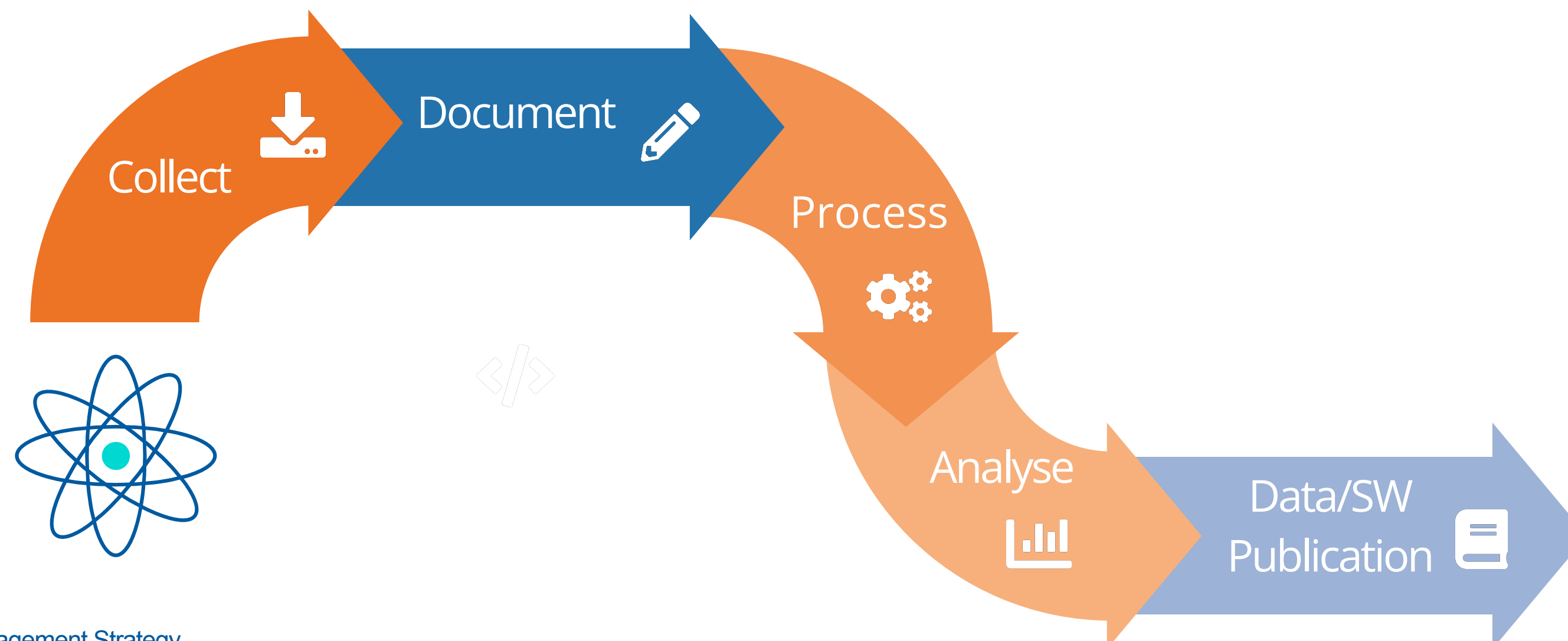# The (emerging) HZDR Data Management Strategy

Data and Software Management Days, October, 2024

Oliver Knodel // contact: o.knodel@hzdr.de
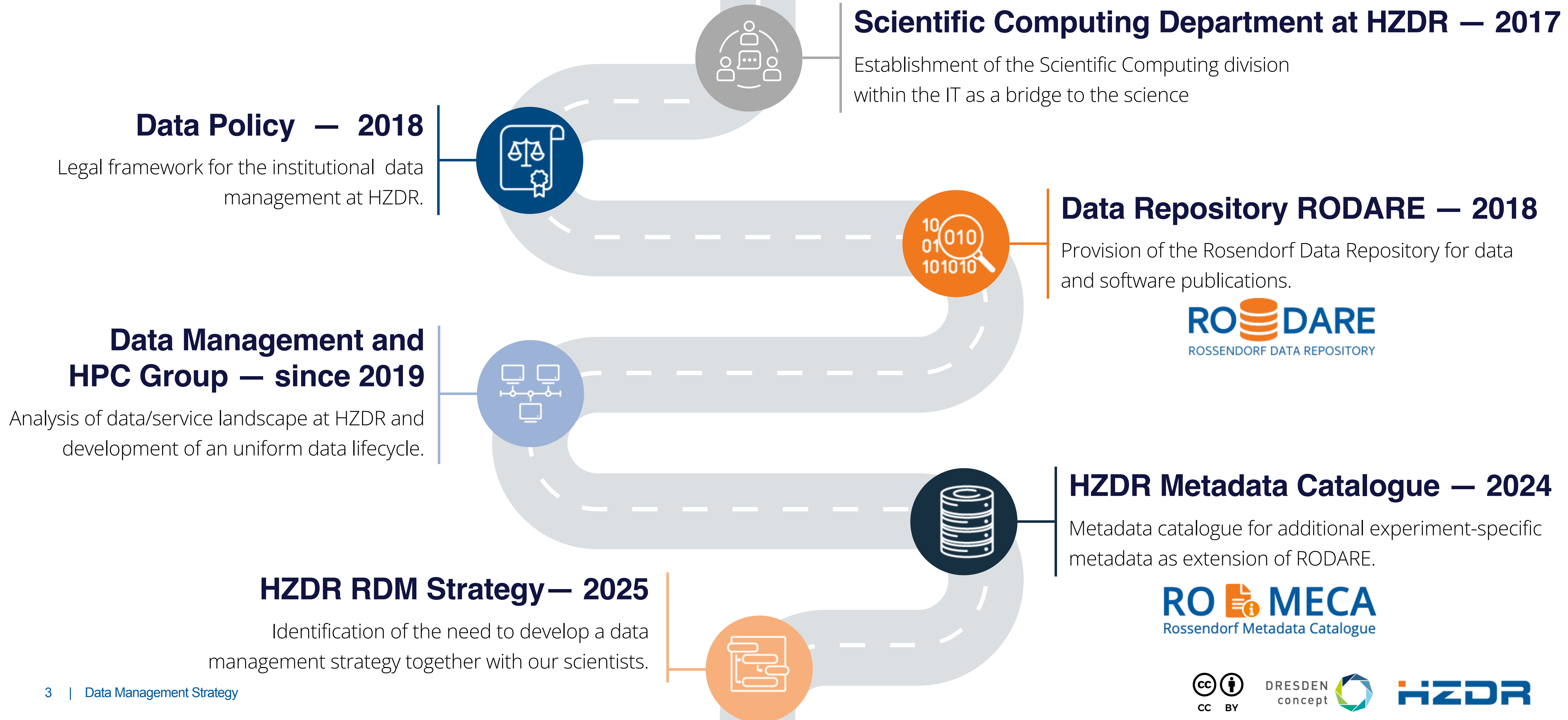
The Data Management Mission of the Computing Department

— 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.

