# *Advancing Digital Workflows in Materials Science*

MATERIALDIGITAL

## The Role of PMDco in Data Integration and Semantic Representation

VMAP Standards Community e.V.

Markus Schilling
Bundesanstalt für Materialforschung und -prüfung (BAM)

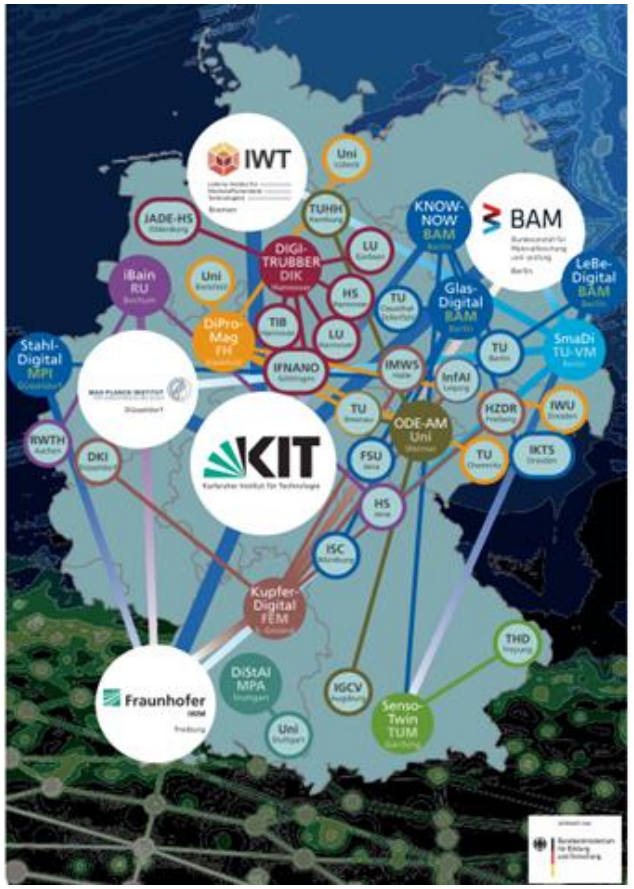
The platform for the digitalization of materials

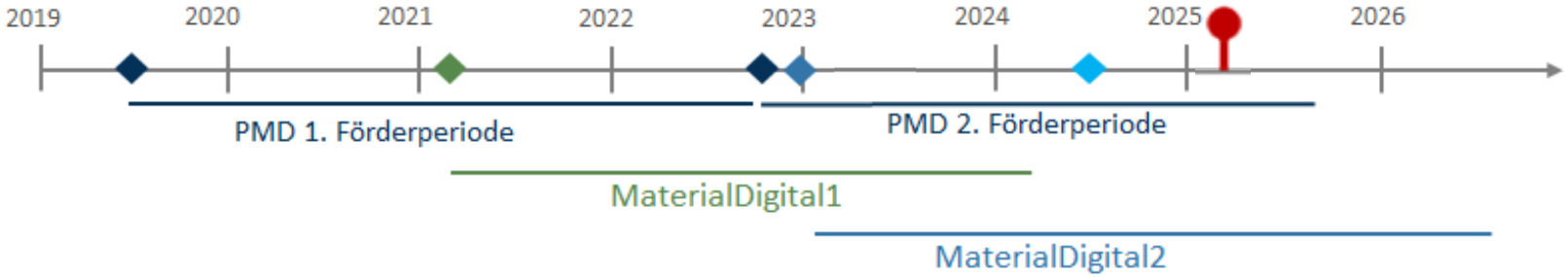
A joint project by:

KIT Karlsruher Institut für Technologie · BAM Bundesanstalt für Materialforschung und -prüfung · MAX-PLANCK-INSTITUT FÜR NACHHALTIGE MATERIALIEN GMBH · Fraunhofer IWM · IWT Leibniz-Institut für Werkstofforientierte Technologien · FIZ Karlsruhe Leibniz-Institut für Informationsinfrastruktur · Bundesministerium für Bildung und Forschung

# Plattform MaterialDigital (PMD)

- PMD Core Team + Partner Projects
- Partner Projects
  - MaterialDigital1: 13 Projects (53 different institutes)
  - MaterialDigital2: 10 Projects
  - MaterialDigital3: 8 Projects


- New Project Call: Material Neutral

# Semantic Technologies, Ontologies | T-Box, A-Box | Description Logics (DL)

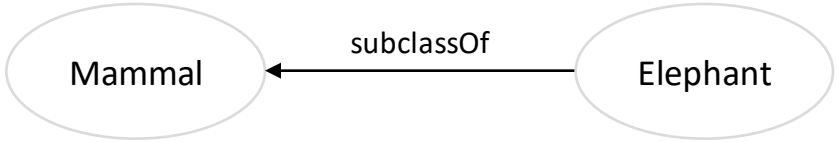
- understand and reuse data across **different contexts and applications → shared vocabulary** and meaning for concepts
- **standardized formats** allowing seamless data exchange
- Facilitating integration of **heterogeneous data sources** without requiring rigid schemas

