Constructing tomorrow with data

NFDI4ING
Research data management in engineering sciences

Technical University of Munich

Archetype DORIS

Data from High-Performance Measurement and Computation (HPMC)

About DORIS

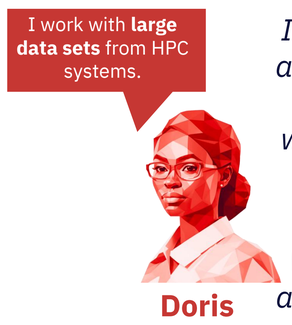
Who is involved in the task area?

The DORIS task area collaborates with all tier 1 (EU) Gauss Centre for Supercomputing facilities on various projects to improve the research data management of researchers from engineering on these systems. Furthermore, we are increasingly aligning ourselves with the National High Performance Computing Alliance (NHR), strengthening our ties and planning close collaborations for the future.

HLRS High-Performance Computing Center Stuttgart · JÜLICH Forschungszentrum · lrz Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities

Who does DORIS support?

To which researchers do we provide services?



*I'm an engineer conducting and post-processing **high-resolution** and **high-performance measurements and simulations on High-Performance Computing systems** (HPC). The data sets I work with are extremely large such that they are immobile and have to be post-processed locally on supercomputers.*

That's why we bring the user to the data!

We advance metadata standards and collection, facilitate data accessibility and enhance reproducibility and re-use of HPC data.

We offer RDM support and solutions for all engineers who generate or process their data on HPC systems. Engineers using these supercomputers mostly come from fluid mechanics, thermal and heat science, materials, and construction engineering.
Researchers looking for solutions or offering innovative approaches are invited to engage with the DORIS task area!

RDM with large (HPMC) Data

What are HPMC research data?

High Performance Measurement
- Measurement data
- Metadata, tools & methods

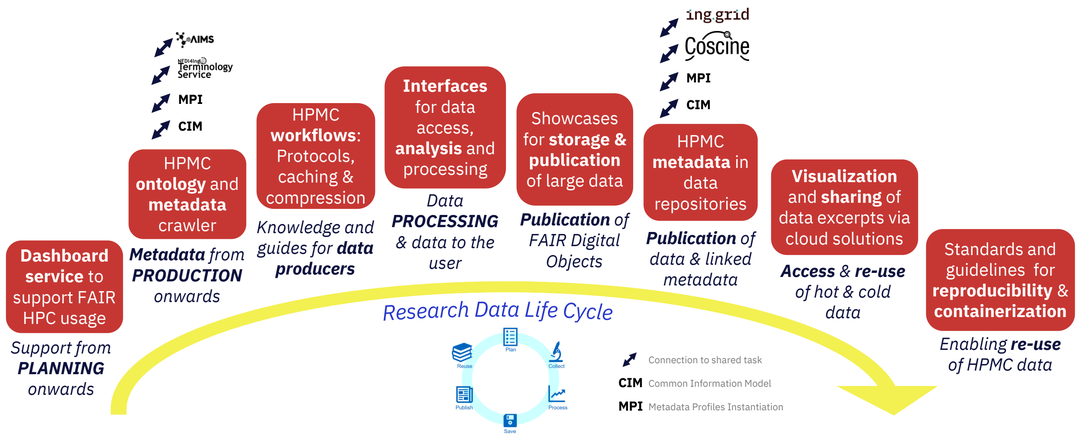
Analysis and processing of measurement data using HPC

High Performance Computing
- Software, script / code
- Input file, output file, log file
- Raw data, processed data
- Data for secondary research (e.g. energy consumption in HPC)
- Metadata, tools & methods

How DORIS serves the community

Providing Tools, Services, Guides, and Support for Researchers

DORIS offers tools and services **across the entire research data lifecycle**, from the planning phase to publication and reuse. Our goal is to improve Research Data Management for engineers in High-Performance Computing while supporting compliance with FAIR principles. The tools are continuously developed in collaboration with our users.



Data literacy from the start

Training and support for students, PhD candidates, and established researchers

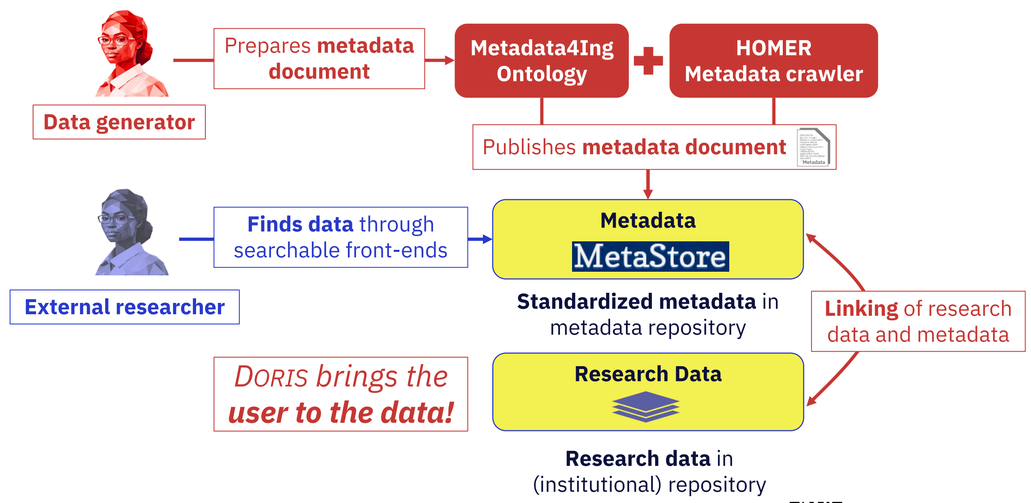
- **For Students:** An elective modules teaches RDM principles and best practices through expert-led sessions and guest lectures. Since 2022, it has introduced hundreds of students to RDM with high satisfaction. Additionally, a mandatory hands-on module at TU Munich integrates RDM into the B.Sc. Mechanical Engineering curriculum through practical lab experiments.
- **For Researchers:** General and domain-specific RDM workshops, including specialized training for High-Performance Computing, are offered in collaboration with HPC centers to meet the unique needs of the community.

