



Bachelor Thesis:

Efficient Metadata Management in Industrial Data Lakes for High Pressure Die Casting

In the Internet of Things, and especially the Internet of Production, massive, heterogeneous streams of semi-structured data, such as sensor readings, process control data etc., are generated from different sources such as production machines and communicated via a wide range of protocols.

To persist this data and enable further analysis, raw data can be captured, processed, and stored in data lakes (Rudack et al., 2022). High Pressure Die Casting (HPDC) is a permanent mold based production technology commonly in use in the automotive industry to facilitate the high volume production of lightweight aluminum components for the body structure and the drive train.

While raw HPDC process data is abundant, extracting information from the heterogeneous production cell is challenging due to the lack of semantic interpretability and connected domain knowledge.

For the data to be useful and reusable later on, it needs to be annotated with rich metadata, describing its context, structure, and origin, for example by reusing and defining suitable ontologies. In order to persist process data and the related, semantic metadata in a reusable and interoperable way, the FactDAG model was recently introduced (Gleim et al., 2021). In this model, resources are globally and persistently identified and the corresponding data is versioned and linked to related resources using provenance information. An [open source implementation](#) of this model, the FactStack, was already developed as a result of previous work.

The goal of this thesis is to develop an ontology which enables semantic understanding of processes and reusability of its (meta-) data for an High Pressure Die Casting use case, based on the requirements from domain experts. Additionally, you should implement the capturing of metadata for industrial manufacturing in a functional demonstrator, combining the existing HPDC processing pipeline and the FactStack.

This thesis will be advised within the Cluster of Excellence "Internet of Production" by the Data Stream Management and Analysis Group (DSMA) in collaboration with the Foundry Institute (GI).

Your profile:

- Computer Science B.Sc. Student
- Interest in interdisciplinary work
- Good programming skills in at least one programming language
- Knowledge in Semantic Web Technologies / Ontologies (not required, but helpful)

Further information:

- M. Rudack, M. Rath et al., "Towards a Data Lake for High Pressure Die Casting," *Metals*, vol. 12, no. 2, p. 349, Feb. 2022, doi: [10.3390/met12020349](https://doi.org/10.3390/met12020349).
- L. Gleim, L. Tirpitz et al., "FactStack: Interoperable Data Management and Preservation for the Web and Industry 4.0," 2021, doi: [10.18420/BTW2021-20](https://doi.org/10.18420/BTW2021-20).

Interested? Questions? Contact Us!

Liam Tirpitz, M.Sc. – tirpitz@dbis.rwth-aachen.de – Tel: +49 241 80-21542

Maximilian Rudack, M.Sc. – m.rudack@gi.rwth-aachen.de – Tel: +49 241 80-95887