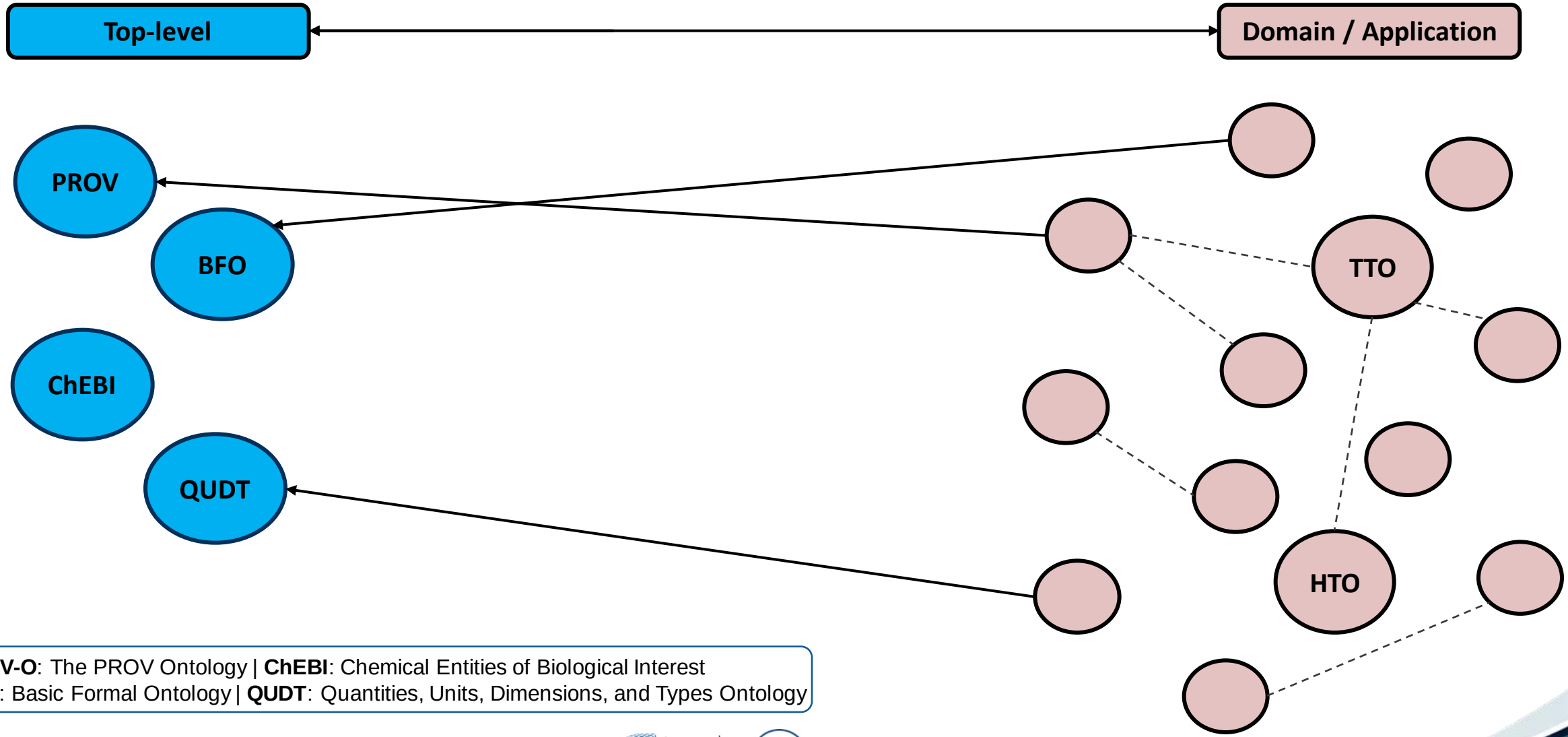
**T-Box** (*Terminological Box*)


- Knowledge about the **concepts of the domain** or field regarded
- Defines which **classes** of objects exist and their **properties**
- Does not change frequently

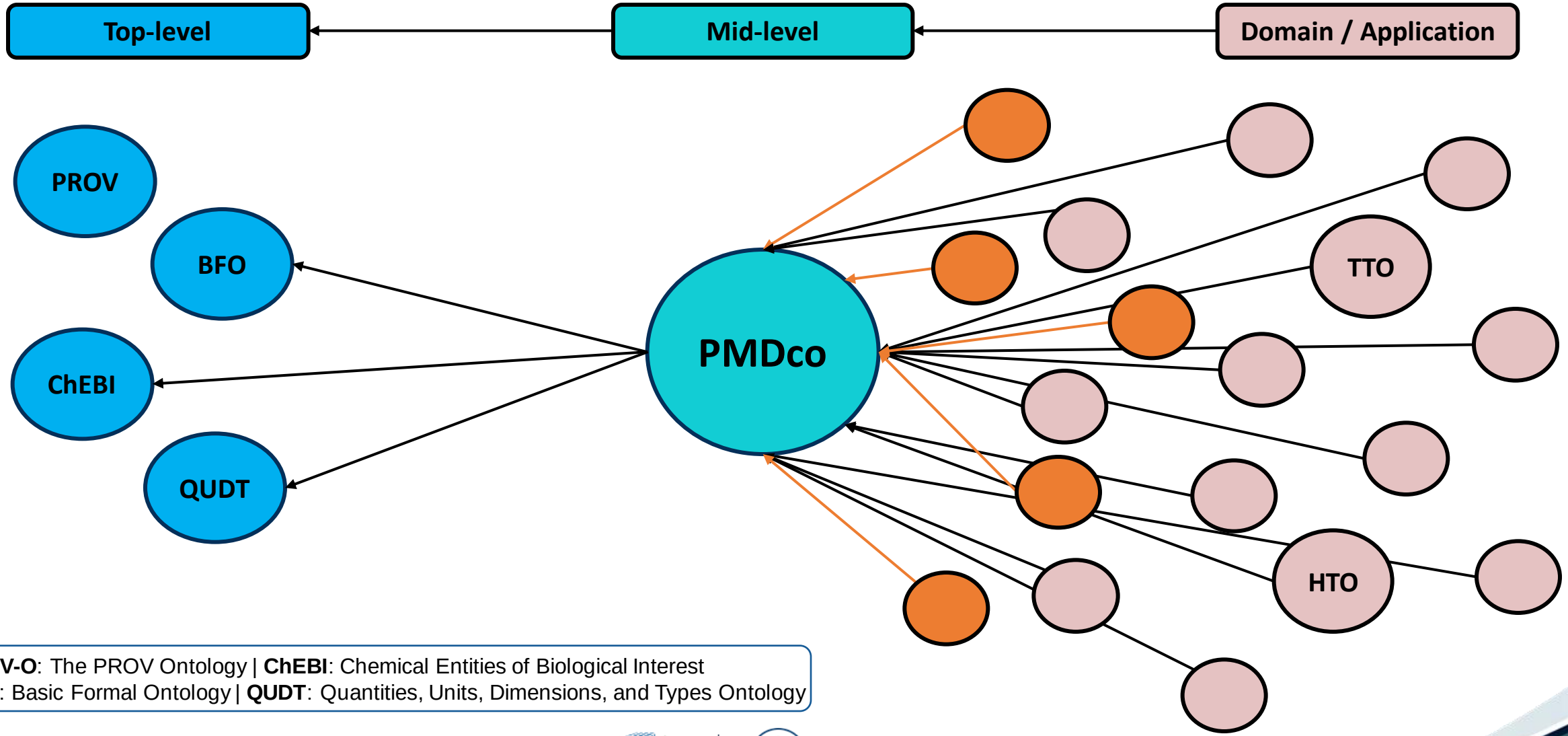
**A-Box** (*Assertional Box*)


- Knowledge about **specific instances** of a domain
- Facts about individuals and relationships
- Represents the state of the modeled world → **data**
- Subject to frequent or constant changes




**PROV-O**: The PROV Ontology | **ChEBI**: Chemical Entities of Biological Interest
**BFO**: Basic Formal Ontology | **QUDT**: Quantities, Units, Dimensions, and Types Ontology




**PROV-O**: The PROV Ontology | **ChEBI**: Chemical Entities of Biological Interest
**BFO**: Basic Formal Ontology | **QUDT**: Quantities, Units, Dimensions, and Types Ontology

# Establish a Semantic Framework for MSE with PMD Core Ontology (PMDco)

Materials & Design 237 (2024) 112603

Contents lists available at ScienceDirect

Materials & Design

journal homepage: www.elsevier.com/locate/matdes

PMD Core Ontology: Achieving semantic interoperability in materials science

Bernd Bayerlein [a,*], Markus Schilling [a], Henk Birkholz [b], Matthias Jung [c], Jörg Waitelonis [d], Lutz Mädler [b,e], Harald Sack [d]

[a] Bundesanstalt für Materialforschung und -prüfung (BAM), Unter den Eichen 87, Berlin, 12205, Germany
[b] Leibniz-Institute for Materials Engineering (IWT), Badgasteiner Str. 3, Bremen, 28359, Germany
[c] Fraunhofer Institute for Mechanics of Materials (IWM), Wöhlerstr. 11, Freiburg im Breisgau, 79108, Germany
[d] FIZ Karlsruhe - Leibniz Institute for Information Infrastructure, Hermann-von-Helmholtz-Platz 1, Eggenstein-Leopoldshafen, 76344, Germany
[e] University of Bremen, Badgasteiner Straße 3, Bremen, 28359, Germany

