Chair of Aerodynamics and Fluid Mechanics **TUM School of Engineering and Design Technical University of Munich** 



# NFDI4Ing - DORIS – Data from High-Performance Measurement and Computation (HPMC)

# FAIR Scientific Data on HPMC Centers

Prof. Dr.-Ing. Christian Stemmer, Benjamin Farnbacher, Vasiliki Sdralia, Friedrich Ulrich | contact: info-doris@nfdi4ing.de

in collaboration with: Dr. Christine Wolter (TUM Library), Dr. Nadiia Huskova (HLRS), Dr. Stephan Hachinger and Johannes Munke (LRZ), JSC, RWTH Aachen (Chair of Fluid Mechanics and Institute of Aerodynamics)

# **Research Data Management with Large (HPMC) Data**



Doris is **conducting projects at HPMC facilities**. Engineers using these supercomputers mostly come from fluid mechanics, thermal and heat science, materials and construction engineering.

# **Our Tools**

• HOMER: a HPMC tool for Ontology-based Metadata Extraction and Re-use



#### What Are HPMC Research Data?

#### High Performance Measurement

- Measurement data
- Metadata (hardware, method, processing steps, etc.)

Analysis and processing of measurement data using HPC

#### FAIR Data Principles in HPMC

- Findable: storage in personalized accounts, little metadata
- Accessible: no access for third parties, insufficient transfer tools
- Interoperable: depending on formats and enriched meta
- Reusable: computing time at HPC centres required or virtualization



#### High Performance Computing

- Script / code
- Input file, output file, log file
- Raw data
- Processed data
- Metadata
- Data for secondary research (e.g. energy consumption in HPC)

#### **Ontology Parsing**

- Reads .owl files, creates property dictionary.

#### **Metadata Generation**

- Uses property dictionary to create accompanying metadata file.
- Allows user instructions for input location.

#### Flexibility

- Avoids hard-coding, accepts user-defined input locations.

#### **Script Integration**

- Designed for easy integration into scripts.



check out the git!

#### **Constant Development**

- Continuously improved and updated.
- Attended Cloud Housing: Compute Cloud server at LRZ for access and re-use of large HPMC data during the acceptance phase: access on request (contact us!)

#### Access to LRZ Data via NFDI4Ing Cloud

- Provides external user access to the
- Data and Science Storage (DSS) at LRZ.



### What Are We Trying to Accomplish?

Metadata Standards and Support to Data-Generating Groups for HPMC

- Establish metadata standards for HPMC environments.
- Convert metadata standards into a **semantic HMPC-sub-ontology** within the Metadata4Ing framework.
- Create and deploy a metadata crawler to extract ontologies, generate a dictionary, and populate metadata files.

#### **Reproducibility and Reusability of HPMC data**

- Promote data reuse projects in HPC centers.
- Provide virtual machine images of HPMC research data per Compute Cloud server (LRZ) and provide reduced datasets (HLRS)
- Evaluate container virtualization for HPC Center

#### **Storage, Access and Transfer**

- Attended Cloud Housing: Compute Cloud server at LRZ for access and re-use of large HPMC data
- Synchronization of encryption, encrypted workflows and other integrity and access management mechanisms within GCS centers

- Enables sharing of results through Paraview Visualizer.

#### **User-Friendly Analysis**

- External users can analyze data without harming it.
- Avoids the need for complex LRZ account setup.

Pilot User: Institute for Combustion Technology, RWTH Aachen

# HLRS JÜLICH

## HPMC Container Guide & Support

#### **HPC Containerization:**

- Tailored containers for LRZ (Charliecloud) and JSC (Singularity).
- Developing solution for University of Stuttgart.

#### Workflow Interoperability:

Performance Computing Systems

Engineering at TUM

- Encapsulates applications for seamless execution.
- Ensures compatibility across computing environments.

### What ELSE Do We Have to Offer?

Find out more on our **website** with the following content:



• Annual Ph.D.-workshop (October): Research Data Management for Science and



Our container guide!



- **Indexable** storage to front-end **interface** with **access management** for large data
- Goal: 3<sup>rd</sup> party users can access and process data **directly at HPC centres**



#### Acknowledgement

- The authors would like to thank the Federal Government and the Heads of Government of the Länder, as well as the Joint Science Conference (GWK), for their funding and support within the framework of the NFDI4Ing consortium. Funded by the German Research Foundation (DFG) - project number 442146713.
- The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. for funding this project by providing computing the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC).
- The authors gratefully acknowledge the Gauss Centre for Supercomputing e.V. for funding this project by providing computing time on the GCS Supercomputer SUPERMUC-NG at Leibniz Supercomputing Centre.
- The authors gratefully acknowledge the support of the Competence Network for Scientific High Performance Computing in Bavaria (KONWIHR) in this research.





Leibniz Supercomputing Centre of the Bavarian Academy of Sciences and Humanities



Subscribe to our newsletter for events and updates:



- **Lecture:** Introduction to Research Data Management for Munich Aerospace & Engineering Students (ED140003 / summer semester)
- **Metadata4Ing**: an **ontology** for describing the generation of research data within a scientific activity more information:

