Digital Research Data Management in Materials Science and Engineering

Hossein Beygi Nasrabadi[1,*], Ebrahim Norouzi[1], Jörg Waitelonis[1], and Harald Sack[1,2]

[1]FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Eggenstein-Leopoldshafen, Germany
[2]Karlsruhe Institute of Technology, Institute of Applied Informatics and Formal Description Methods, Eggenstein-Leopoldshafen, Germany
[*]Corresponding author: hossein.beygi_nasrabadi@fiz-karlsruhe.de.

Introduction

The rapid digital transformation of research has profoundly changed the way materials science and engineering (MSE) generates, manages, and shares data [1]. Research Data Management (RDM) in MSE plays a central role in ensuring that complex and heterogeneous data, ranging from experimental measurements to computational simulations, remain findable, accessible, interoperable, and reusable (FAIR) [2]. Digitalization in this context refers to the systematic conversion of research workflows into machine-readable and semantically linked data ecosystems. In MSE, such digital infrastructures enable researchers to connect materials properties, experimental parameters, and processing conditions between institutions and disciplines [3]. A key enabler of digital RDM is the use of ontologies and knowledge graphs (KGs). Ontologies provide structured vocabularies and logical frameworks to describe entities, processes, and relationships in a standardized manner, while knowledge graphs integrate diverse datasets into an interconnected network [4]. Together, they form the semantic backbone for effectively managing and reusing MSE data. Existing tools such as electronic lab notebooks (ELNs), data repositories, and metadata portals can be connected through ontological models to enhance interoperability and long-term accessibility of research outcomes.

Objective

This study aims to strengthen digital RDM in MSE through three main goals: 1. Structuring research data types in materials science to identify key entities, relations, and metadata across experimental, computational, and organizational contexts. 2. Designing an ontology for RDM that provides a harmonized, BFO-compliant framework for representing MSE-specific research data and processes. 3. Constructing a knowledge graph (KG) that connects distributed datasets, enabling semantic integration and retrieval of materials data across diverse sources.

Methods

A community survey was conducted among MSE researchers to distinguish the most relevant categories of research data and related metadata requirements. Based on the survey findings, an ontology was developed following OBO Foundry best practices, reusing upper and mid-level ontologies such as BFO 2020 [5], PMDco [3], and NFDIcore [2] to ensure semantic interoperability. The ontology was populated using example datasets collected in spreadsheets, representing common MSE workflows such as MSE-related datasets, software, publications, etc. These structured examples were then translated into knowledge graph triples, connecting materials, instruments, processes, datasets, and publications. The resulting MSE Knowledge Graph (MSE-KG) was evaluated for machine readability, consistency, and data recovery using SPARQL queries derived from competency questions (e.g., "Which instruments were used for dataset X?" or "Which projects contributed to dataset Y?"). These queries verified the ontology's ability to retrieve information accurately and link related entities across the research lifecycle. List of relating competency questions and more details about ontology development approch can be find in our other paper [4].

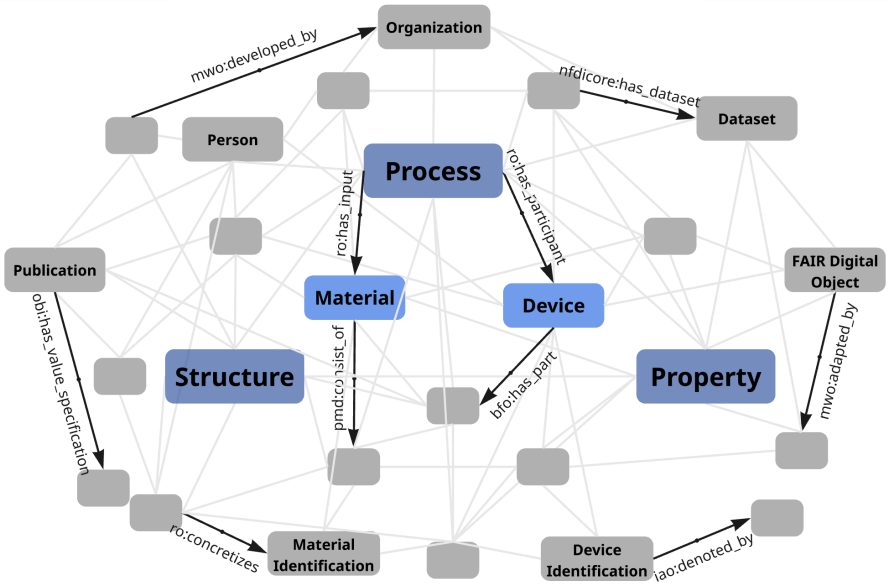

Figure 1: Semantic representation of MSE concepts in relation with RDM entities.

Results and discussion

The developed NFDI MatWerk ontology (MWO) provides a coherent and reusable semantic structure for representing MSE research data, combining elements of organizational management and experimental practice. It captures entities such as materials, specimens, instruments, and facilities, along with data products like datasets, workflows, and publications. MWO and its related resources can be accessed via the MWO GitHub repository (https://github.com/ISE-FIZKarlsruhe/mwo), Widoco documentation (https://ise-fizkarlsruhe.github.io/mwo/), and the ODK-based MkDocs documentation (https://ise-fizkarlsruhe.github.io/mwo/docs/). The resulting MSE-KG enables integrated exploration of data across various projects, supporting linking MSE researches (basically process-property-structure relationship) to the affecting materials and devices, as well as relating RDM metadata like relating datasets, publications, software, etc (Fig. 1). Evaluation results confirmed that the ontology-supported knowledge graph allows for efficient querying, logical reasoning, and automatic metadata generation. MSE-KG is also available for exploration and reuse at https://nfdi.fiz-karlsruhe.de/matwerk/.

Acknowledgments

The NFDI MatWek project is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the National Research Data Infrastructure – NFDI 38/1 – project number 460247524.

References

[1] H. Beygi Nasrabadi, T. Hanke, M. Weber, M. Eisenbart, F. Bauer, R. Meissner, G. Dziwis, L. Tikana, Y. Chen, and B. Skrotzki, "Toward a digital materials mechanical testing lab," *Computers in Industry*, vol. 153, p. 104016, 2023.

[2] O. Bruns, T. Tietz, J. Waitelonis, E. Posthumus, and H. Sack, "Nfdicore 2.0: A bfo-compliant ontology for multi-domain research infrastructures," *arXiv preprint arXiv:2410.01821*, 2024.

[3] B. Bayerlein, M. Schilling, H. Birkholz, M. Jung, J. Waitelonis, L. Mädler, and H. Sack, "Pmd core ontology: Achieving semantic interoperability in materials science," *Materials & Design*, vol. 237, p. 112603, 2024.

[4] H. Beygi Nasrabadi, E. Nourouzi, K. Hubaiev, J. Waitelonis, and H. Sack, "Nfdi matwerk ontology (mwo): A bfo-compliant ontology for research data management in materials science and engineering," *Advanced Engineering Materials*, 2026.

[5] I. ISO, "Iec 21838-2: 2021 information technology—top-level ontologies (tlo)—part 2: Basic formal ontology (bfo)," 2022.