Collect → Document → Process → Analyse → Data/SW Publication

Development of the HZDR   Open Data 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.

Data Repository RODARE — 2018

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

RODARE
ROSSENDORF DATA REPOSITORY

Data Management and HPC Group — since 2019

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

HZDR Metadata Catalogue — 2024

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

RO MECA
Rossendorf Metadata Catalogue

HZDR RDM Strategy— 2025

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

DRESDEN concept

HZDR

The Consequence of Open Science and FAIR: Institutional Data Policies

— 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!

DOI 10.14278/rodare.2269

HZDR	HZDR Data Policy	Date:	01.05.18
		Rev.:	0
	HZDR-Regulation No. B 220	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

CC BY DRESDEN concept HZDR

The Data and Software Publication Platform RODARE

Powered by:

INVENIO

Harvested via OAI-PMH by:

B2FIND   EUDAT

Registered in:

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

http://doi.org/10.17616/R3BR40
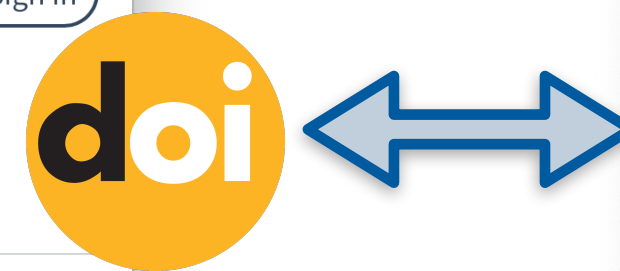
Metadata and Data

— 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**.

— In Theory: max 50 GB per file, max 100 GB per dataset (largest dataset: 7,2 TByte).

File Access:
Rest API

Filesystem

Metadata API:
Rest API, OAI-PMH
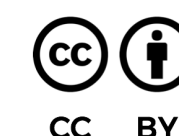
The RODARE Statistics and Worldwide Accesses

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

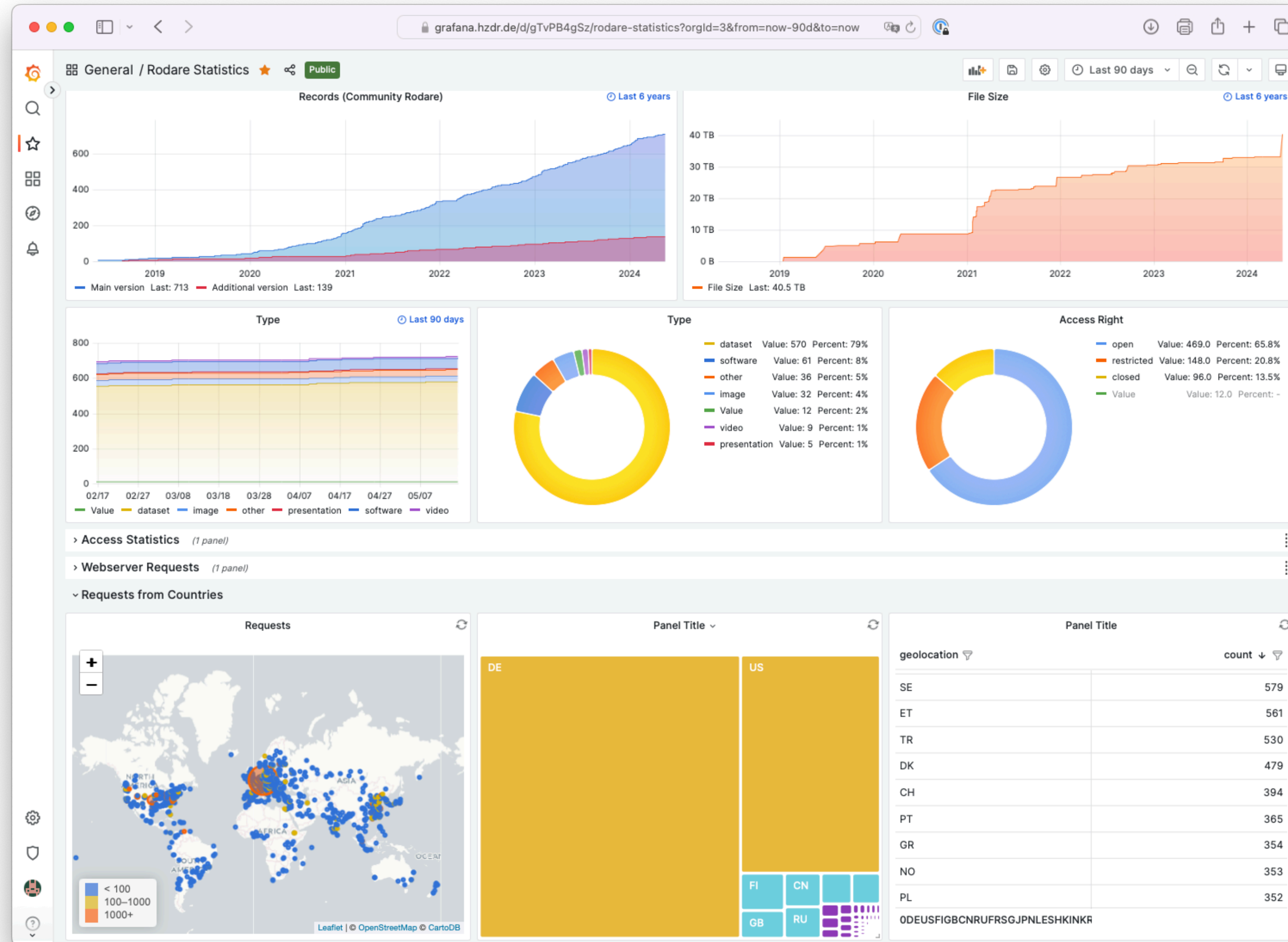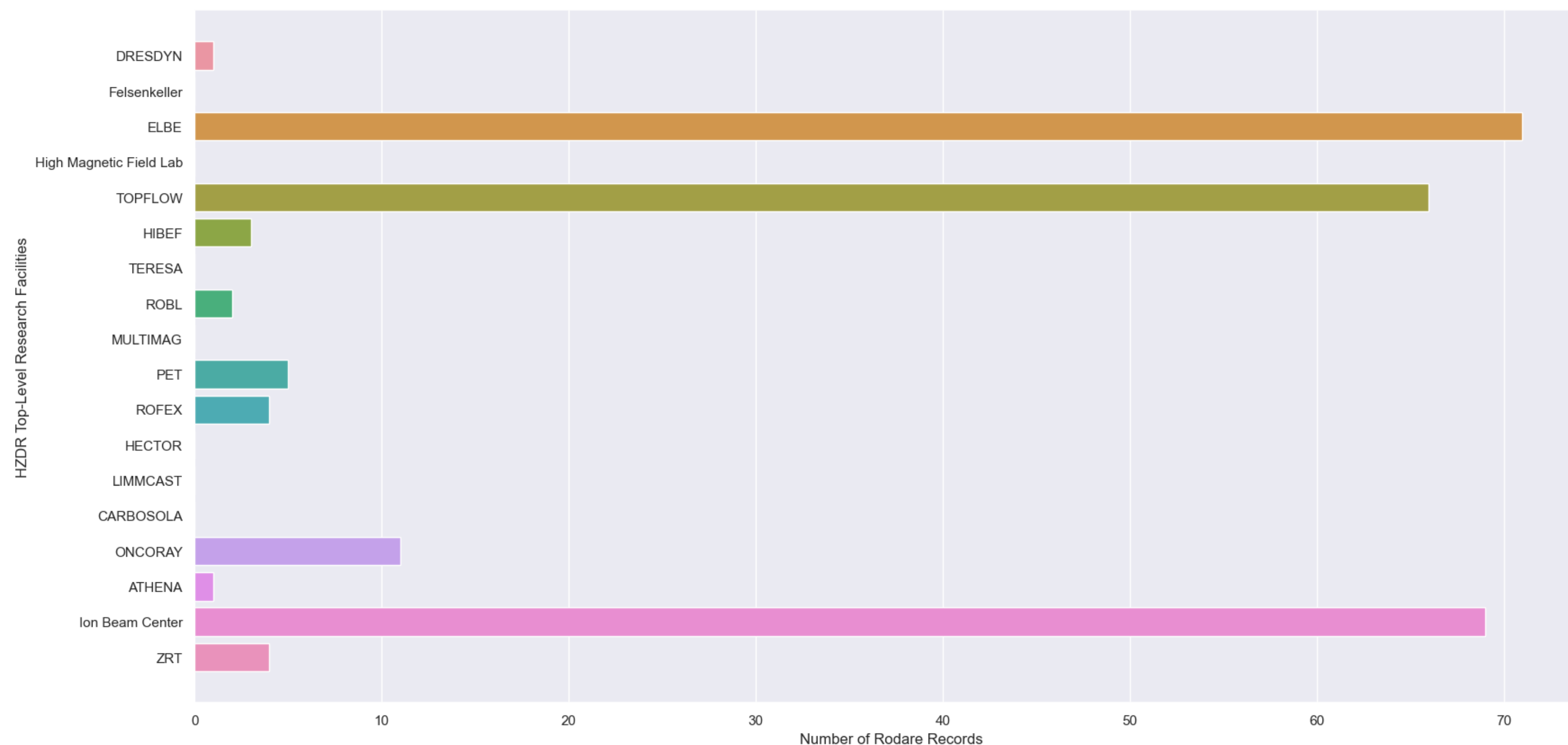In order to make our large-scale HZDR facilities more visible, Rodare now offers specific communities and an overview on our front page!