ARTICLE INFO

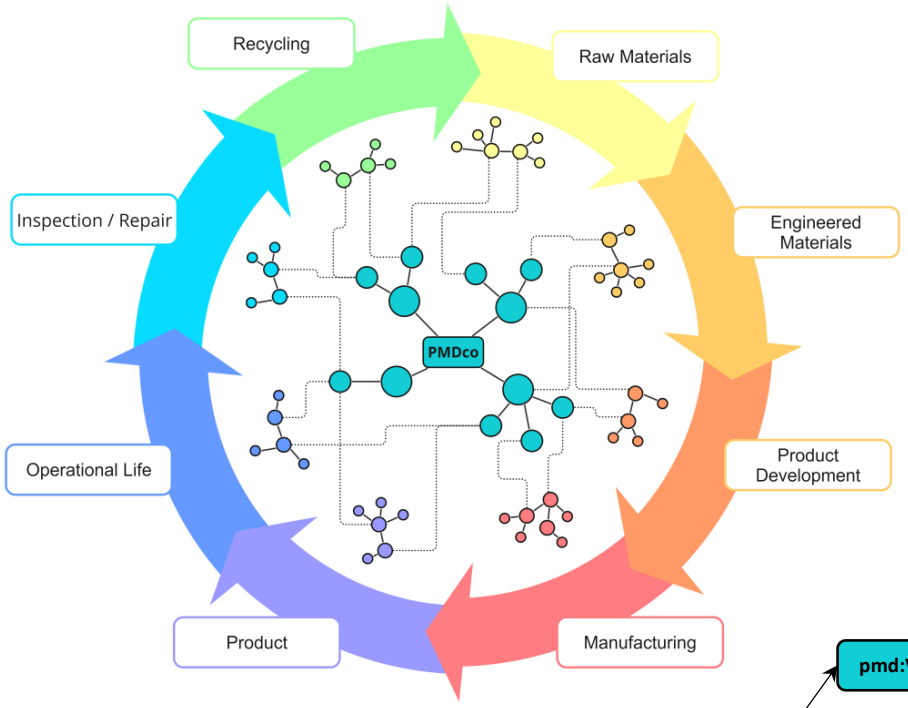
Dataset link: https://github.com/materialdigital/core-ontology

*Keywords:*
Ontology
Materials science and engineering
Knowledge representation
Reproducibility
Semantic interoperability
Semantic data integration

ABSTRACT

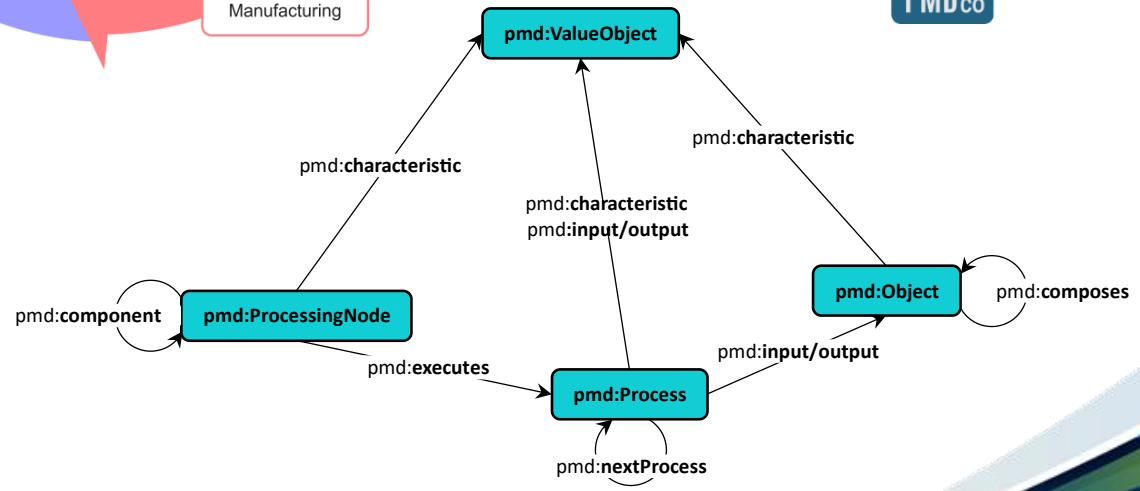
https://doi.org/10.1016/j.matdes.2023.112603

Knowledge representation in the Materials Science and Engineering (MSE) domain is a vast and multi-faceted challenge: Overlap, ambiguity, and inconsistency in terminology are common. Invariant (consistent) and variant (context-specific) knowledge are difficult to align cross-domain. Generic top-level semantic terminology often is too abstract, while MSE domain terminology often is too specific. In this paper, an approach how to maintain a comprehensive MSE-centric terminology composing a mid-level ontology–the Platform MaterialDigital Core Ontology (PMDco)–via MSE community-based curation procedures is presented. The illustrated findings show how the PMDco bridges semantic gaps between high-level, MSE-specific, and other science domain semantics. Additionally, it demonstrates how the PMDco lowers development and integration thresholds. Moreover, the research highlights how to fuel it with real-world data sources ranging from manually conducted experiments and simulations with continuously automated industrial applications.




