Pioneering Digital Research Landscapes: Innovations at HZDR

Towards Open Digital Research Ecosystems – Interconnecting Infrastructures, Open Science Forum, February 14, 2024

Oliver Knodel // contact: 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 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
We support many steps of our different research experiment (matter, energy and health) with a wide range of tools:

- Electronic lab notebook (E-Logbook),
- Interactive analysis,
- FAIR Publication of datasets,
- Scientific workflow management,
- Handle (PID) generation and management.

A uniform and smooth access to and between all services and systems in our ecosystem is necessary.

The documentation of all these linked resources is essential to create a comprehensible and FAIR data lifecycle.
The Foundation for our Digital Data Lifecycle: HZDR Data Policy

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

© Helmholtz-Gemeinschaft 2023, https://os.helmholtz.de/open-research-data/forschungsdaten-policies/
Starting Point in 2018: Data Publication Platform RODARE

Image of a website interface with recent uploads:
- **March 19, 2021 (r1)** Dataset
  - Title: "Fluorination of graphene leads to susceptibility for nanopore formation by highly charged ion impact"
  - Authors: Schäkm, Hübner, René, Faccio, Stefan
  - Description: The repository contains STEM images, experimental data from charge exchange measurements, and data from charge exchange simulations.
  - Uploaded on March 11, 2021

- **March 9, 2021 (r1)** Software
  - Title: PICongGPU setup: LPWF4 cownramp injection
  - Authors: Bauer, Behrens, Belica, Cheung, Peter, Rostocki, Siegel, Rosemann, Schmuck, Thomas, Schöbel, Schramm, Ulrich, Strehlitz, Klaus, Fritzsche, Patrick, Widera, René, Debuss, Alexander
  - Description: PICongGPU source code and setup files used for the LPWF4 cownramp injection simulation study.
  - Uploaded on March 9, 2021

- **March 3, 2021 (r1)** Dataset
  - Title: Data for: Experimental studies on bubble aspect ratio and corresponding correlations under bubble swarm condition
  - Authors: Liu, Liu, Zhang, Heyng, Yan, Honglei, Ziegelmehl, Thomas, Heidenkemp, 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
- Have a look at the restructured documentation and blog system of RODARE. We now can more easily notify about news and features. You can find tutorials there. Visit https://rodare.hzdr.de/about.

RODARE now offers usage statistics!
- 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.

Registered in: re3data.org
http://doi.org/10.17616/R36R40

Findable with:
- DRESデン concept
- HZDR

Powered by:
- zenodo
- INVENIO
- re3data.org
- B2FIND
- EUDAT
Our Observations Over the Following Years…

— We need to support the entire experiment with reliable interconnected tools to enable comprehensible and FAIR science.

— The resulting IT infrastructures are complex.

— Documentation is necessary, but typically time is missing.

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

— An overarching system guiding our scientists (and visitors) through the lifecycle of their research project is essential.
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.
Requirements and Expectations

- HELIPORT was intended to provide only the proposal’s metadata, from internal and external scientists, to allow the assignment of resources.

- Over time, we realised that HELIPORT can also answer our scientists’ most important questions, such as:

  - How can we automate recurring processes and keep track of status and data products?
  - How can we bring new team members or visiting/external scientists into our project lifecycle and all associated tools?
  - Which datasets or software can be published (and how)?
  - Where are data, software and how can I gain access to both of them?

And how we can support them?!

What are the necessary steps towards a full comprehensible and FAIR research experiment ensuring data provenance?
Example: HELIPORT Supported Experiment at The TELBE Beamline

— HELIPORT provides access to our services and guides external/visiting scientists through the entire experiment.
— This requires the experiment to be mapped to systems in the HZDR infrastructure.

Proposer (Principal Investigator)

I. Apply

II. Configure

HZDR Beamline Scientist

Data Source (Cluster Storage, network drive)

DAQ LabView

Time Series Data (Visualization with Grafana)

Proposal Management (GATE)

Project-Level Metadata

Digital Object/Project Graph

Data Sources

Workflow Dashboard

Lab Documentation

Heliport (Plugin)

E-Labbook Frontend (MediaWiki)

SciCat (Plugin)

Workflow (HPC Cluster)

SciCat Metadata Catalogue

Publication

Cites

Registered

Public

HZDR PIDs (Handle.net)

SciCat (Plugin)

SciCat (Plugin)

SciCat (Plugin)

HZDR PIDs (Handle.net)

HZDR PIDs (Handle.net)

HZDR PIDs (Handle.net)

WorkflowHub.eu

RODARE

ROBIS

Publication Database

Data Sources

Project - Level Metadata

Workflow Dashboard

Digital Object/Project Graph

Helxport

Experimental Team (Visiting Scientists)

E-Labbook

Automated logging

> 10GB per min

Data analysis

Data Analysis: Sorting, Binning (HPC/OpenStack)

I. Apply

II. Configure

Proposer (Principle Investigator)

Experimental Team (Visiting Scientists)
I. Proposal Submission

Automated transfer of project metadata from the proposal system (GATE) into HELIPORT:

— Title, Authors, Description,
— Beamtime schedule,
— Large-scale facility used,
— Scientific method (PaNET)
II. Project List and Dashboard

— Typically, a beam line scientist is the owner of a HELIPORT project and the proposer has the role of the manager and can add additional project members.