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

- **ATHENA** — Accelerator Technology 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

- **HECToR** — High-power ultrafast computed tomography

- **HIBEF** — Helmholtz International Beamline for ~~me Fields~~

The Research Data Management Landscape at HZDR

— Our IT infrastructures can support various experiments, but they are complex...

— Scientists often don't know which services are available and how to use them.

— Offers to our scientists to guide them through the services:

  · IT-Service catalogue: hzdr.de/fdm-services

  · HELIPORT: heliport.hzdr.de
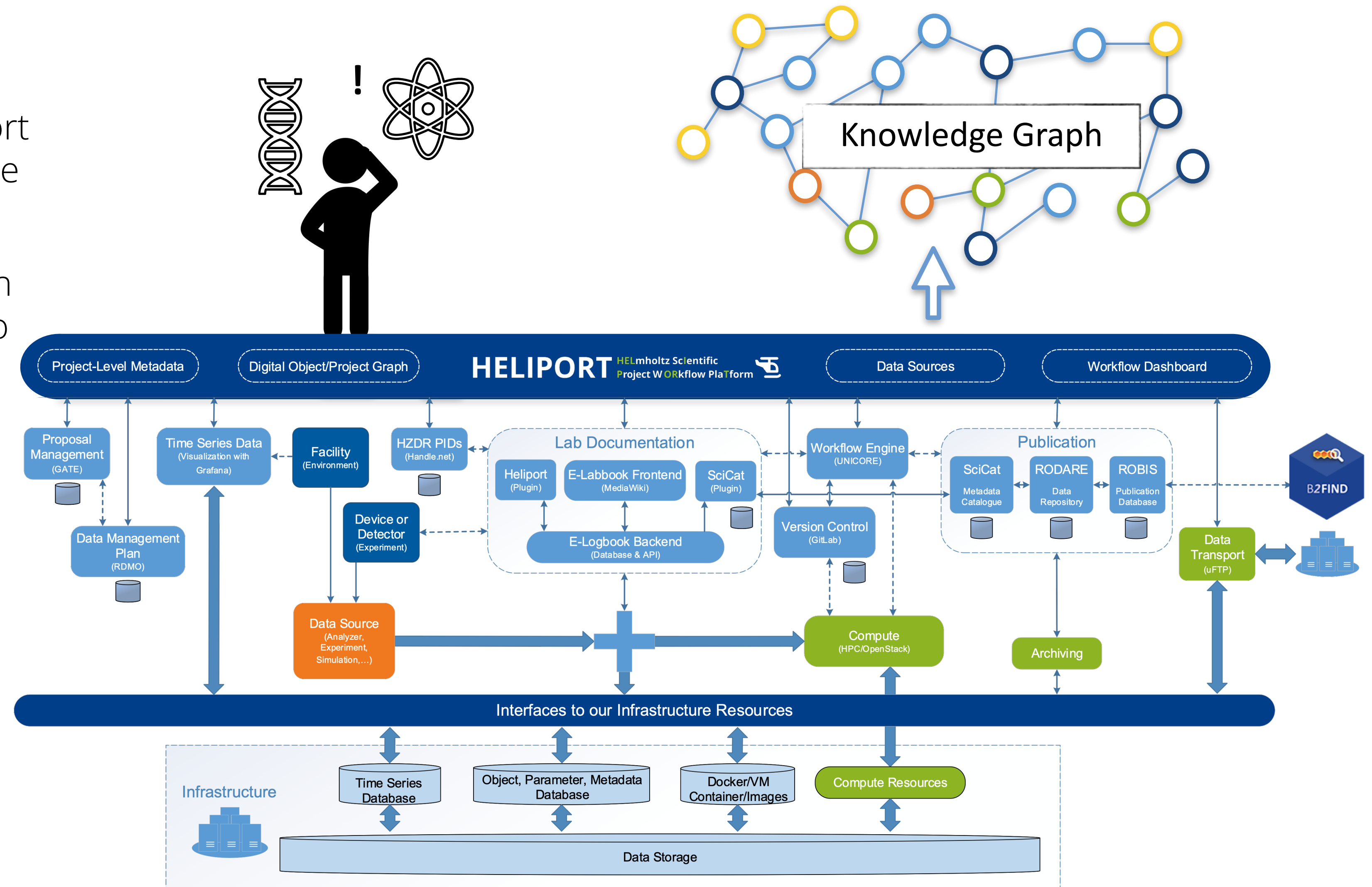
Gap Analysis of our Metadata Ecosystem

GAP

Some approach to provide additional searchable experiment- and discipline-specific metadata

CURRENT STATE

DataCite metadata with only a few domain/experiment-specific fields (e.g. facility) in our data repository RODARE

PLAN

Integrate SciCat in between RODARE and the experiment, transfer of experiment-specific metadata using automated pipelines and link the final data in RODARE, archive or filesystem

DESIRED STATE

Additional community-specific metadata to search for and reference to the datasets (in RODARE, our archive or filesystem)
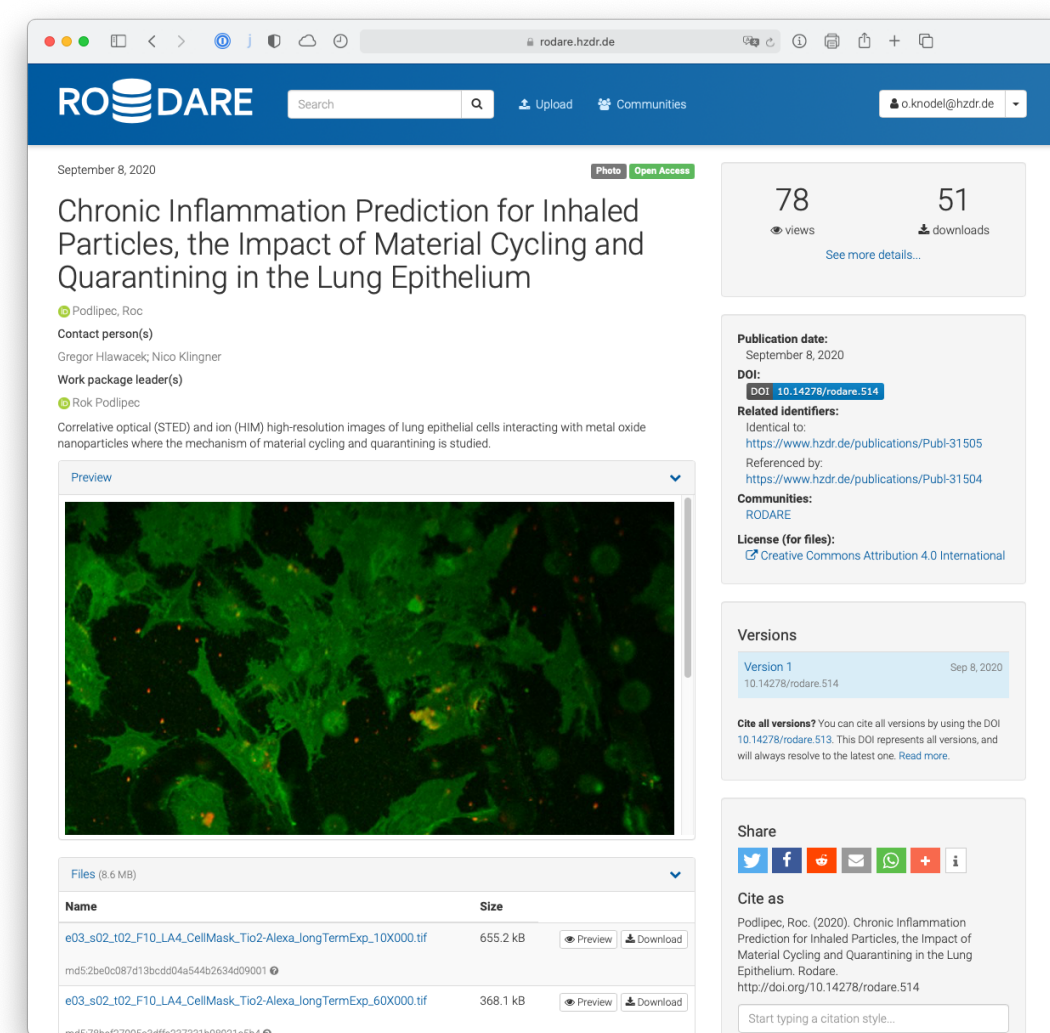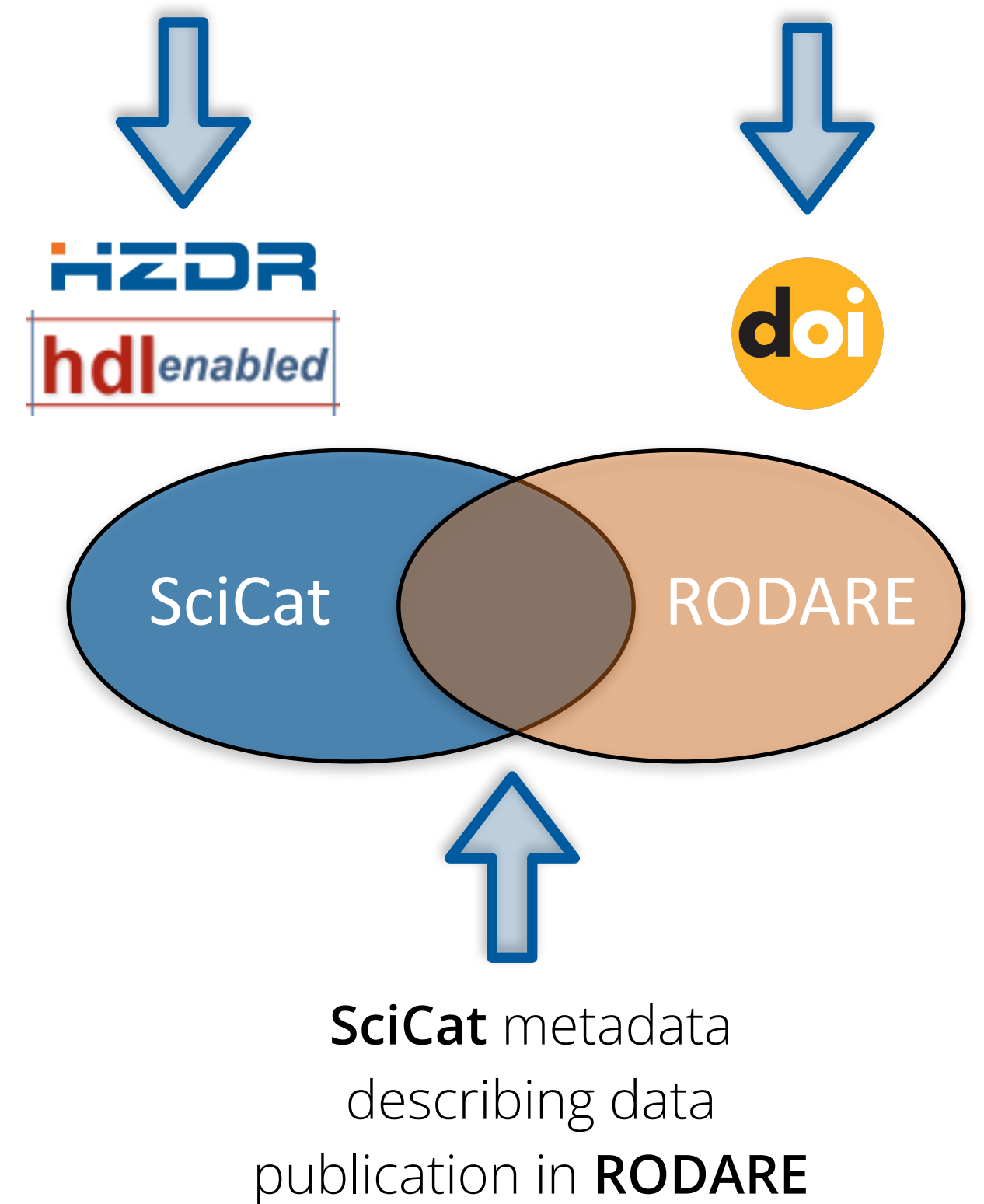
DRESDEN concept

