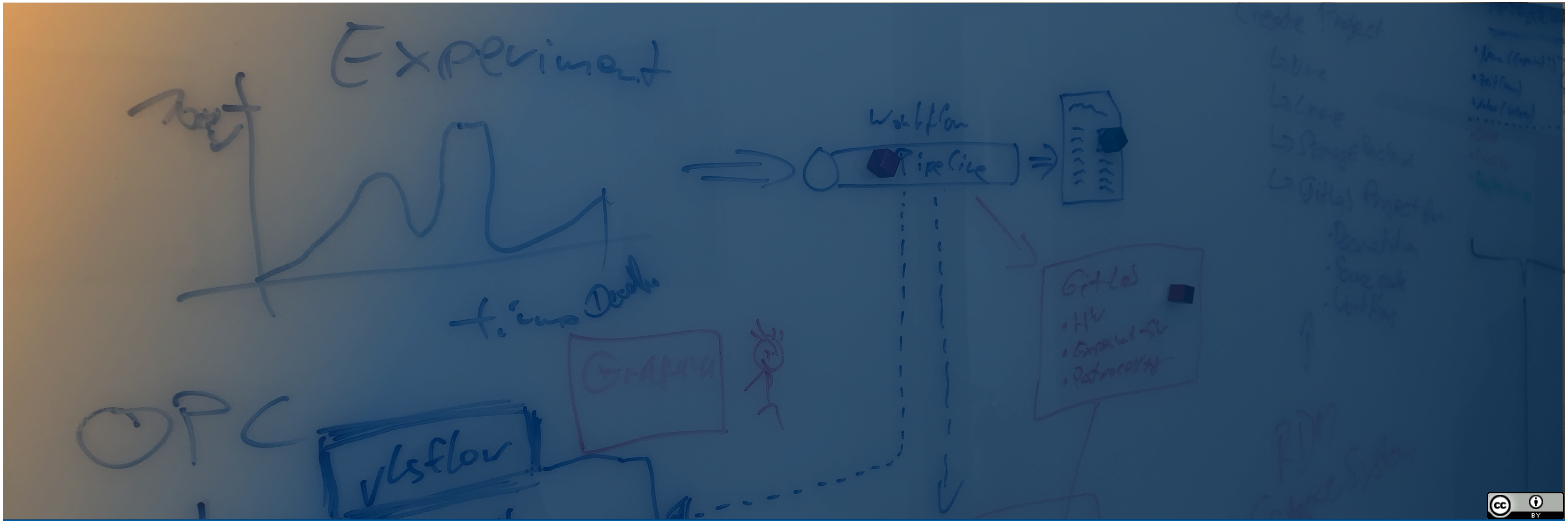


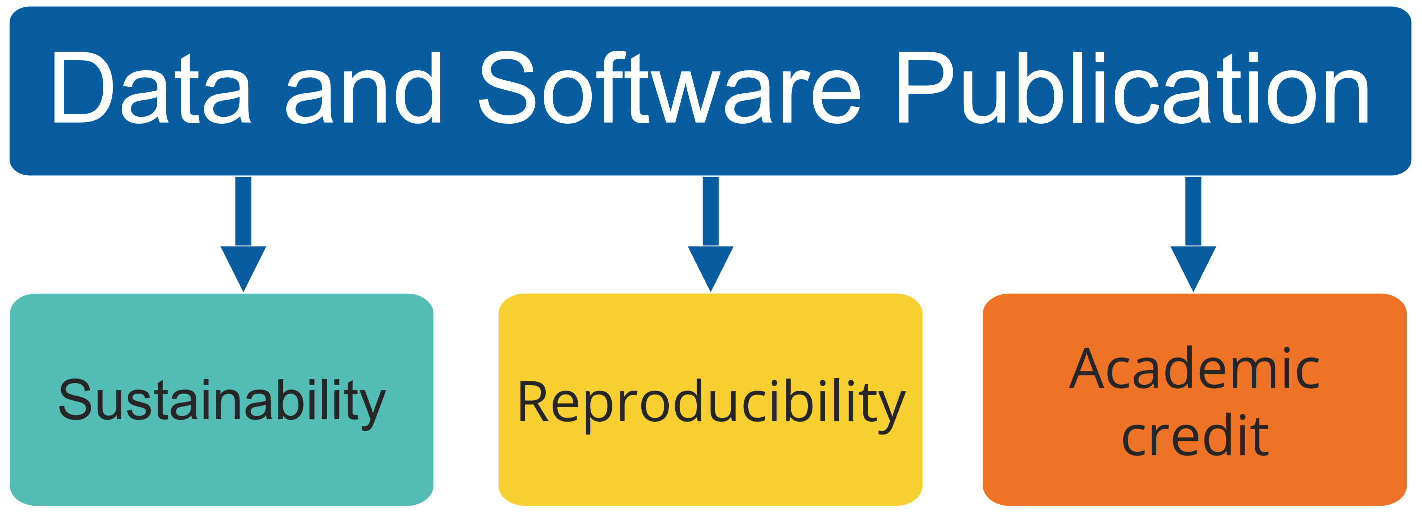
Towards a Seamlessly Interlinked Research Data and Software Ecosystem at HZDR

2nd Practice Forum Research Data Management, October 20, 2022

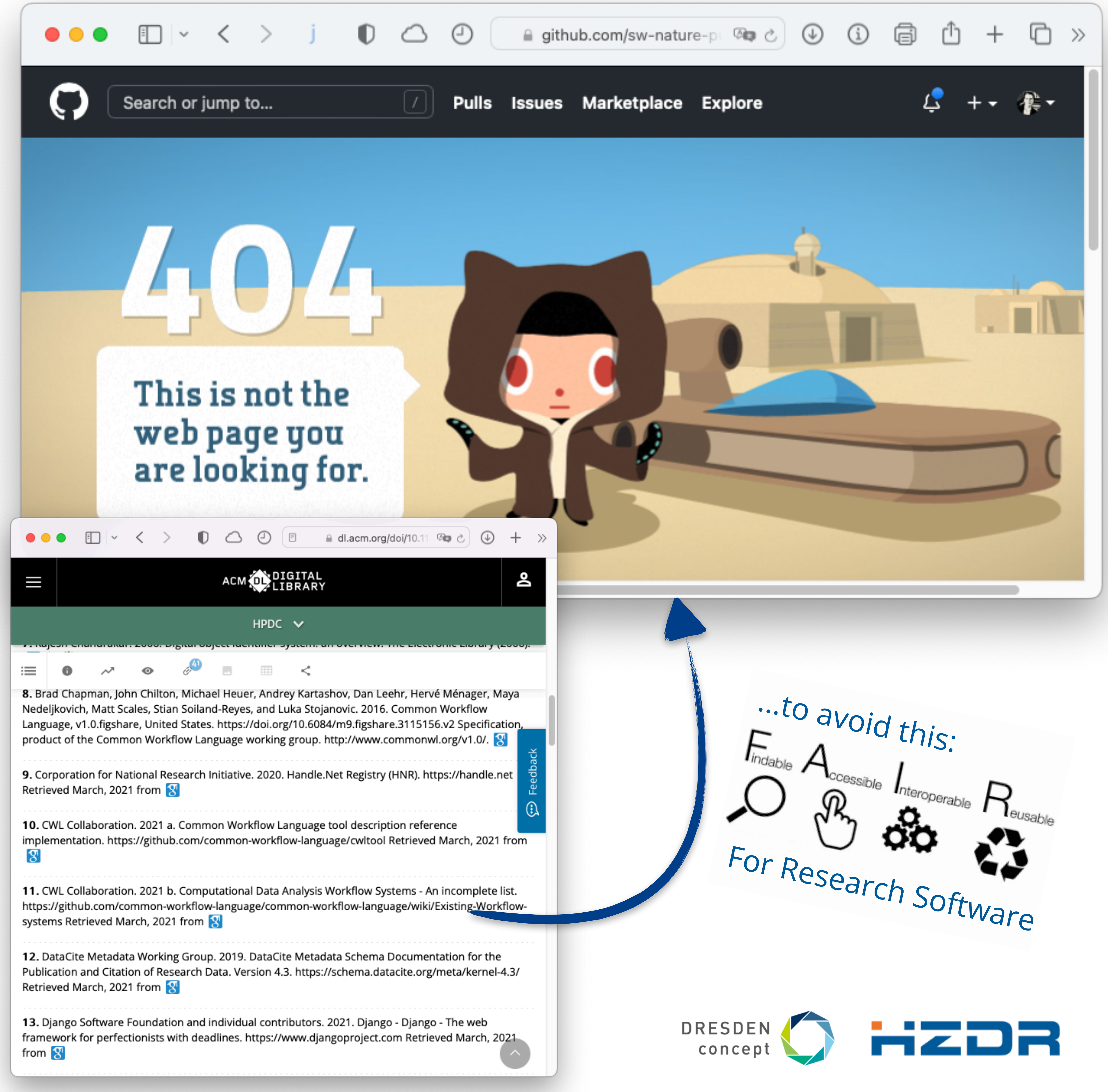
Oliver Knodel // contact: o.knodel@hzdr.de

Motivation — Software, Data and Everything in Between

- Data and software are an important result of a scientific experiment.
- Scientific publication, research software and data must receive the same academic credit:



- FAIR principles also exist for research software and should be taken into account [1].
- In addition to the publication itself a seamless interlinking between all available data products is also necessary to improve findability.



8. Brad Chapman, John Chilton, Michael Heuer, Andrey Kartashov, Dan Leehr, Hervé Ménager, Maya Nedeljkovich, Matt Scales, Stian Soiland-Reyes, and Luka Stojanovic. 2016. Common Workflow Language, v1.0.figshare, United States. https://doi.org/10.6084/m9.figshare.3115156.v2 Specification, product of the Common Workflow Language working group. http://www.commonwl.org/v1.0/.