GitHub

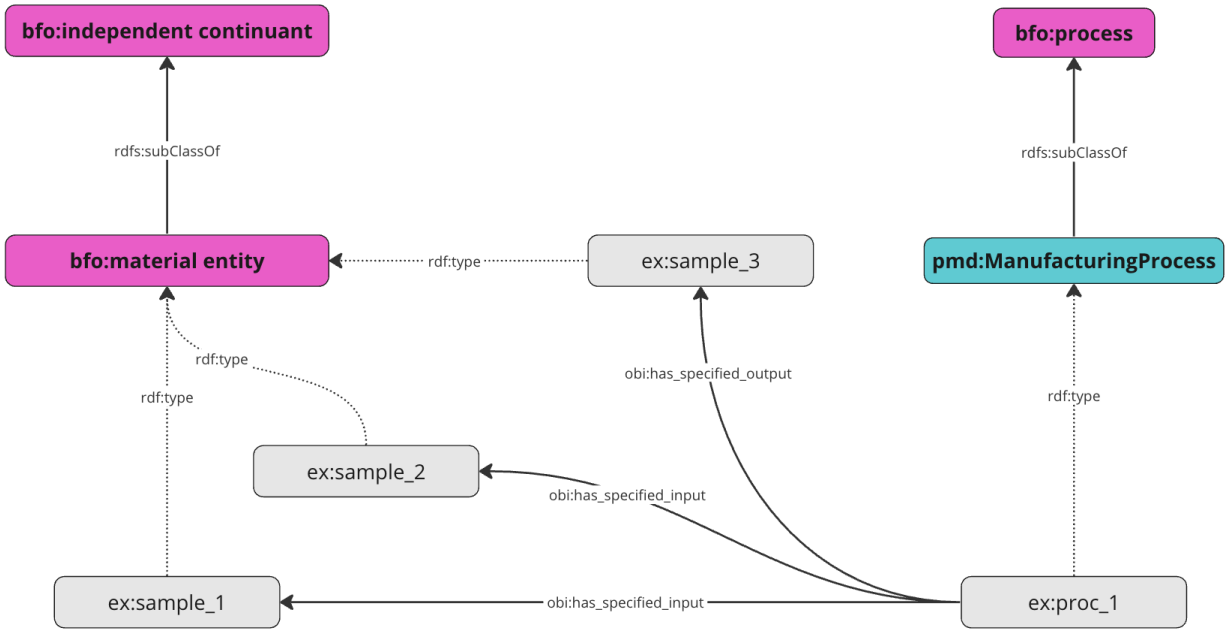
PMDco



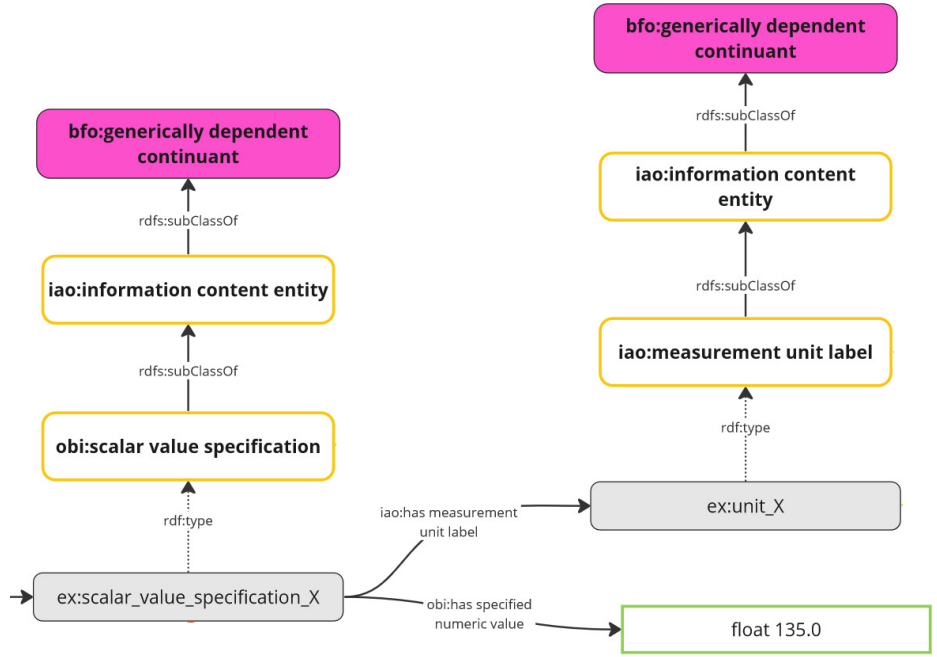

- Make materials and processes data interoperable
- Support FAIR principles and foster reproducibility
- Facilitate traceability of (meta)data throughout value chains
- Bridge semantic gaps in MSE and establish a shared vocabulary