HZDR

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

Data publication system RODARE
rodare.hzdr.de

SciCat metadata catalogue
scicat.hzdr.de

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)

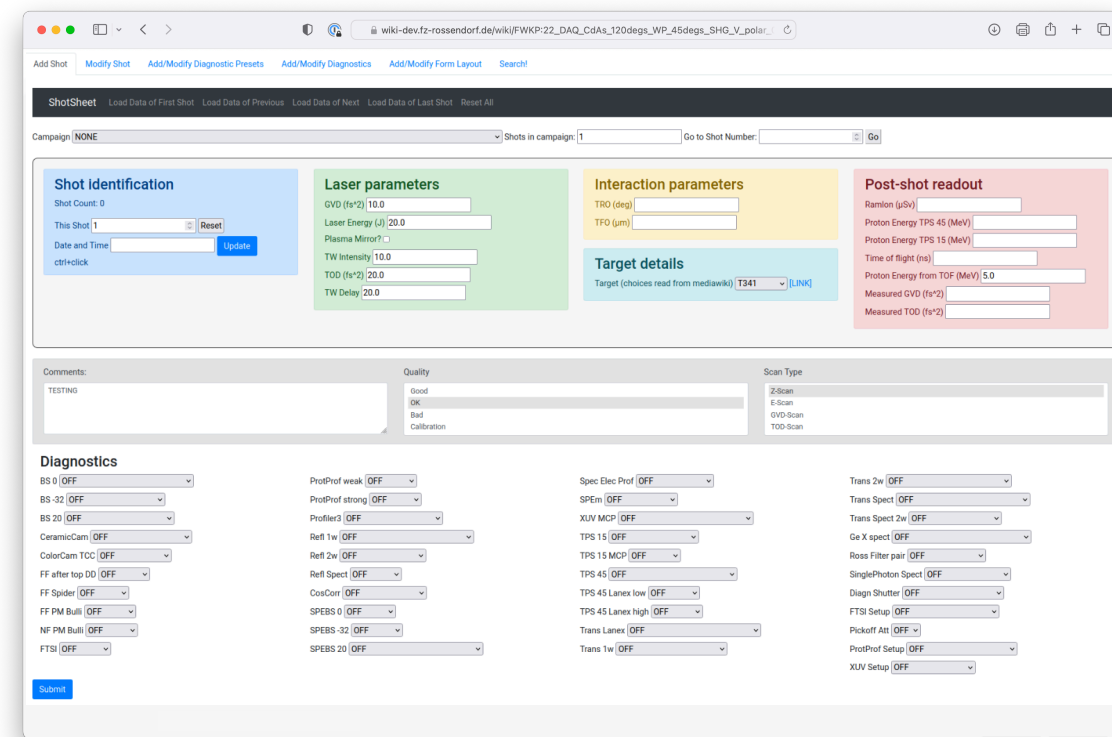**SciCat** metadata describing data publication in **RODARE**

From the Experiment over the Metadata to the Publication of Data

Curated Metadata Sources

Public Metadata Catalogue

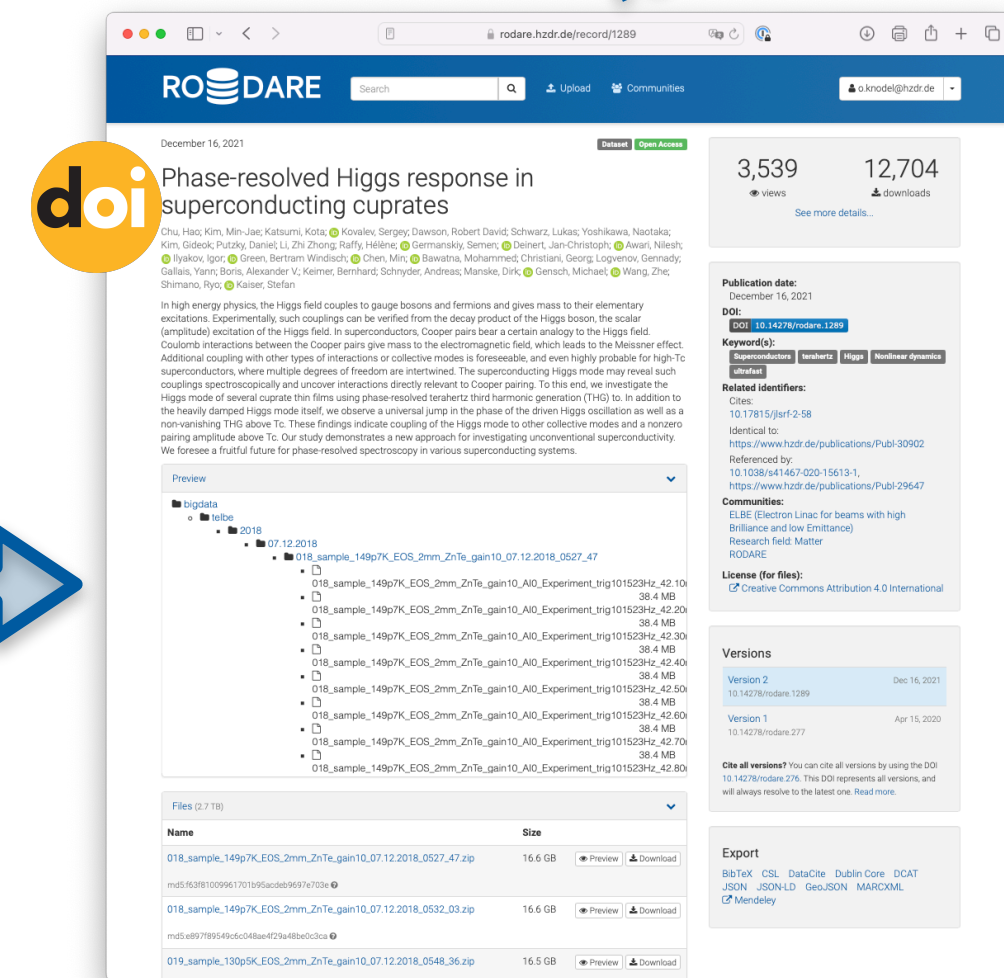## Subsequent Access to Data

ExperimentLogging app (ExL)

E-Logbook

SciCat

panosc data portal

RODARE (Invenio)

B2FIND EUDAT
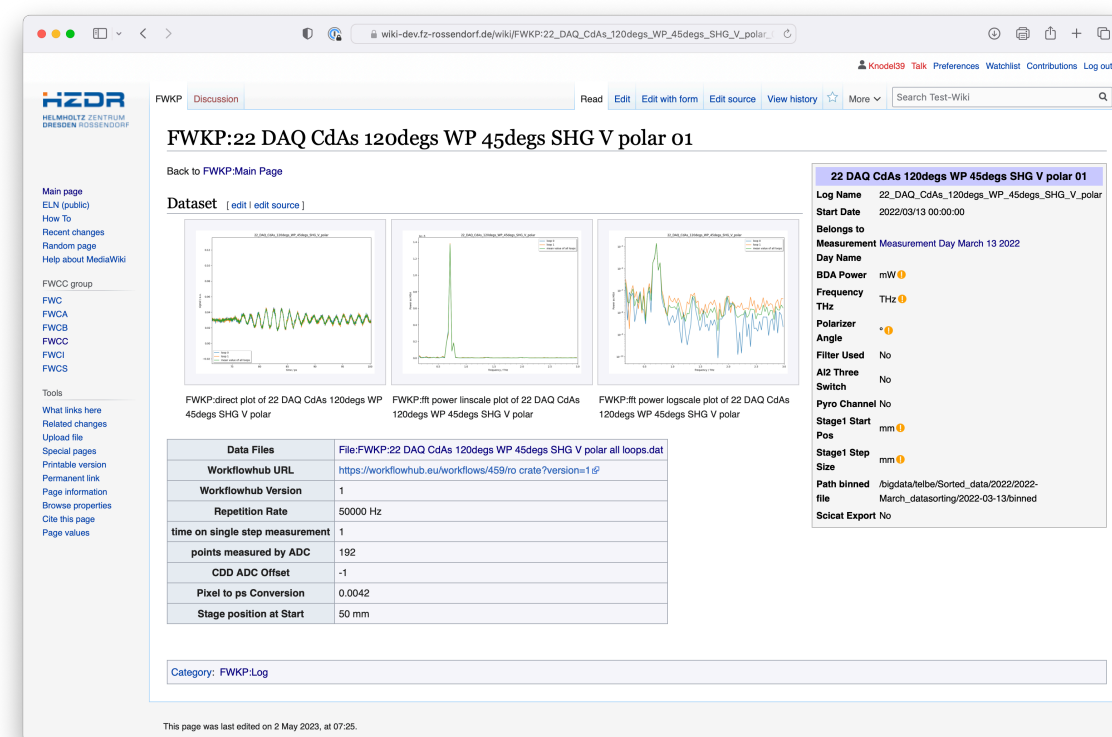
Dataset

Filesystem

Tape Archive

HELIPORT
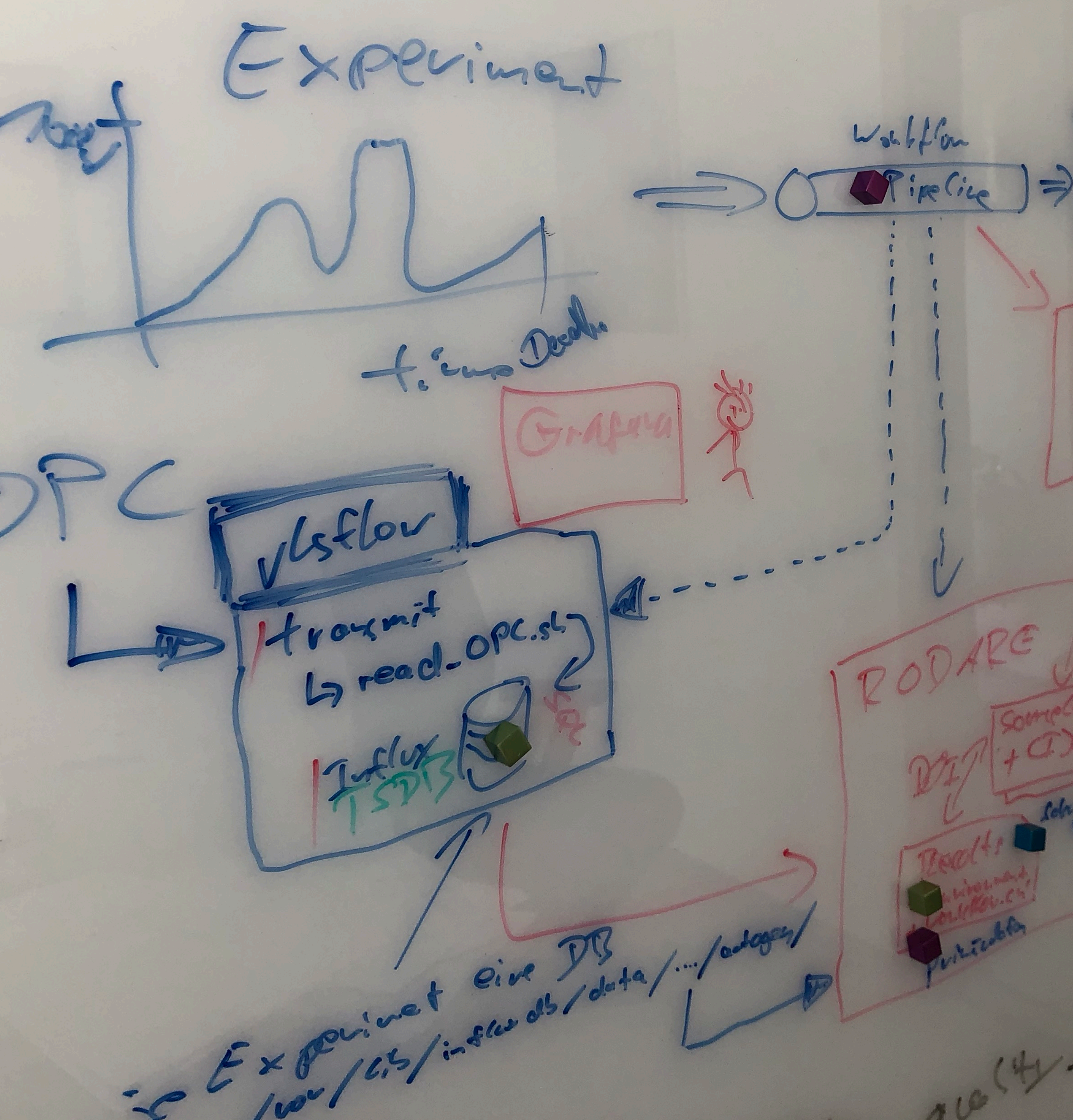
Metadata from Experiment/Simulation
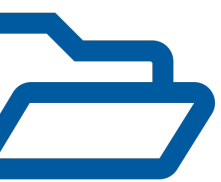
Fully Automated Process for DRACO