9. Corporation for National Research Initiative. 2020. Handle.Net Registry (HNR). https://handle.net Retrieved March, 2021 from

10. CWL Collaboration. 2021 a. Common Workflow Language tool description reference implementation. https://github.com/common-workflow-language/cwltool Retrieved March, 2021 from

11. CWL Collaboration. 2021 b. Computational Data Analysis Workflow Systems - An incomplete list. https://github.com/common-workflow-language/common-workflow-language/wiki/Existing-Workflow-systems Retrieved March, 2021 from

12. DataCite Metadata Working Group. 2019. DataCite Metadata Schema Documentation for the Publication and Citation of Research Data. Version 4.3. https://schema.datacite.org/meta/kernel-4.3/ Retrieved March, 2021 from

13. Django Software Foundation and individual contributors. 2021. Django - Django - The web framework for perfectionists with deadlines. https://www.djangoproject.com Retrieved March, 2021 from

[1] Barker, M., Chue Hong, N.P., Katz, D.S. *et al.* Introducing the FAIR Principles for research software. *Sci Data* **9**, 622 (2022). doi.org/10.1038/s41597-022-01710-x

The Git Repository is not a Publication…

Software Repository

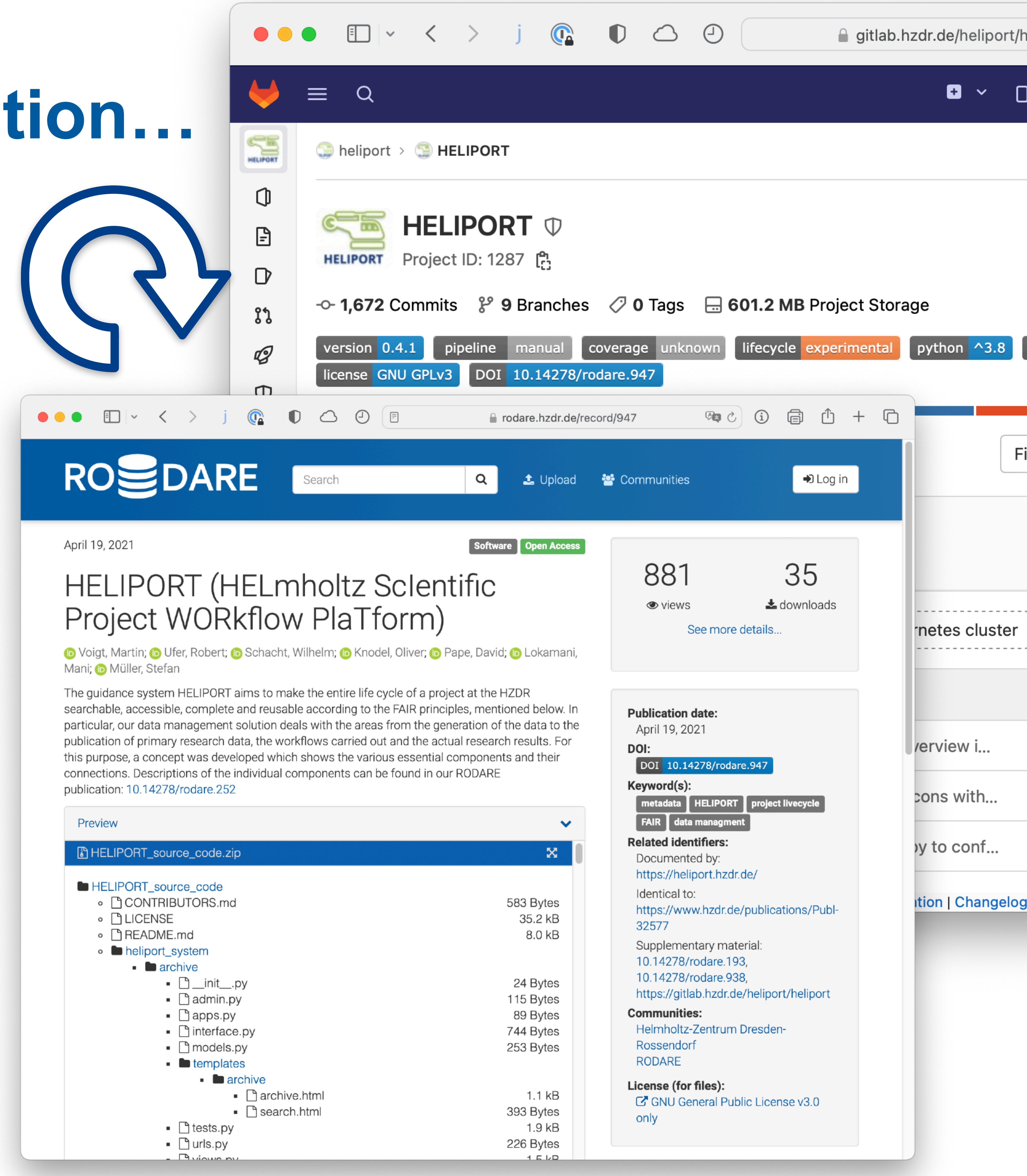
- Software is typically available (not *published*) in version control systems with open or restricted access:
- We need workflows or methods to publish software and data to ensure long-term *availability* and to meet the FAIR principles.

Software Publication

- Software must be cited in a similar way to scientific publications.
- Common data repositories (e.g. institutional, domain-specific, Zenodo) support typically the publication type *software*.
- With an additional *software publication* we can cite specific versions of a software including rich metadata:
  - Title, authors (including ORCIDs), Abstract, license, …
  - Related Identifiers to link additional resources.
  - Typically it would not be practical to link *all* scientific datasets produced or analyzed by the software to the software publication.

Repository Structure

- Software and date mixen in one repository,
- Separation between data and software repositories,
- Something in between…



gitlab.hzdr.de/heliport/h

heliport › HELIPORT

HELIPORT

Project ID: 1287

1,672 Commits · 9 Branches · 0 Tags · 601.2 MB Project Storage

version 0.4.1 · pipeline manual · coverage unknown · lifecycle experimental · python ^3.8 · license GNU GPLv3 · DOI 10.14278/rodare.947

rodare.hzdr.de/record/947

April 19, 2021

Software · Open Access

HELIPORT (HELmholtz ScIentific Project WORkflow PlaTform)

Voigt, Martin; Ufer, Robert; Schacht, Wilhelm; Knodel, Oliver; Pape, David; Lokamani, Mani; Müller, Stefan

The guidance system HELIPORT aims to make the entire life cycle of a project at the HZDR searchable, accessible, complete and reusable according to the FAIR principles, mentioned below. In particular, our data management solution deals with the areas from the generation of the data to the publication of primary research data, the workflows carried out and the actual research results. For this purpose, a concept was developed which shows the various essential components and their connections. Descriptions of the individual components can be found in our RODARE publication: 10.14278/rodare.252

Preview

HELIPORT_source_code.zip

```
HELIPORT_source_code
  CONTRIBUTORS.md                 583 Bytes
  LICENSE                          35.2 kB
  README.md                         8.0 kB
  heliport_system
    archive
      __init__.py                  24 Bytes
      admin.py                    115 Bytes
      apps.py                      89 Bytes
      interface.py                744 Bytes
      models.py                   253 Bytes
      templates
        archive
          archive.html              1.1 kB
          search.html             393 Bytes
      tests.py                      1.9 kB
      urls.py                     226 Bytes
```

881 views · 35 downloads

See more details…

Publication date: April 19, 2021

DOI: DOI 10.14278/rodare.947

Keyword(s): metadata · HELIPORT · project livecycle · FAIR · data managment

Related identifiers:

Documented by: https://heliport.hzdr.de/

Identical to: https://www.hzdr.de/publications/Publ-32577

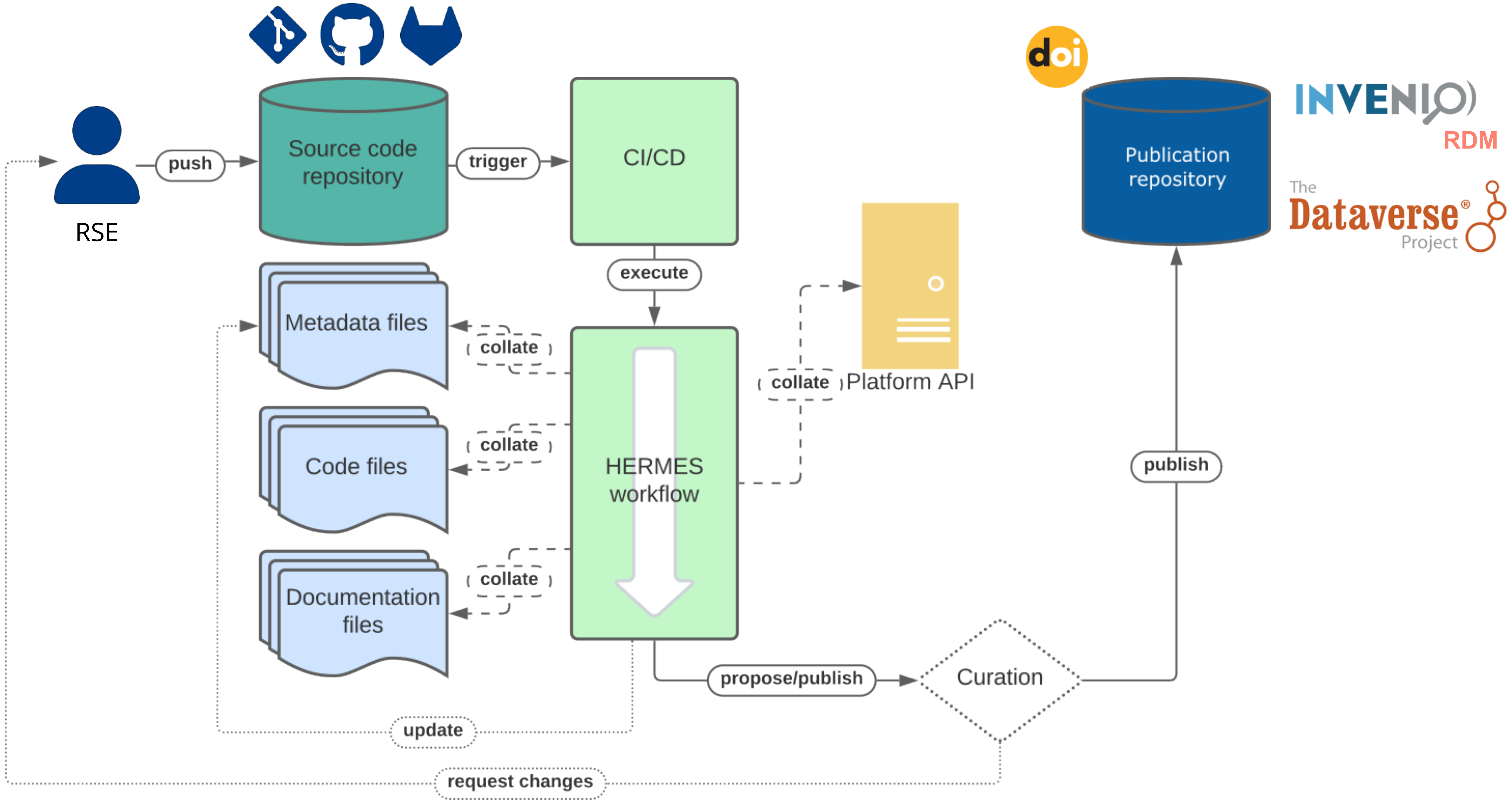
Supplementary material: 10.14278/rodare.193, 10.14278/rodare.938, https://gitlab.hzdr.de/heliport/heliport