# Systematic and Uniform (Re)Usage via Semantic Graph Patterns

- **Consistency & Reusability**: standardized structures creating uniformity – integrate, extend, and reuse
- **Enhanced Querying & Reasoning**: improve automated reasoning, data linking, and querying efficiency



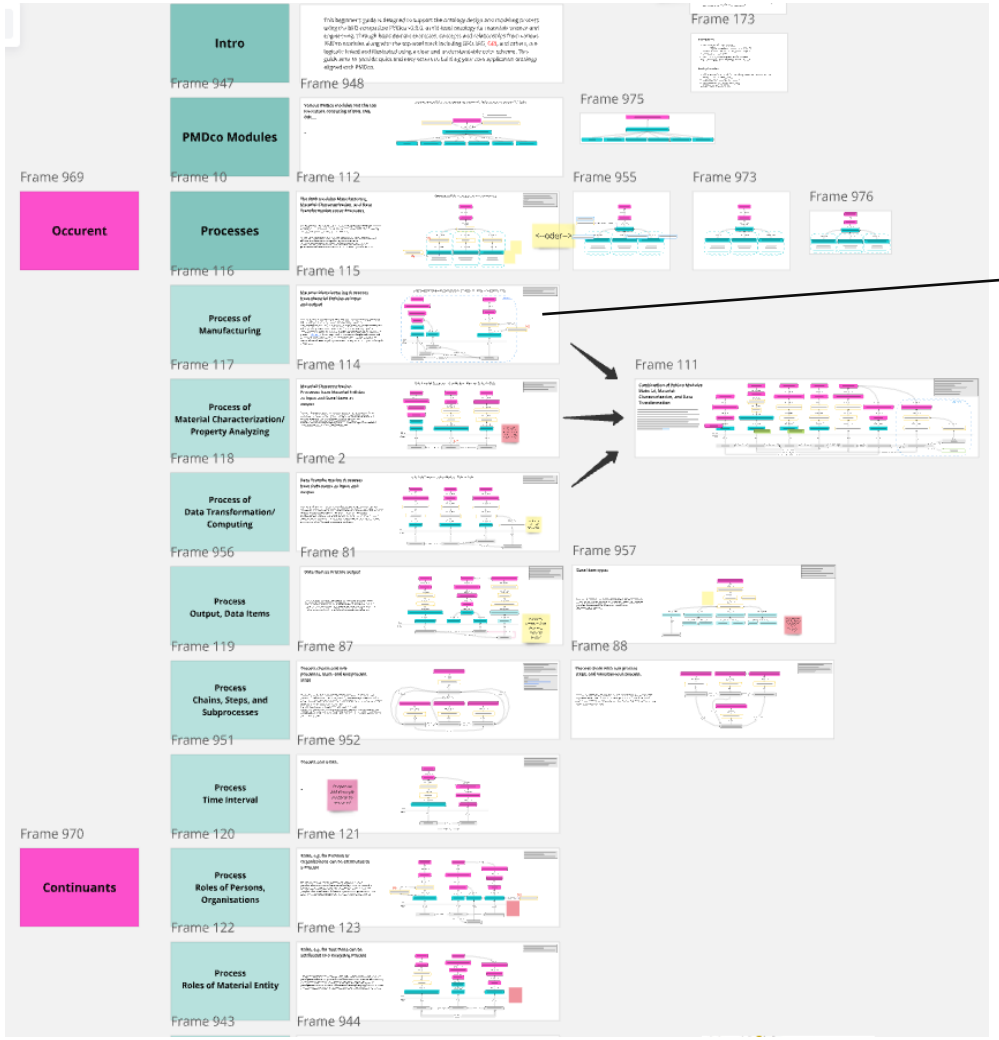

Pattern 1: input and output of processes

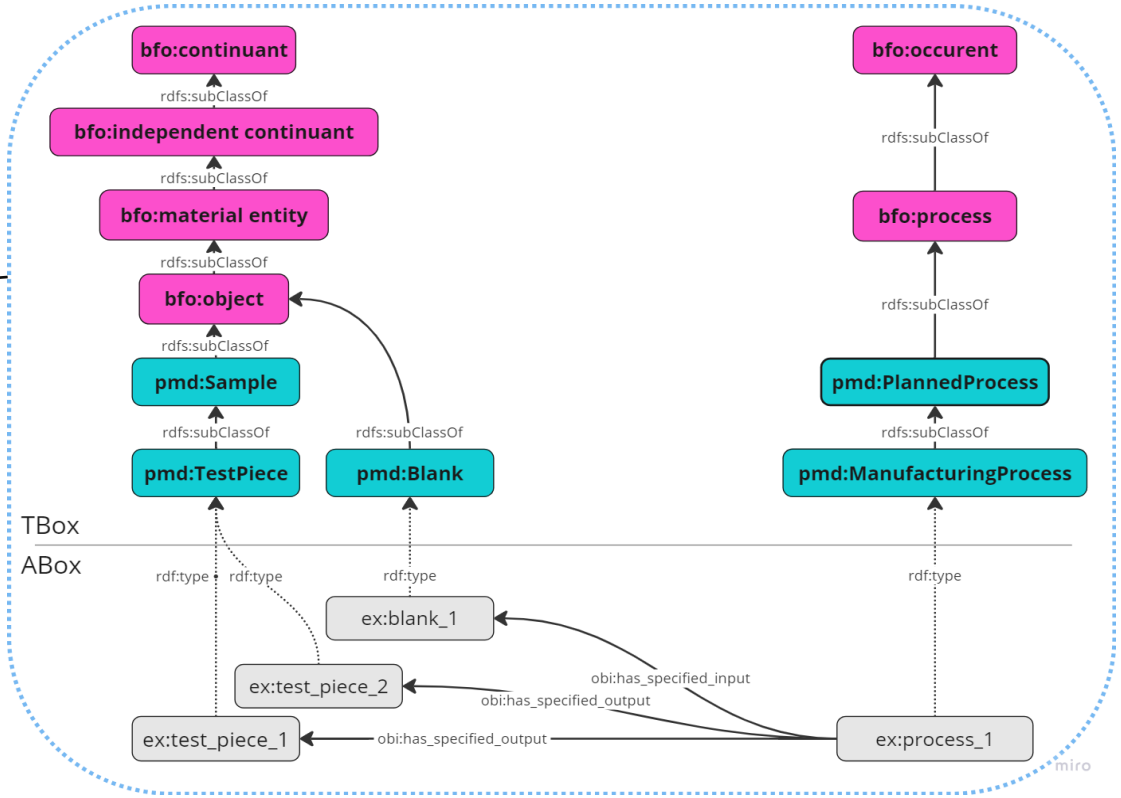