**Access: Private/Internal**

Access: Public

Why a
HZDR Data Management Strategy
is necessary

# Motivation for the Development of a HZDR Data Management Strategy

— We have a complex ecosystem of tools and services at the HZDR, within HiFiS and beyond: NFDI and EOSC.

— A clear and transparent roadmap for our developments in coordination with our interred and external stakeholders is necessary.

— The HZDR scientists, large-scale instruments and experiments must always be at the top of the agenda.
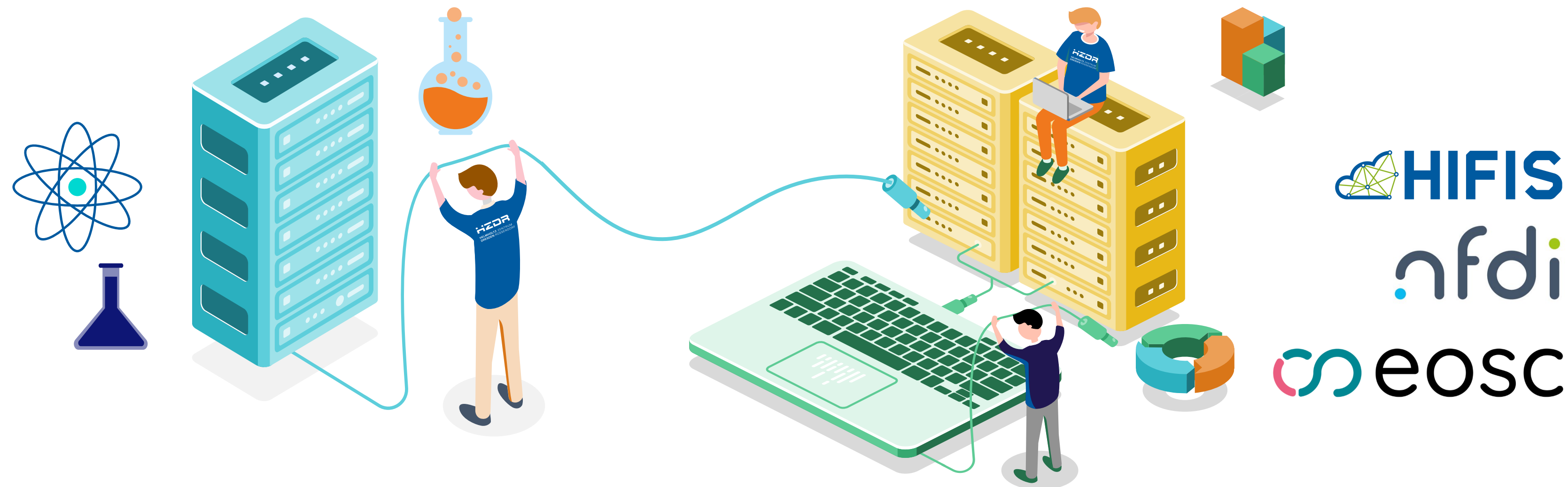
➜ **We want to work with you to develop a clear and transparent roadmap to define priorities, allocate resources and obtain appropriate funding.**