Communities: Helmholtz-Zentrum Dresden-Rossendorf, RODARE

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

The HERMES Project: Automated Software Publication Workflow

- — A simple and transparent software publication workflow for open and closed access software can be a platform for an understandable science.
- — The metadata harvesting is essential to create a findable software publication.



HERMES

project.software-metadata.pub

github.com/hermes-hmc

team@software-metadata.pub

HMC HELMHOLTZ METADATA COLLABORATION

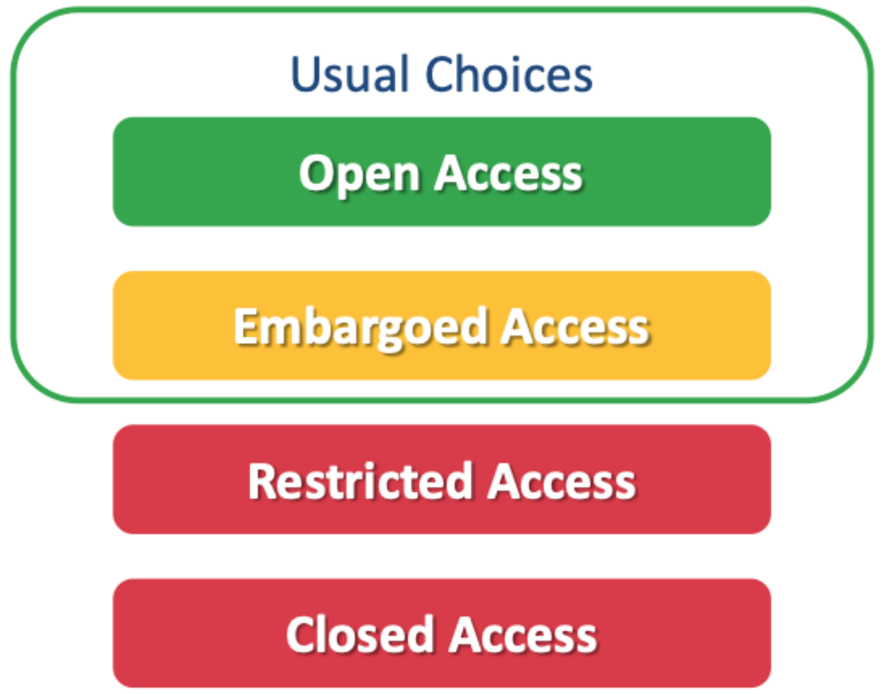
DLR JÜLICH Forschungszentrum HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

- 07/2021 – 06/2023
- Aim: Support RSEs in automatedly publishing their software with rich metadata

Data (and Software) Publication Repositories

- The data can be published in the same or a different repository as the software (possibly there is a domain-specific repository for the scientific data).
- A dataset (at least the metadata) should always be published to provide sustainable scientific evidence.
- The data itself can also be published under restricted access.*

*Nevertheless it fulfills the FAIR principles, because the steps to access the data are documented.



Usual Choices

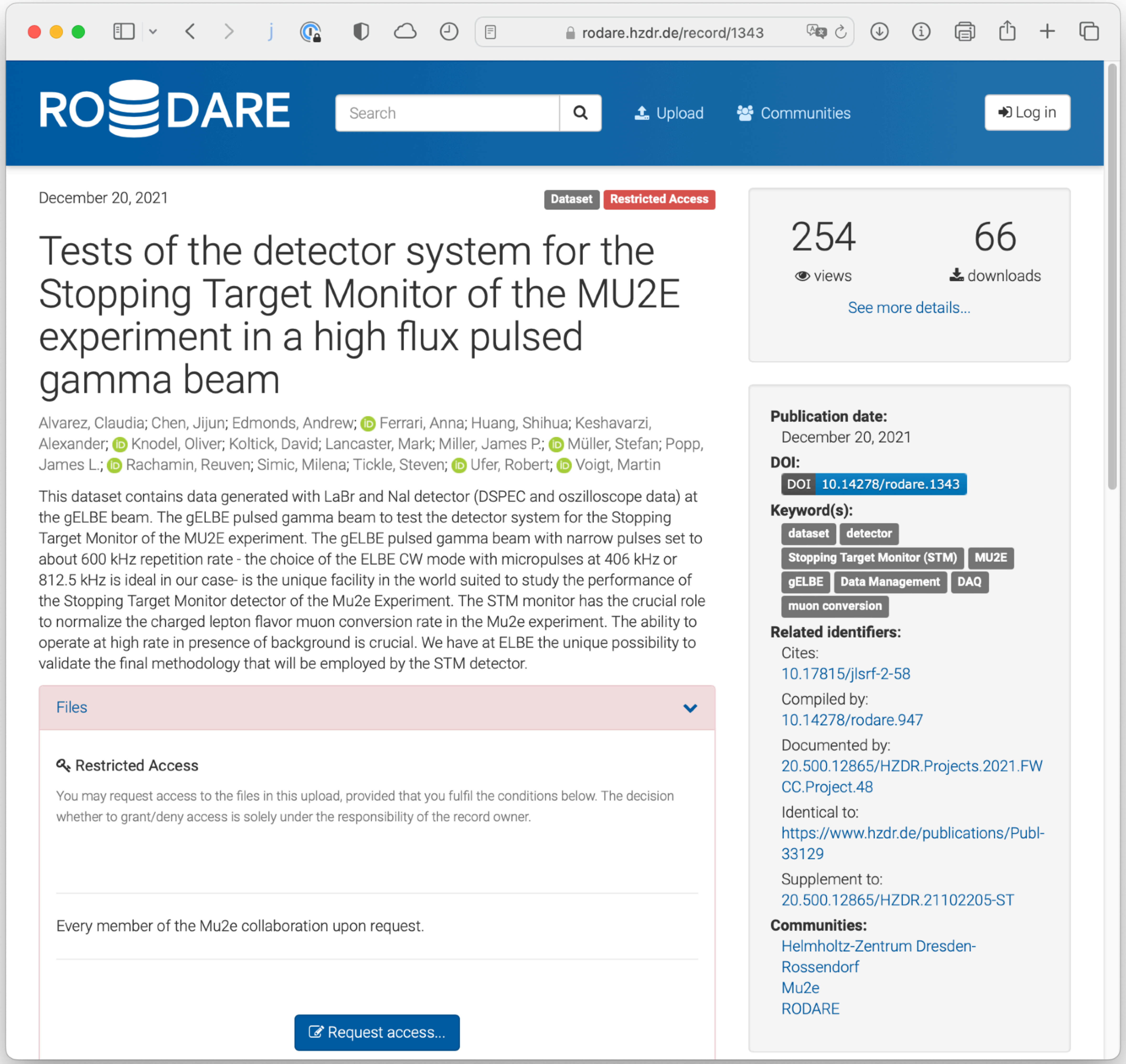
Open Access

Embargoed Access

Restricted Access

Closed Access

- The data publication should reference:
  - The software repository used to create or analyse the dataset.
  - The scientific publication based on the datasets.
  - The instrument or facility where the data was generated…



December 20, 2021

Dataset | Restricted Access

Tests of the detector system for the Stopping Target Monitor of the MU2E experiment in a high flux pulsed gamma beam

Alvarez, Claudia; Chen, Jijun; Edmonds, Andrew; Ferrari, Anna; Huang, Shihua; Keshavarzi, Alexander; Knodel, Oliver; Koltick, David; Lancaster, Mark; Miller, James P.; Müller, Stefan; Popp, James L.; Rachamin, Reuven; Simic, Milena; Tickle, Steven; Ufer, Robert; Voigt, Martin

This dataset contains data generated with LaBr and NaI detector (DSPEC and oszilloscope data) at the gELBE beam. The gELBE pulsed gamma beam to test the detector system for the Stopping Target Monitor of the MU2E experiment. The gELBE pulsed gamma beam with narrow pulses set to about 600 kHz repetition rate - the choice of the ELBE CW mode with micropulses at 406 kHz or 812.5 kHz is ideal in our case- is the unique facility in the world suited to study the performance of the Stopping Target Monitor detector of the Mu2e Experiment. The STM monitor has the crucial role to normalize the charged lepton flavor muon conversion rate in the Mu2e experiment. The ability to operate at high rate in presence of background is crucial. We have at ELBE the unique possibility to validate the final methodology that will be employed by the STM detector.