— Tags and sub-projects including inheritance are possible in the project list.
III. Resources: Documentation and Repositories

The documentation section is typically used to refer to all internal and external systems or services used:

— E-Labbook (Mediawiki),
— GitLab, Github, Workflowhub, ...

Documentation

<table>
<thead>
<tr>
<th>ID</th>
<th>Description</th>
<th>System</th>
</tr>
</thead>
<tbody>
<tr>
<td>02</td>
<td>Present documentation in Mediawiki</td>
<td>Mediawiki</td>
</tr>
</tbody>
</table>

Add a Documentation

FWKP:22 DAQ GaAs 120deg's WP 45deg's SHG V polar 01

Version Control

<table>
<thead>
<tr>
<th>ID</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>Sorting and registration of Terahertz ELBE raw data version 1</td>
</tr>
</tbody>
</table>

Add a Source Code Repository

Choose where you want to create your new repository

- [ ] HZDR Gitlab repository
- [ ] GitHub repository
IV. Detector Control and Workflows

— The HELIPORT REST-API enables the transfer of metadata between HELIPORT and external systems (e.g. detector control in LabView).

— The integrated job/workflow submission system collects metadata with **provenance information** provided by HELIPORT.

— Workflows (on our HPC cluster) can be accessed by any project member using the HELIPORT web frontend.
V. Data Sources

— Folders and files in our internal filesystems can be registered in HELIPORT as **data source**.

— Each **project member** has access to the files and folders using HELIPORT.

— The provenance of the data sets generated from an experiment is entirely comprehensible.
Automated data publication with:

- Metadata from Proposal System,
- Files and folders registered and selected in HELIPORT.
VII. Search & Reuse: Metadata Catalogue SciCat

— With our data publication repository RODARE we provide a platform for publishing datasets.
— In RODARE, bibliographic metadata is based on DataCite.
— For additional scientific metadata we use SciCat and reference the specific datasets in RODARE or HELIPORT.
Metadata Catalogue SciCat and Data Repository RODARE (Draft)

Curated Metadata Source
- ExperimentLogging app (ExL)
- E-Logbook

Public Metadata Catalogue
- SciCat

Data Access
- RODARE
- HELIPORT

Metadata from Experiment/Simulation

Direct API Call

Dataset

Filesystem

Tape Archive
Conclusions

— Close and automated interaction of services and systems in our digital research landscape is essential to provide FAIR and comprehensible research projects.

— HELIPORT describes and collects metadata from services and systems involved in experiments.

— APIs and workflows are used to transfer metadata between our services and systems.
Resources

Website: heliport.hzdr.de

Repository: codebase.helmholtz.cloud/heliport

API Doc: heliport.hzdr.de/redoc/
Example: Data Management View of the TELBE Experiment

— Terahertz facility at the ELBE center for High-Power Radiation Sources.
— In the future HELIPORT guides (external) scientists through the complete experiment.
— Submission of data analysis Jobs from LabView to UNICORE with visualisation in HELIPORT.
Documentation Ecosystem at HZDR

— Our **Electronic Lab Documentation** is a central database gathering information from various systems:
  • Proposal management (GATE),
  • Environmental data,
  • Devices (e.g. Labview),
  • Workflows,
  • (Meta)data databases.

— Different frontends are available:
  • SciCat (metadata only),
  • MediaWiki (structured user-definable views).

— The system is build on-top:

---

- **Data Source** (Analyzer, Experiment, Simulation,...)
- **Device or Detector** (Experiment)
- **Facility** (Environment)
- **HZDR PIDs** (Handle.net)
- **Lab Documentation**
  - **E-Logbook Frontend** (MediaWiki)
  - **SciCat** (Metadata)
  - **E-Logbook Backend** (Database & API)
- **Version Control** (GitLab)
- **Workflow Engine**
- **Compute** (HPC/OpenStack)
- **HELIPORT** (GATE and Workflow API)

---

*Interfaces to our Infrastructure Resources*
Relations Between Digital Objects and

— Relations between digital objects are visualized to provide a top-level view on the project with dependencies.
— The relationships between simulation (surrogate model) and experiment can also be demonstrated.
— The versioning of an experiment is an essential extension, and first approaches via a timeline are being evaluated.
HELIPORT offers an infrastructure which permits the integration of various workflow languages and access modes to HPC infrastructures.

The infrastructure keeps track of and collects the metadata and enables access to all resources involved.

Next steps:
- Python library sending workflow information directly to HELIPORT,
- Provision of provenance information from Jupyter notebooks,
- Use case: PIConGPU