Data Management Strategy

— A document that describes the HZDR data management,

— integrated into our guidelines and policies,

— including future developments based on

— a clear roadmap and subproject,

— with appropriate prioritisation and

— embedded in a larger context beyond the HZDR.
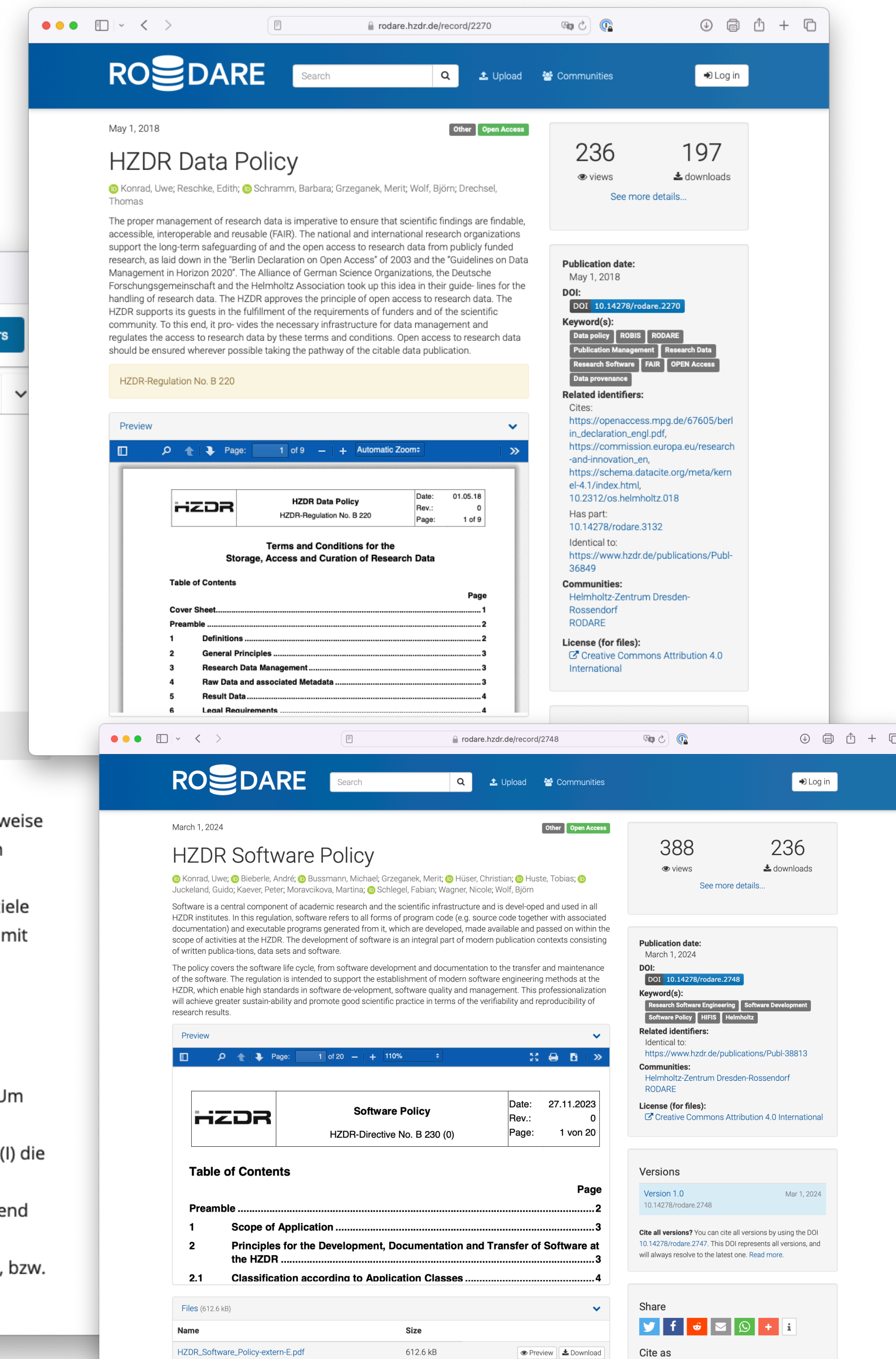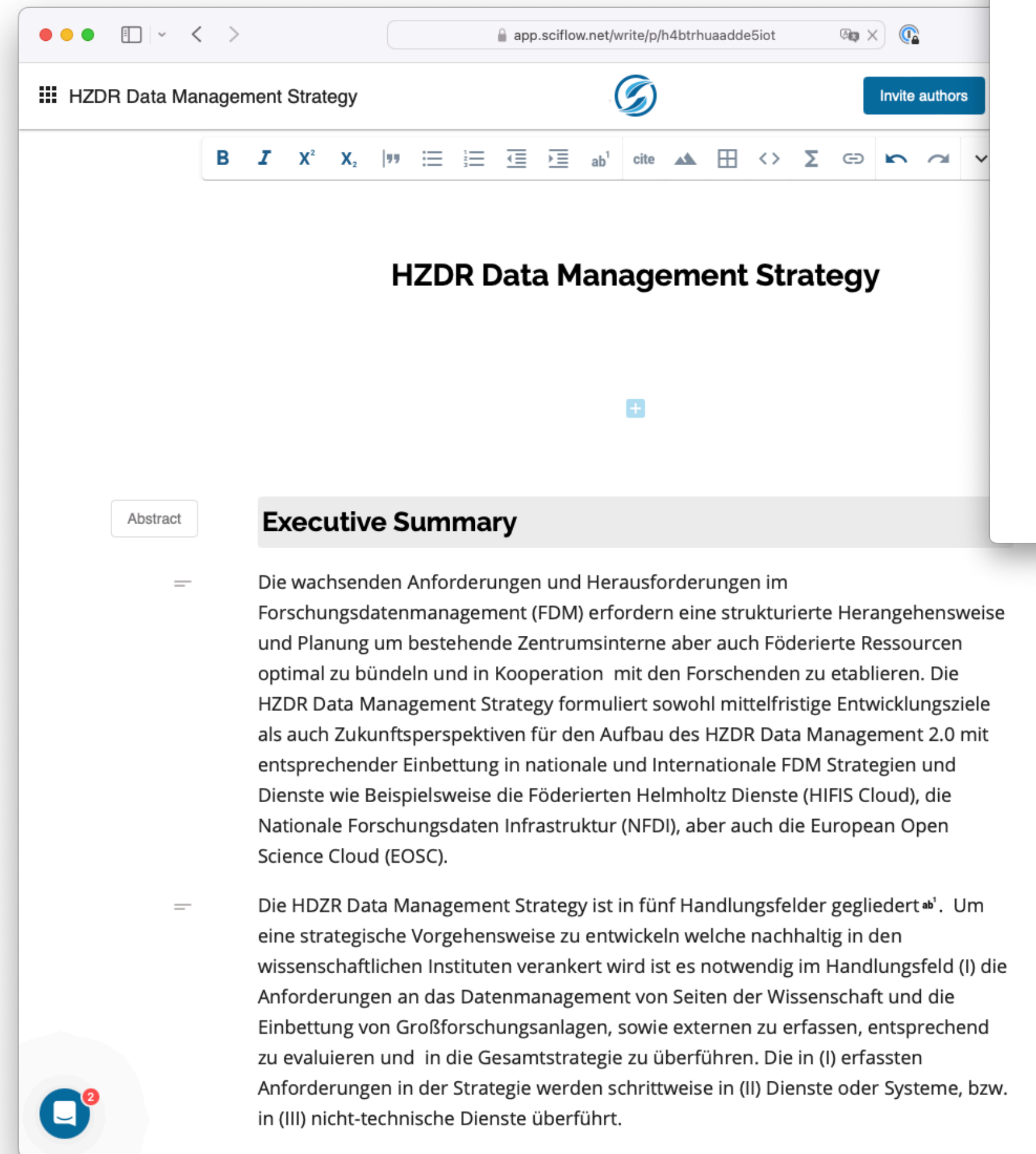
Conclusions

— With the workshops today and tomorrow we want to initially discuss the requirements of our scientists and experiments.

— We want to finalise a first draft by the end of the year.

— Together with our scientists and stakeholders, we will iteratively transform the draft into a comprehensive document.