

## Thesis / Interdisciplinary Project (IDP) / Research Practice / Study Project

# Test & Improvement of a Metadata-Crawler for HPC Simulations

for students within

## Informatics, Aerospace, Mechanical Engineering, Data Science

Our research group is working on solutions to make data findable and reusable. This is supported by attaching appropriate metadata to the produced research data. In this work, the metadata extraction is to be automated using a python based metadata crawler. The tool should first extract metadata from computational fluid dynamics (CFD) codes and then successively extend to data from the HPC system used, as well as post-processing algorithms. Therein, flexibility to work on inputs from other sources, e.g. other codes or experimental data is to be ensured. The goal is then to publish the metadata in the correct format and schema in a repository and to automate this process as far as possible.

### Tasks

- Improvement of an (existing) python toolkit and accompanying workflows to extract metadata
- Familiarization with metadata-ontologies
- Familiarization with CFD and HPC workflows
- Provide interoperable (HPC) metadata through a repository (MetaStore).



### Requirements

- Knowledge of Python and Linux-CL

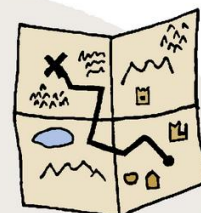
### Benefits

- Flexible working hours, remote work
- Insight into a nationwide, state-of-the-art research project
- Recognition within your study program (thesis / internships / projects etc.)
- Joint-mentoring by TUM and LRZ: exclusive experience with HPC-clusters

DATA



METADATA



<https://dataedo.com/cartoon/data-vs-metadata-2>

Protr@Dataedo

### Links

- Journal article: <https://preprints.inggrid.org/repository/view/12/>
- NFDI4Ing research group: <https://www.epc.ed.tum.de/en/aer/research-groups/nfdi4ing/>

### Contact

Benjamin Farnbacher  
benjamin.farnbacher@tum.de  
089.289.16094