Files

Restricted Access

You may request access to the files in this upload, provided that you fulfil the conditions below. The decision whether to grant/deny access is solely under the responsibility of the record owner.

Every member of the Mu2e collaboration upon request.

Request access...

254 views | 66 downloads

See more details...

Publication date: December 20, 2021

DOI: DOI 10.14278/rodare.1343

Keyword(s): dataset | detector | Stopping Target Monitor (STM) | MU2E | gELBE | Data Management | DAQ | muon conversion

Related identifiers:

Cites: 10.17815/jlsrf-2-58

Compiled by: 10.14278/rodare.947

Documented by: 20.500.12865/HZDR.Projects.2021.FWCC.Project.48

Identical to: https://www.hzdr.de/publications/Publ-33129

Supplement to: 20.500.12865/HZDR.21102205-ST

Communities:

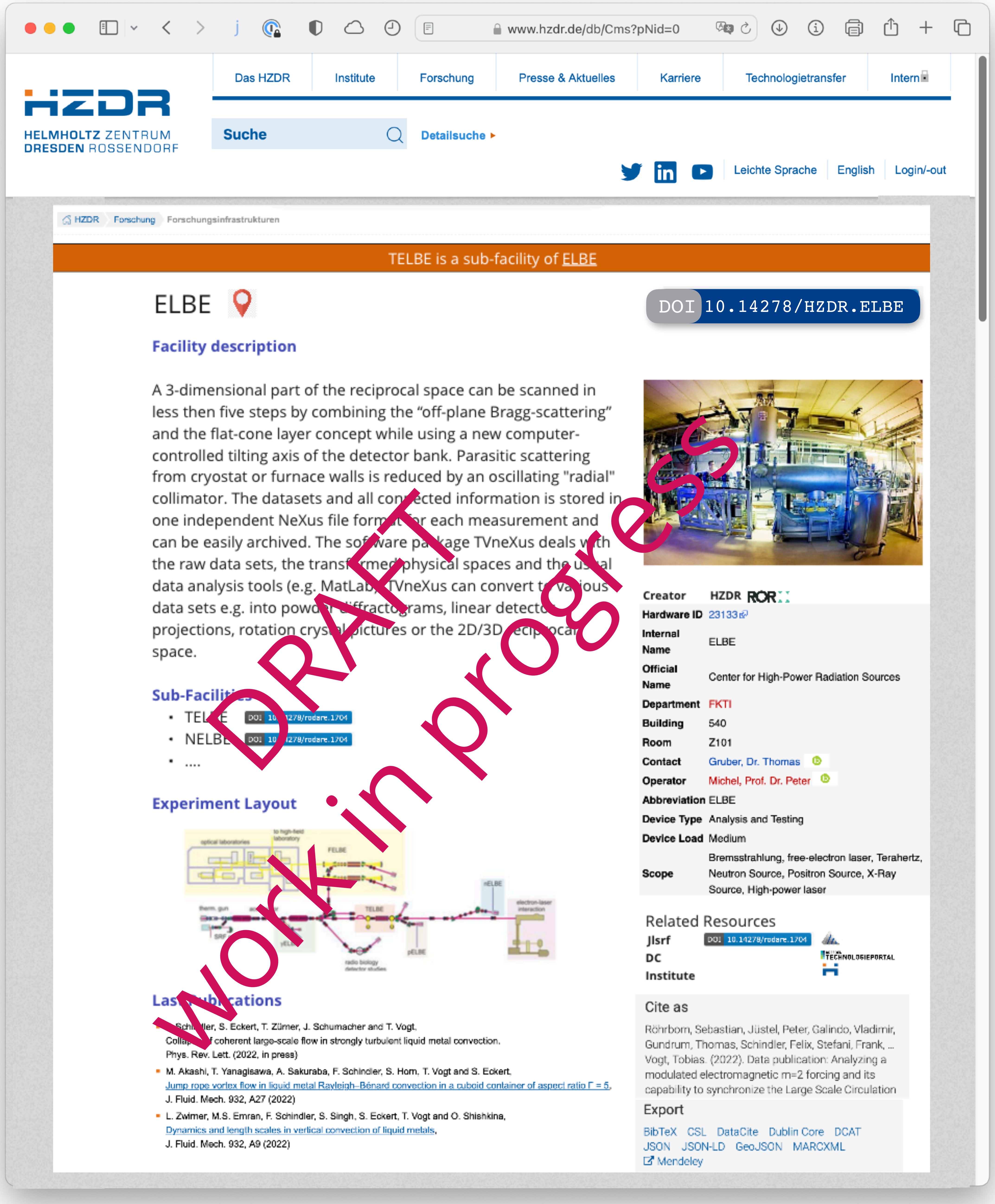
Helmholtz-Zentrum Dresden-Rossendorf

Mu2e

RODARE

Instrument DOIs and Landing Page

- For data publications we have the field *related identifiers*, were we can refer research facilities and instruments.
- Therefore, we plan to assign DOIs to instruments and provide DataCite records [2] and additional metadata on public landing pages.
- Components of the landing pages:
  - Mandatory: DOI, name, description, contacts, scope, location, ROR, device type.
  - Optional: Image, layout, sub-facilities, additional resources (JLSRF publication, internal website, …) and the latest publications.
  - Citation export to BibTex, JSON, …



TELBE is a sub-facility of ELBE

ELBE

DOI 10.14278/HZDR.ELBE

Facility description

A 3-dimensional part of the reciprocal space can be scanned in less then five steps by combining the "off-plane Bragg-scattering" and the flat-cone layer concept while using a new computer-controlled tilting axis of the detector bank. Parasitic scattering from cryostat or furnace walls is reduced by an oscillating "radial" collimator. The datasets and all connected information is stored in one independent NeXus file format for each measurement and can be easily archived. The software package TVneXus deals with the raw data sets, the transformed physical spaces and the usual data analysis tools (e.g. MatLab). TVneXus can convert to various data sets e.g. into powder diffractograms, linear detector projections, rotation crystal pictures or the 2D/3D reciprocal space.

Sub-Facilities

- TELBE DOI 10.14278/rodare.1704
- NELBE DOI 10.14278/rodare.1704
- ....

Experiment Layout

Latest Publications

- F. Schindler, S. Eckert, T. Zürner, J. Schumacher and T. Vogt, Collapse of coherent large-scale flow in strongly turbulent liquid metal convection. Phys. Rev. Lett. (2022, in press)
- M. Akashi, T. Yanagisawa, A. Sakuraba, F. Schindler, S. Horn, T. Vogt and S. Eckert, Jump rope vortex flow in liquid metal Rayleigh–Bénard convection in a cuboid container of aspect ratio Γ = 5, J. Fluid. Mech. 932, A27 (2022)
- L. Zwirner, M.S. Emran, F. Schindler, S. Singh, S. Eckert, T. Vogt and O. Shishkina, Dynamics and length scales in vertical convection of liquid metals, J. Fluid. Mech. 932, A9 (2022)

Creator	HZDR ROR
Hardware ID	23133
Internal Name	ELBE
Official Name	Center for High-Power Radiation Sources
Department	FKTI
Building	540
Room	Z101
Contact	Gruber, Dr. Thomas
Operator	Michel, Prof. Dr. Peter
Abbreviation	ELBE
Device Type	Analysis and Testing
Device Load	Medium
Scope	Bremsstrahlung, free-electron laser, Terahertz, Neutron Source, Positron Source, X-Ray Source, High-power laser

Related Resources

Jlsrf	DOI 10.14278/rodare.1704
DC	TECHNOLOGIEPORTAL
Institute	

Cite as

Röhrborn, Sebastian, Jüstel, Peter, Galindo, Vladimir, Gundrum, Thomas, Schindler, Felix, Stefani, Frank, ... Vogt, Tobias. (2022). Data publication: Analyzing a modulated electromagnetic m=2 forcing and its capability to synchronize the Large Scale Circulation

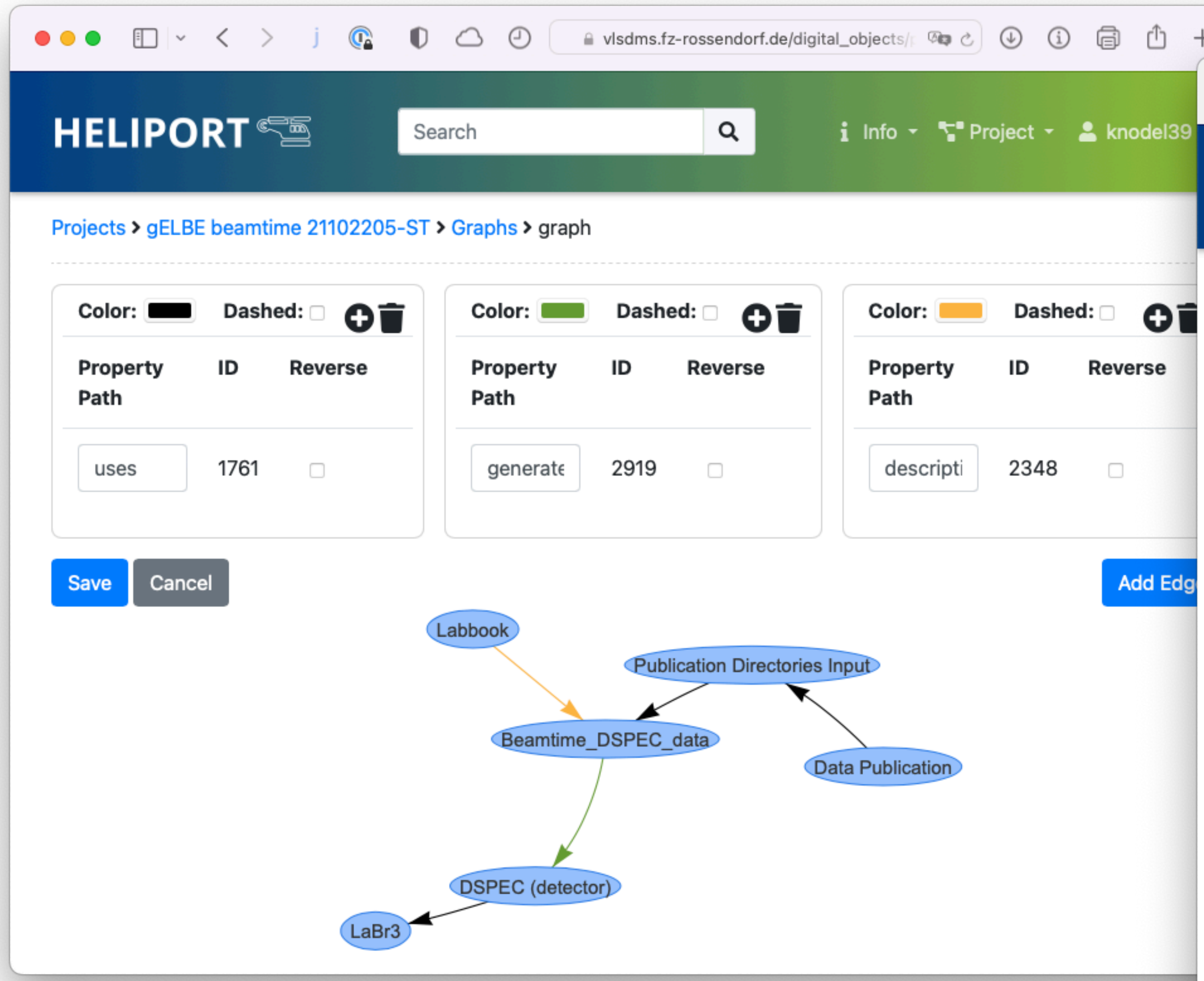
Export

BibTeX CSL DataCite Dublin Core DCAT JSON JSON-LD GeoJSON MARCXML Mendeley

[2] Bunakov, Vasily, Krahl, Rolf, Matthews, Brian, Vizcaino, Noeland Vukolov, Andrey, "Advanced infrastructure for PIDs in Photon and Neutron RIs", ExPaNDS project deliverable D2.5, Zenodo, Mar. 2022. doi: 10.5281/zenodo.5905351.

Digital Objects and Handles Enable Long-term Sustainability

- At the HZDR, we use DOIs for resources containing the whole set of bibliographic metadata:
  - Scientific articles,
  - Published datasets and software,
  - Instruments.
- Other identifiers included in the metadata are ORCID and ROR available in our internal databases for almost every scientist at HZDR.
- Further digital objects can request PIDs from our Handle hdl*enabled* server.
- The digital objects in our ecosystem can be correlated with each to create a comprehensible experiment providing data provenance.



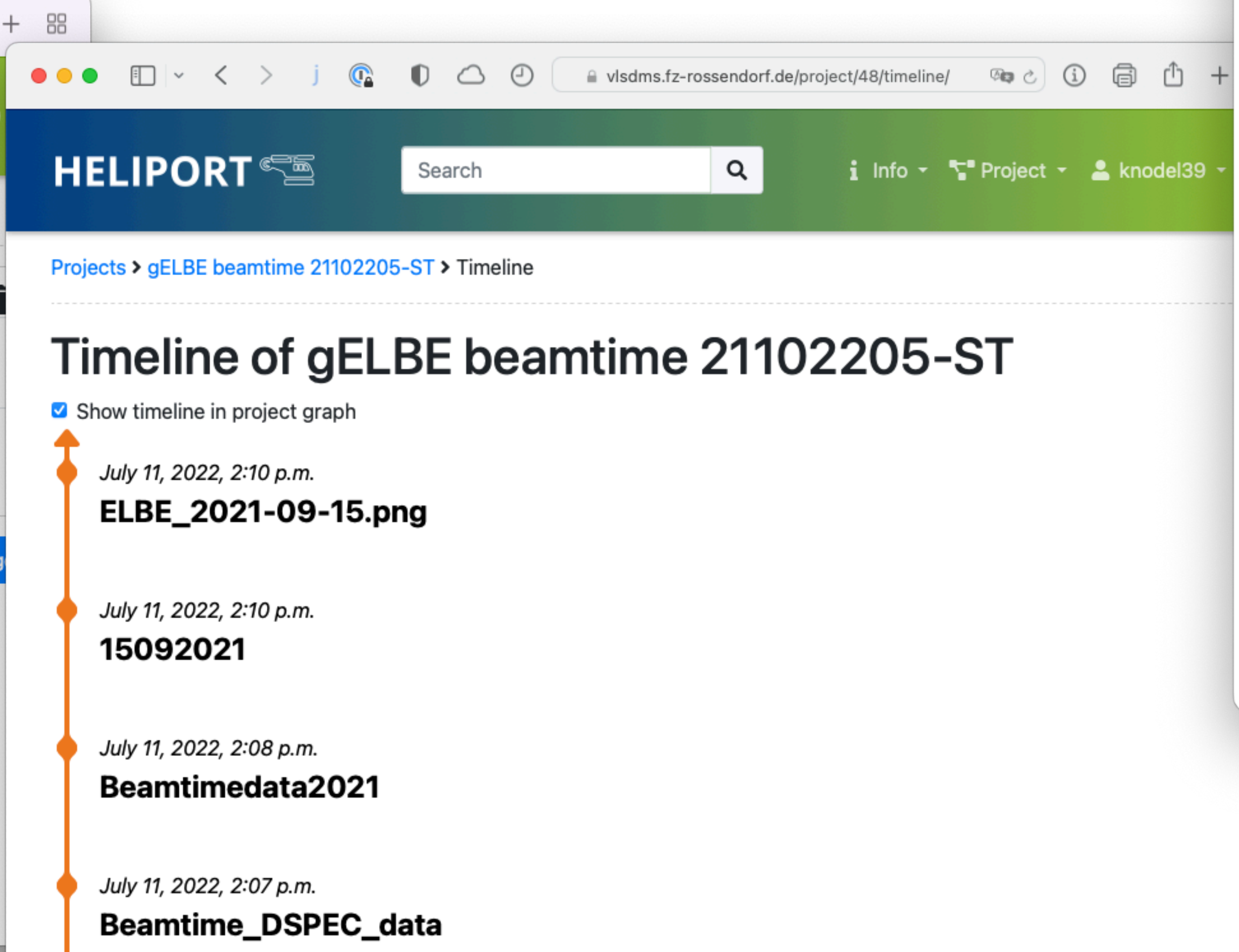
Projects > gELBE beamtime 21102205-ST > Graphs > graph

Color: ⬛	Dashed: □	
Property Path	ID	Reverse
uses	1761	□

Color: 🟩	Dashed: □	
Property Path	ID	Reverse
generate	2919	□

Color: 🟧	Dashed: □	
Property Path	ID	Reverse
descripti	2348	□

Save Cancel Add Edg



Projects > gELBE beamtime 21102205-ST > Timeline

Timeline of gELBE beamtime 21102205-ST

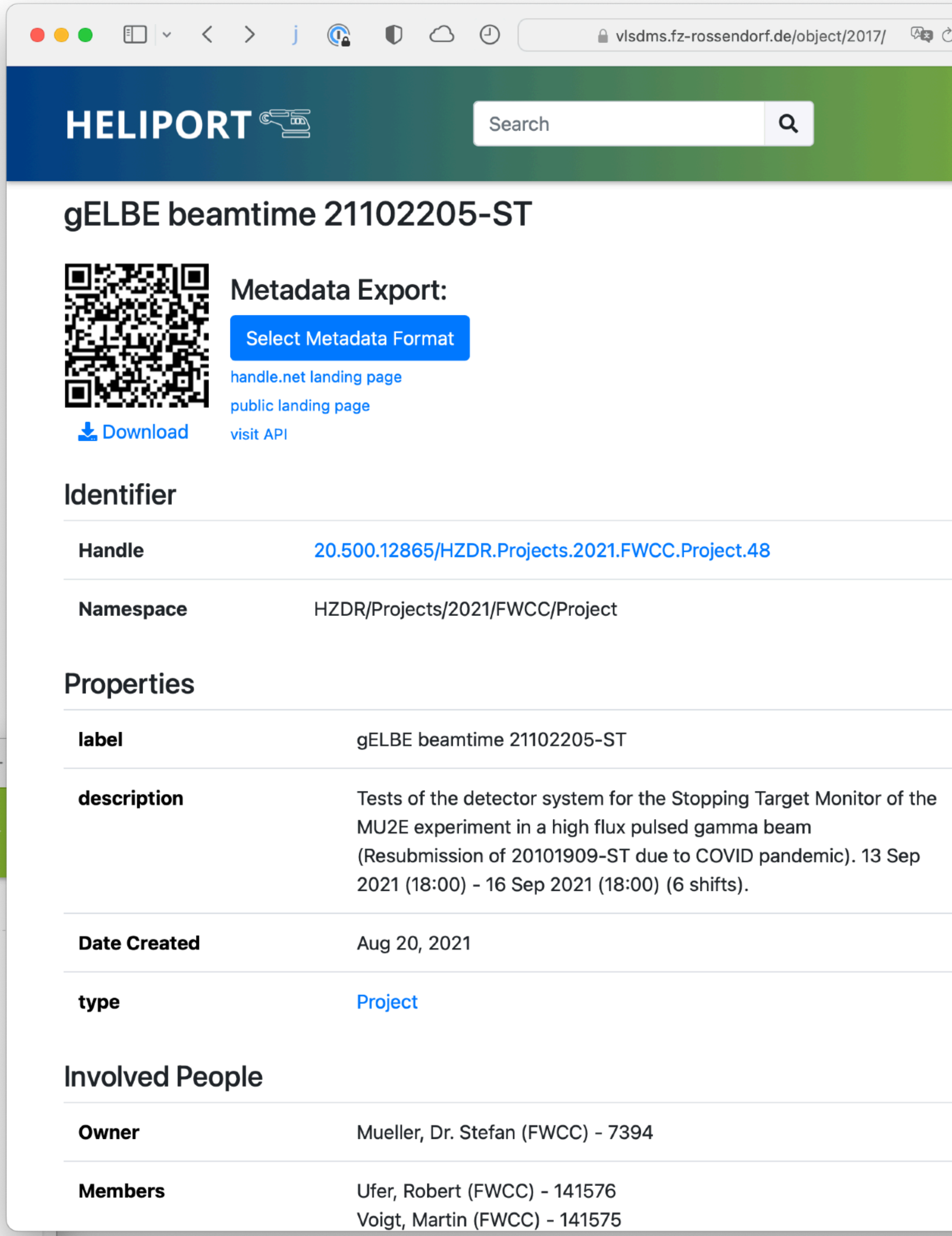
☑ Show timeline in project graph

July 11, 2022, 2:10 p.m.
ELBE_2021-09-15.png

July 11, 2022, 2:10 p.m.
15092021

July 11, 2022, 2:08 p.m.
Beamtimedata2021

July 11, 2022, 2:07 p.m.
Beamtime_DSPEC_data



gELBE beamtime 21102205-ST

Metadata Export:
Select Metadata Format
handle.net landing page
public landing page
visit API
Download

Identifier

Handle	20.500.12865/HZDR.Projects.2021.FWCC.Project.48
Namespace	HZDR/Projects/2021/FWCC/Project

Properties

label	gELBE beamtime 21102205-ST
description	Tests of the detector system for the Stopping Target Monitor of the MU2E experiment in a high flux pulsed gamma beam (Resubmission of 20101909-ST due to COVID pandemic). 13 Sep 2021 (18:00) - 16 Sep 2021 (18:00) (6 shifts).
Date Created	Aug 20, 2021
type	Project

Involved People

Owner	Mueller, Dr. Stefan (FWCC) - 7394
Members	Ufer, Robert (FWCC) - 141576 Voigt, Martin (FWCC) - 141575

Linking Data, Text and Research Software Together

Software:

I. HERMES can extract the metadata provided by Github or GitLab.

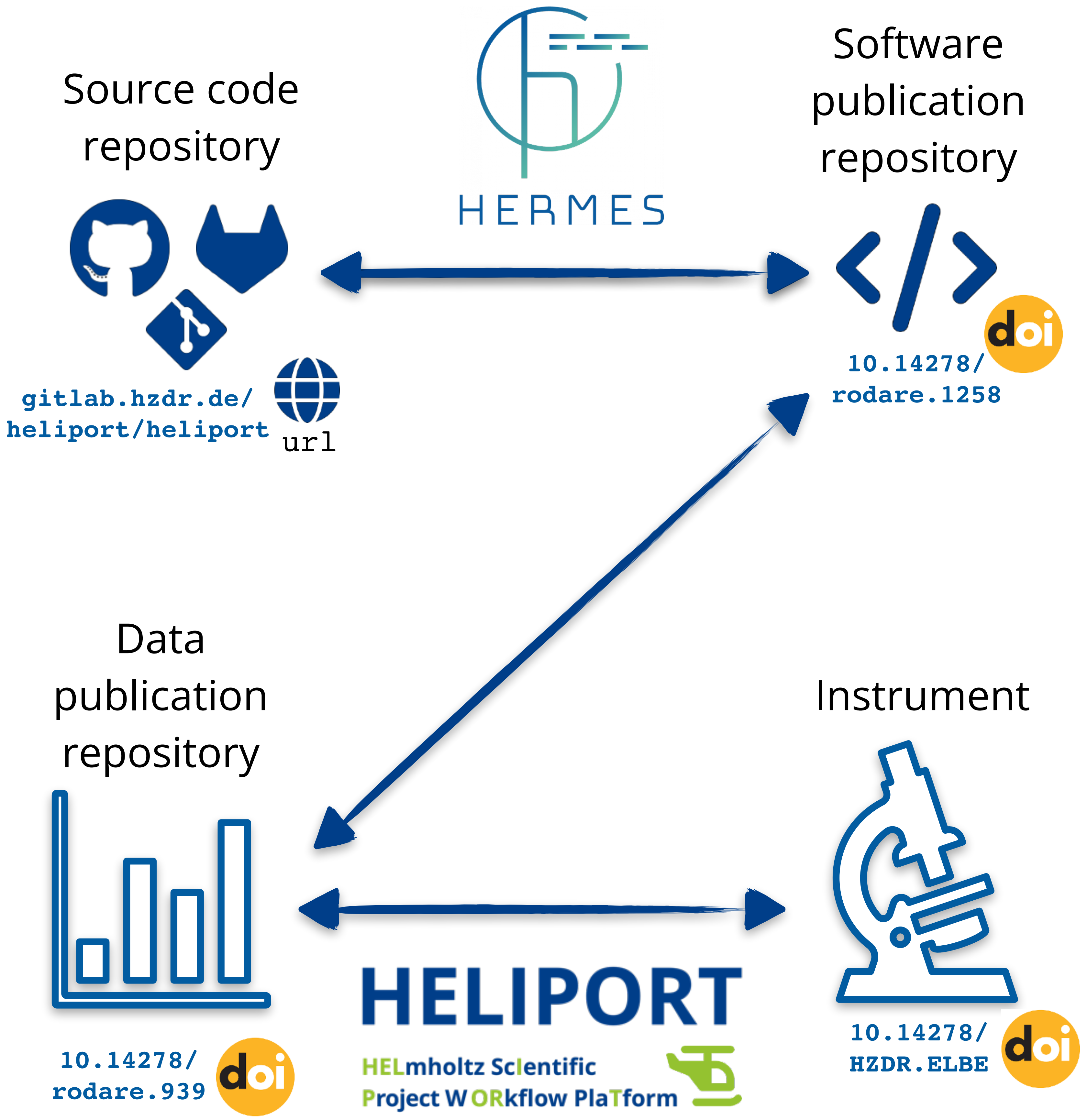
II. A software release can trigger a pipeline that initialises a publication with DOI based on the available (and third-party) metadata.

III. In a subsequent step, the DOI is added to the Readme file in the Git repository and the cross-linking is completed.

Data:

I. After data collection or processing, a pipeline can start collecting metadata from a proposal system or other related services.

II. The metadata and information from a computational workflow can be used to create a data publication with references to a specific software version (DOI) and the instrument where the data was taken.

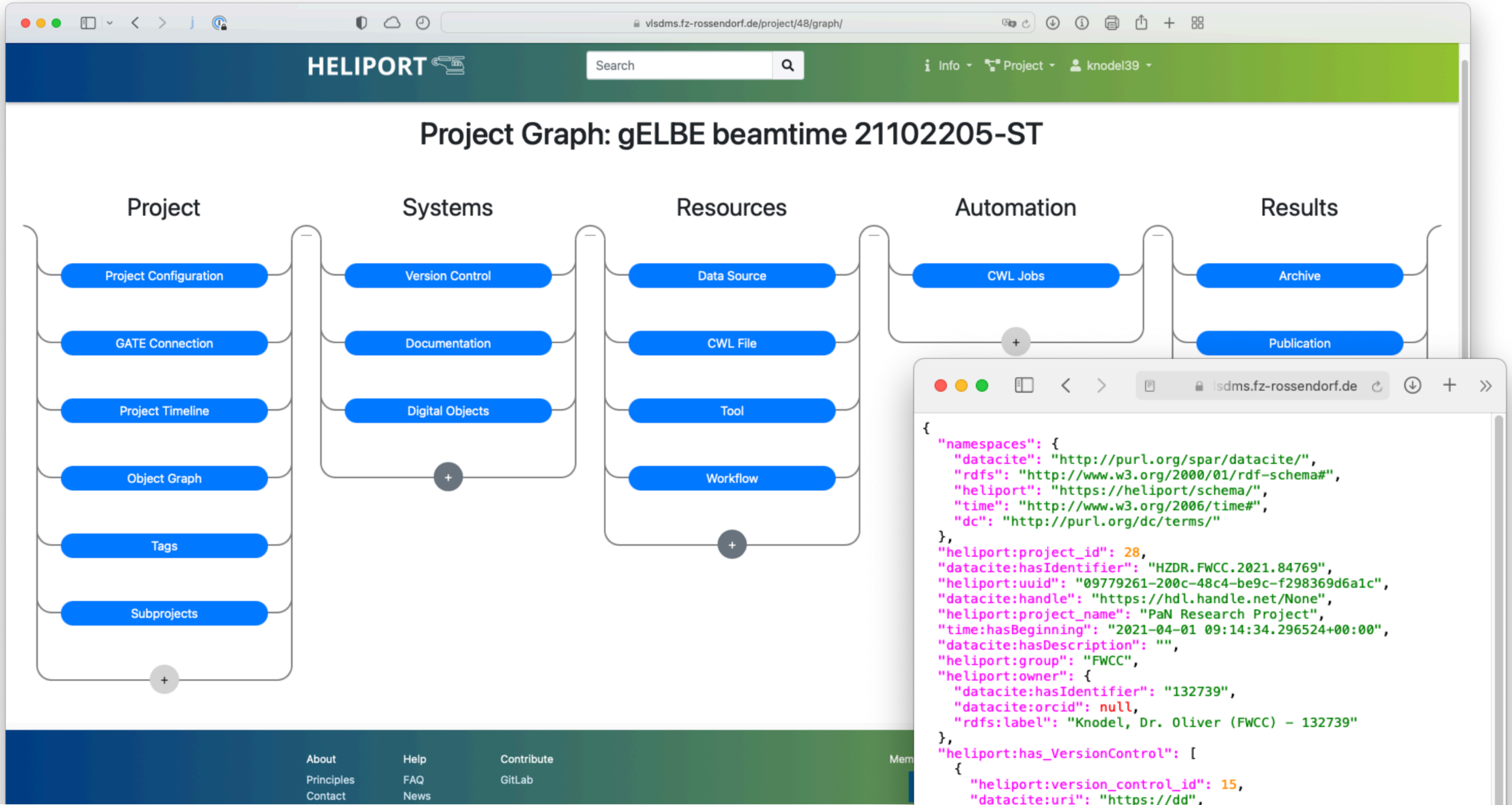

Source code repository
HERMES
Software publication repository
gitlab.hzdr.de/ heliport/heliport
url
doi
10.14278/ rodare.1258
Data publication repository
Instrument
10.14278/ rodare.939
doi
HELIPORT
HELmholtz ScIentific Project WORkflow PlaTform
10.14278/ HZDR.ELBE
doi

Overview of the Project Resources from a Higher Level

> The HELIPORT project aims at developing a platform which accommodates the **complete life cycle** of a scientific project and links all corresponding programs, systems and workflows to create a more **FAIR** and comprehensible project description.

HELIPORT

HELmholtz ScIentific Project WORkflow PlaTform



Project Graph: gELBE beamtime 21102205-ST

Project
- Project Configuration
- GATE Connection
- Project Timeline
- Object Graph
- Tags
- Subprojects

Systems
- Version Control
- Documentation
- Digital Objects

Resources
- Data Source
- CWL File
- Tool
- Workflow

Automation
- CWL Jobs

Results
- Archive
- Publication

```
{
  "namespaces": {
    "datacite": "http://purl.org/spar/datacite/",
    "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
    "heliport": "https://heliport/schema/",
    "time": "http://www.w3.org/2006/time#",
    "dc": "http://purl.org/dc/terms/"
  },
  "heliport:project_id": 28,
  "datacite:hasIdentifier": "HZDR.FWCC.2021.84769",
  "heliport:uuid": "09779261-200c-48c4-be9c-f298369d6a1c",
  "datacite:handle": "https://hdl.handle.net/None",
  "heliport:project_name": "PaN Research Project",
  "time:hasBeginning": "2021-04-01 09:14:34.296524+00:00",
  "datacite:hasDescription": "",
  "heliport:group": "FWCC",
  "heliport:owner": {
    "datacite:hasIdentifier": "132739",
    "datacite:orcid": null,
    "rdfs:label": "Knodel, Dr. Oliver (FWCC) - 132739"
  },
  "heliport:has_VersionControl": [
    {
      "heliport:version_control_id": 15,
      "datacite:uri": "https://dd",
```

heliport.hzdr.de

codebase.helmholtz.de/heliport

heliport@hzdr.de

<HMC> HELMHOLTZ METADATA COLLABORATION

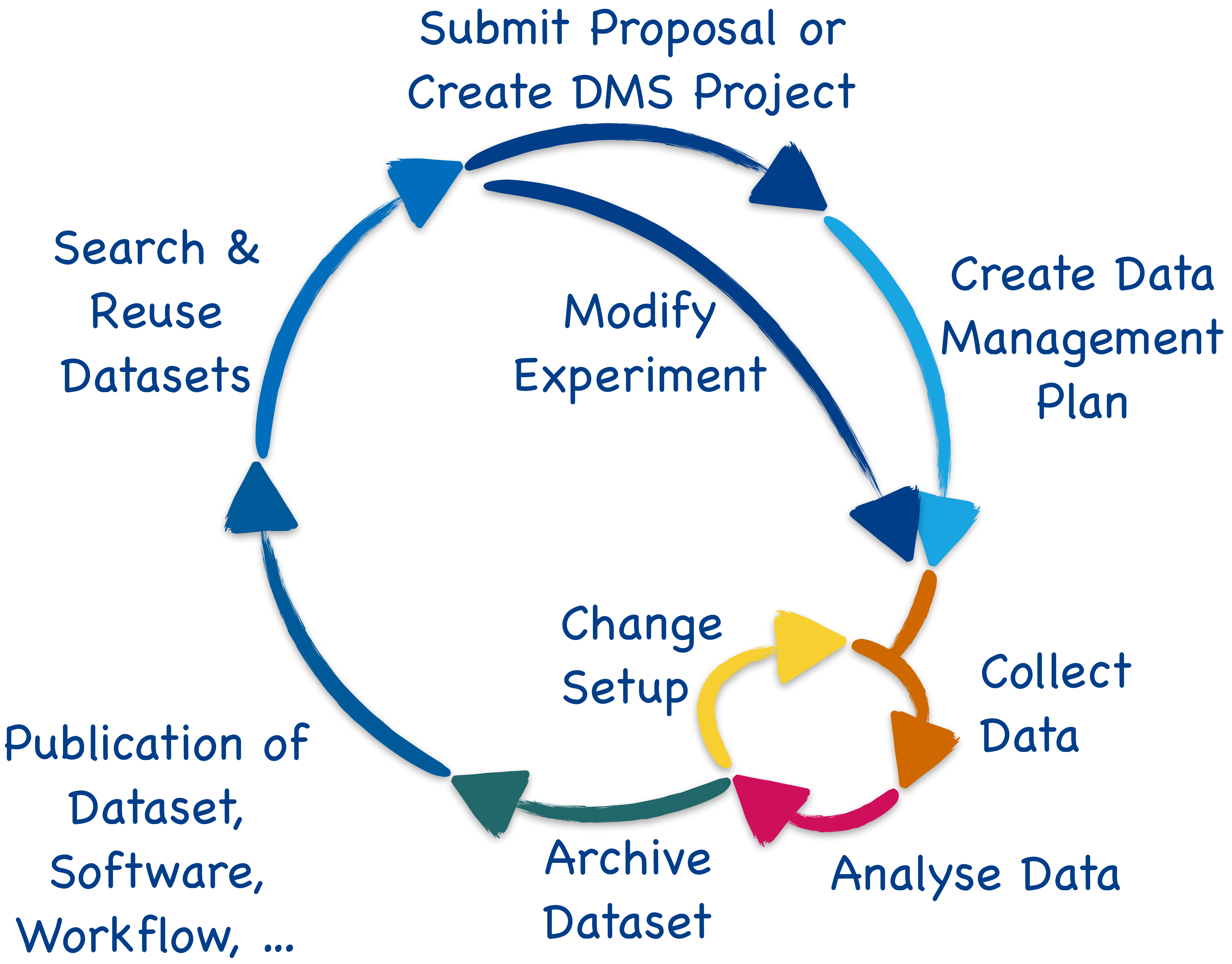
HZDR HELMHOLTZ ZENTRUM DRESDEN ROSSENDORF

HI JENA Helmholtz Institute Jena

JÜLICH FORSCHUNGSZENTRUM

- 07/2021 – 06/2023
- Aim: Collect every system, service or digital product of a research project in an uniform metadata package.