We actively contribute to relevant committees and invite RDM education experts to collaborate with us!

Metadata Toolchain

Enabling FAIR metadata in the HPMC domain

Together with NFDI services, we can cover the entire **metadata workflow for engineers in the HPMC** domain - from data generation to publication - making HPMC research data findable and reusable.



A (sub-)ontology for engineers on HPC systems

The Metadata4Ing ontology provides a framework for describing scientific data generation. Developed by CFD and HPC experts from DORIS, the HPMC sub-ontology extends this for High-Performance Measurement and Computing, using a **modular approach to model processing steps, methods, and tools**.

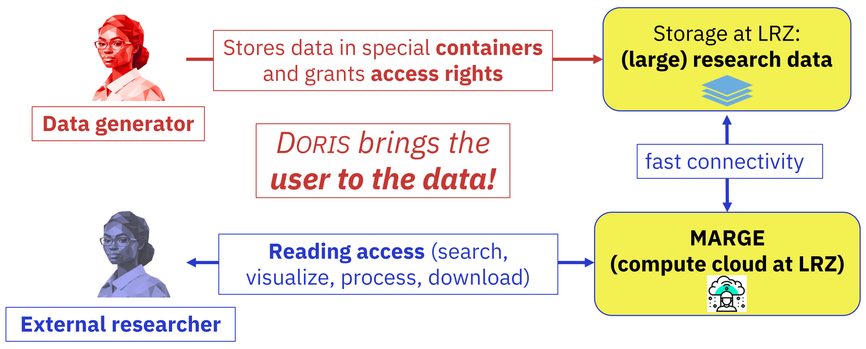
Metadata crawler HOMER

A HPMC tool for Ontology-based Metadata Extraction and Re-use that offers:
- **Ontology Parsing**: Reads .owl files and creates a property dictionary.
- **Metadata Generation**: Uses the dictionary to generate metadata files with user-defined input locations.
- **Flexibility**: No hard-coding; supports custom inputs.
- **Script Integration**: Easily integrates into scripts.
- **Continuous Development**: Regular improvements and updates.

Compute Cloud MARGE

The Multi-Access Research Gateway for HPC Experts

HPC data access and reuse are challenging due to size and hardware limits. MARGE enables direct access and analysis: **Remote access** allows accessing and re-using research data without HPC credentials. **Selective download** enables visualizing and extracting relevant data**. On-site processing** allows editing data without downloading. **VM deployment** allows running virtual machines for analysis, and fast access ensures quick connection to the HPC storage.



HPC container guide & support

Best practices to improve the reproducibility & reusability of research data in HPC
- **Tailored Containers**: Optimized for LRZ and JSC.
- **Expanding Support**: Developing solutions for HLRS and GPU applications.
- **Encapsulates applications** for seamless execution.
- **Ensures compatibility** across diverse computing environments.
- **Workflow Interoperability**

Charliecloud

Comprehensive support for RDM in HPC

Guides, Tools, and Services to Enhance RDM and FAIR Principles in HPC

DORIS offers a wide range of resources to enhance research data management and support the implementation of FAIR principles in HPC. These include **guides and templates** (e.g., an HPC-specific Data Management Plan template), **scientific publications** on RDM**, testing and comparison** of existing tools (e.g., for data transfer), and the **synchronization of workflows** across the Gauss Centers and, in the future, with NHR. Additionally, we **provide networking events**, **consultations**, and **support** for our target audience.

Visit our website and subscribe to our newsletter to stay updated!

Website *Newsletter*

Contact:
contact@nfdi4ing.de
https://nfdi4ing.de/

NFDI4ING is funded by the German Research Foundation (DFG) - project number 4421467

Benjamin Farnbacher,
Technical University of Munich,
https://orcid.org/0000-0002-1489-6501

Prof. Dr.-Ing. Christian Stemmer,
Technical University of Munich,
https://orcid.org/0000-0002-6904-8315

co-applicant institutions

Technische Universität Braunschweig · Leibniz Universität Hannover · Technische Universität München · KIT Karlsruher Institut für Technologie · Deutsches Zentrum für Luft- und Raumfahrt · Technische Universität Darmstadt · JÜLICH Forschungszentrum · TIB Leibniz Information Centre for Science and Technology University Library · RWTH Aachen University · Universität Stuttgart · Technische Universität Dresden