Our Challenge: An End-to-End Digital Data Lifecycle



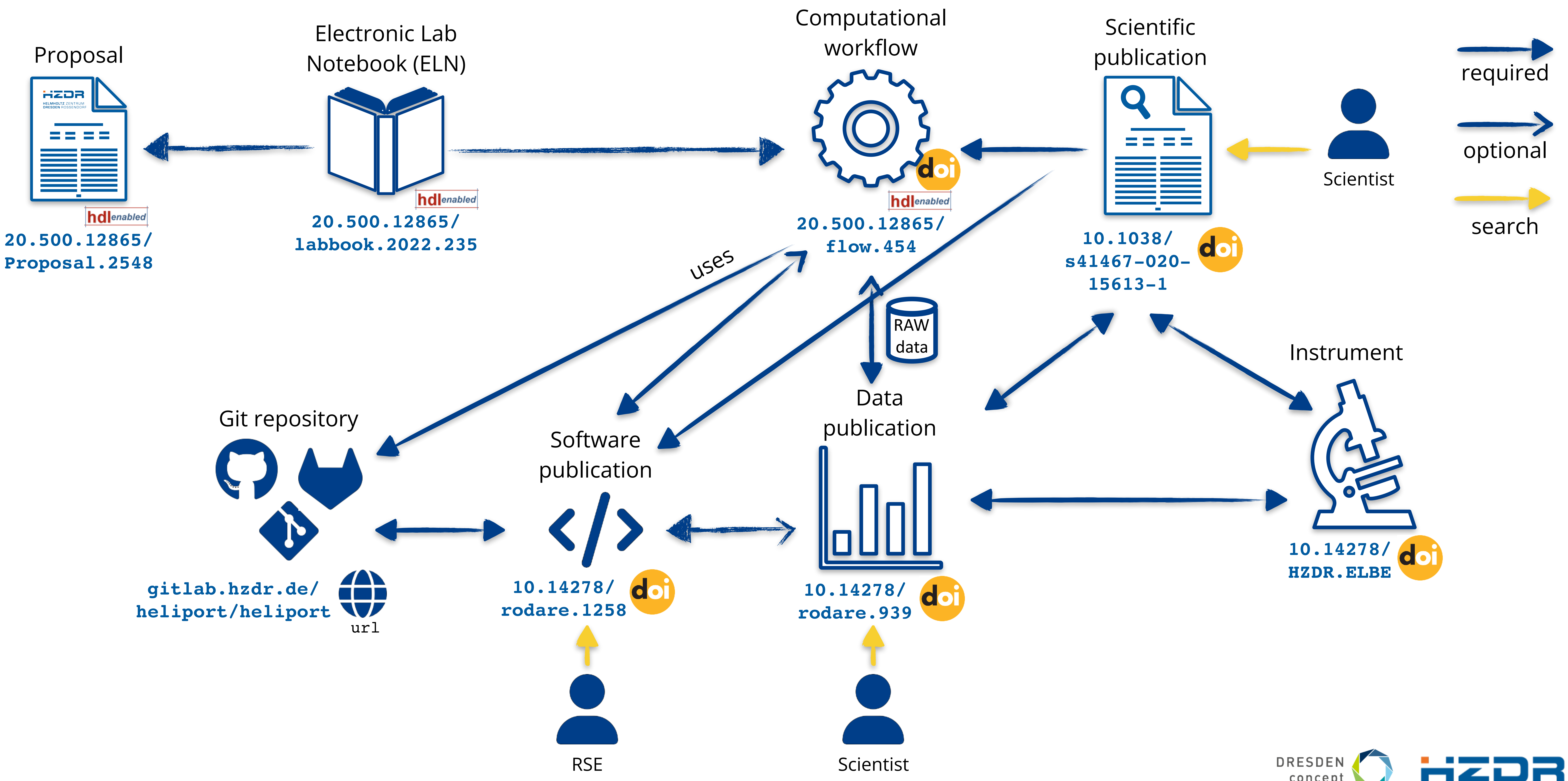
HELIPORT

HELmholtz ScIentific Project WORkflow PlaTform

HERMES

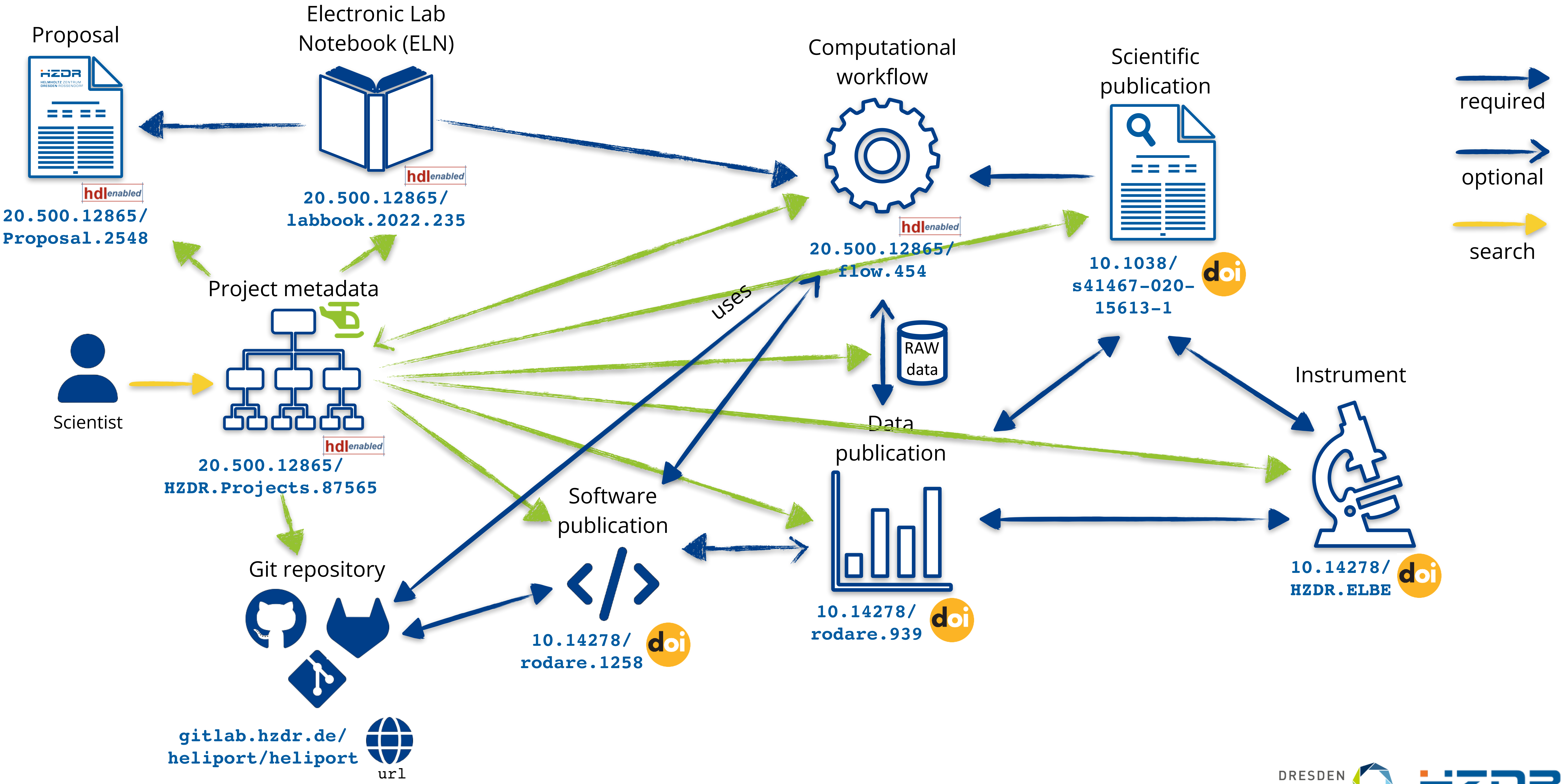
DRESDEN concept HZDR

Top-Level View of the Interlinked Digital Objects of an Experiment at the HZDR



Including HELIPORT

Top-Level View of the Interlinked Digital Objects of an Experiment at the HZDR



Proposal

hdl enabled

20.500.12865/
Proposal.2548

Electronic Lab
Notebook (ELN)

hdl enabled

20.500.12865/
labbook.2022.235

Computational
workflow

hdl enabled

20.500.12865/
flow.454

Scientific
publication

10.1038/
s41467-020-
15613-1

doi

required

optional

search

Scientist

Project metadata

hdl enabled

20.500.12865/
HZDR.Projects.87565

uses

RAW
data

Data
publication

10.14278/
rodare.939

doi

Instrument

10.14278/
HZDR.ELBE

doi

Software
publication

10.14278/
rodare.1258

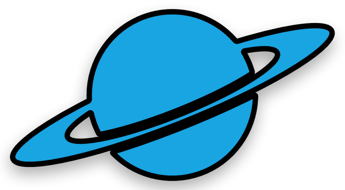
doi

Git repository

gitlab.hzdr.de/
heliport/heliport

url

Conclusions and Outlook



Conclusion:

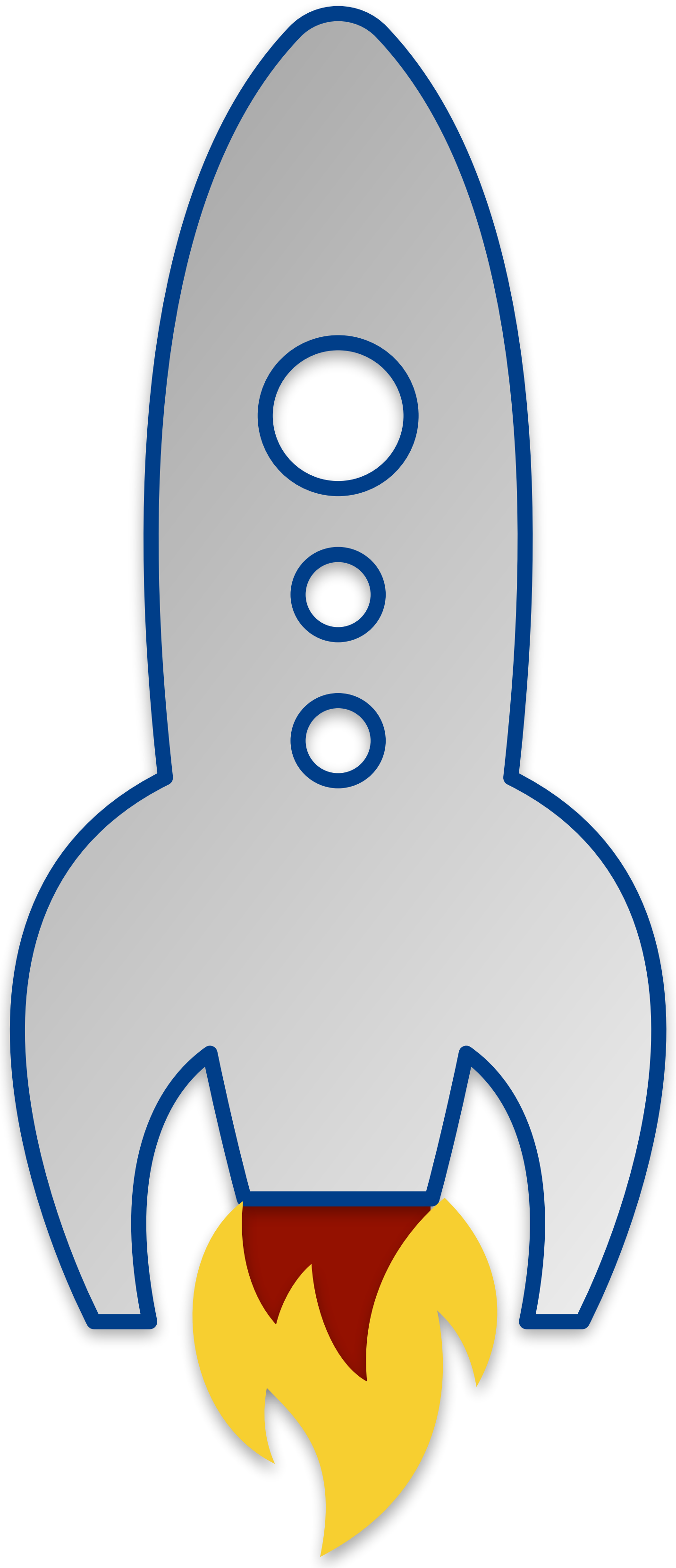
- For an interlinked ecosystem, it is necessary to consider different entry points for the provision of metadata.
- The cross-linking of the services and systems is unavoidable to enable comprehensible science.

➔ Automated pipelines and workflows are the key to exchange metadata and support scientists and RSEs.

Status:

- We provide DOIs for software and data (instrument DOIs are work in progress),
- Handles can be created for all types of digital objects.

➔ With HELIPORT and HERMES, we develop systems that automate the exchange of metadata between internal and external systems and services.



HELIPORT
HELmholtz ScIentific Project WORkflow PlaTform

RODARE
ROSSENDORF DATA REPOSITORY

HERMES