Pattern 2: scalar value specification with value and measurement unit

(multi)colored: T-Box concepts / classes | grey: A-Box instances

# Systematic and Uniform (Re)Usage via Semantic Graph Patterns

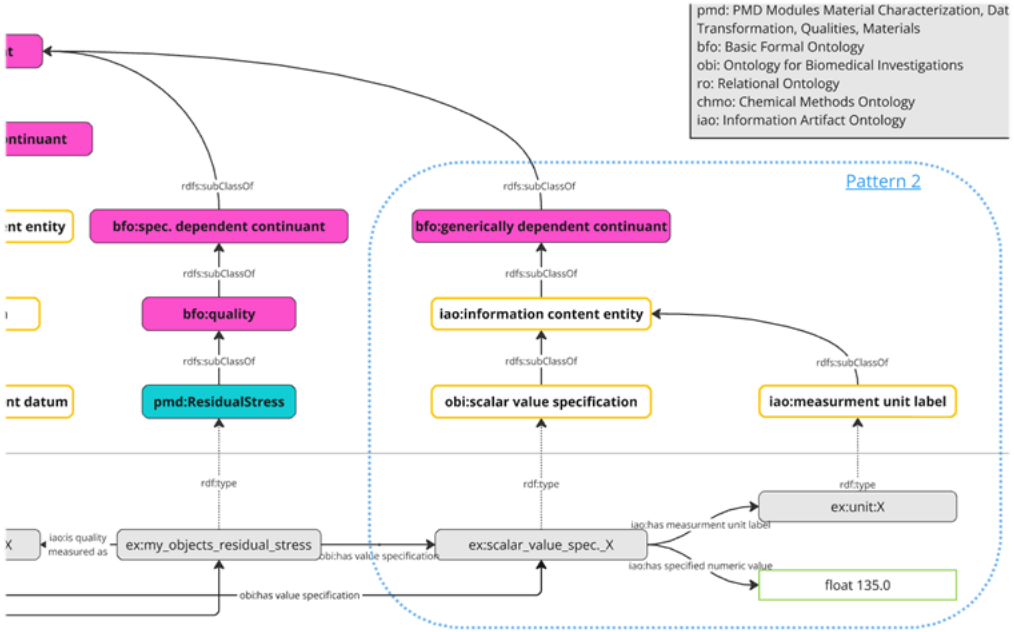


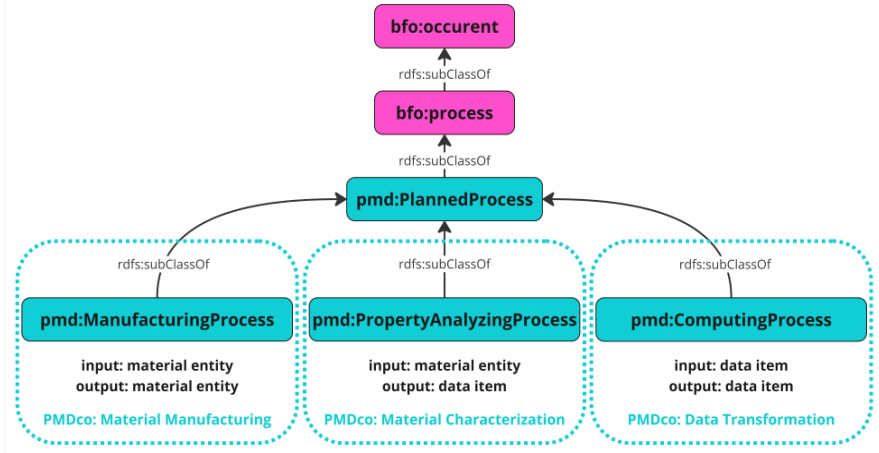



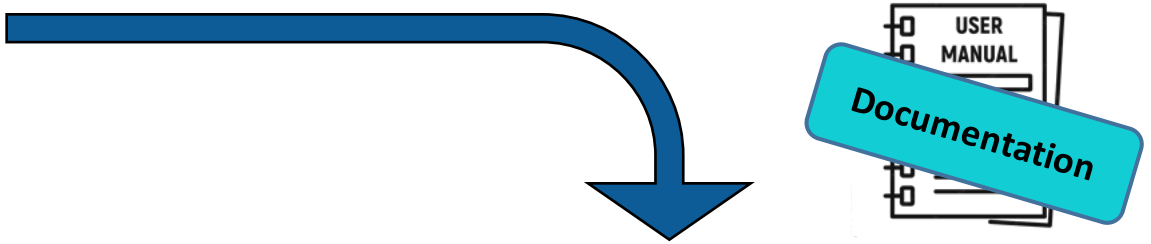

**Guide with essential ontology design patterns**

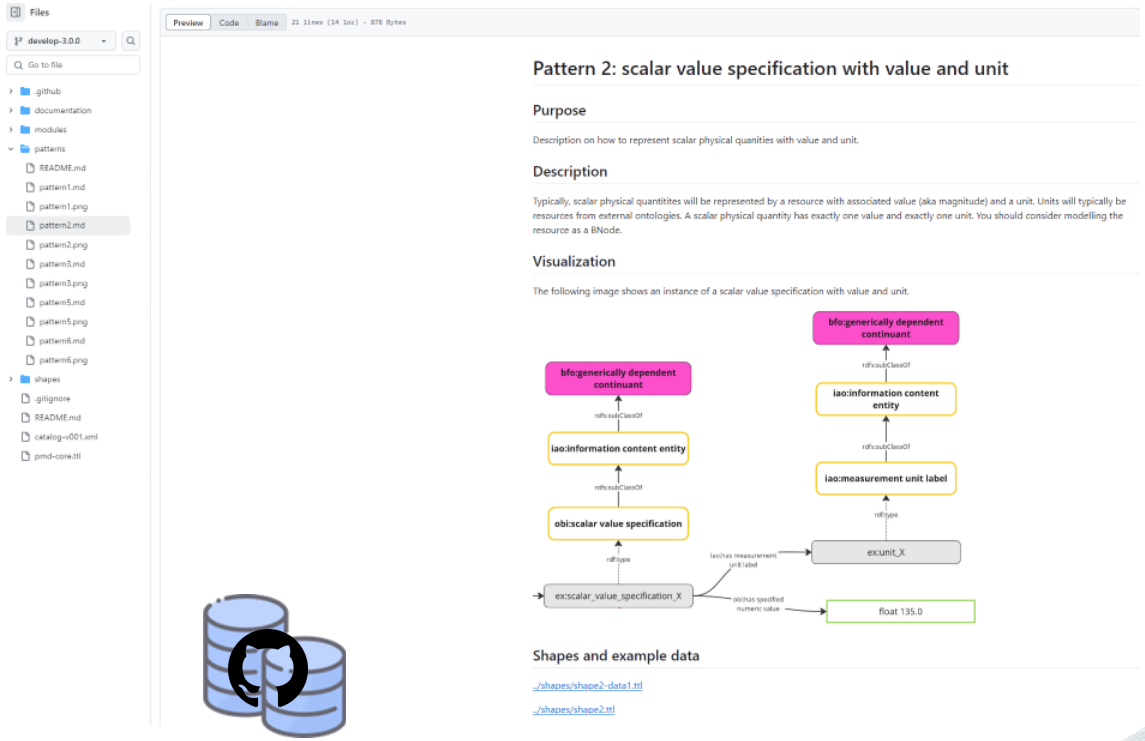
- GitHub repository: Hosts patterns and functional semantic shapes
- SPARQL queries: Shape-specific queries provided

# Systematic and Uniform (Re)Usage via Semantic Graph Patterns



# Community Effort | PMD Ontology Playground

**Bi-weekly public sessions**

*Fridays, 1-2 pm*

**Audience**

- Ontology practitioners & MSE domain experts
- Interested parties

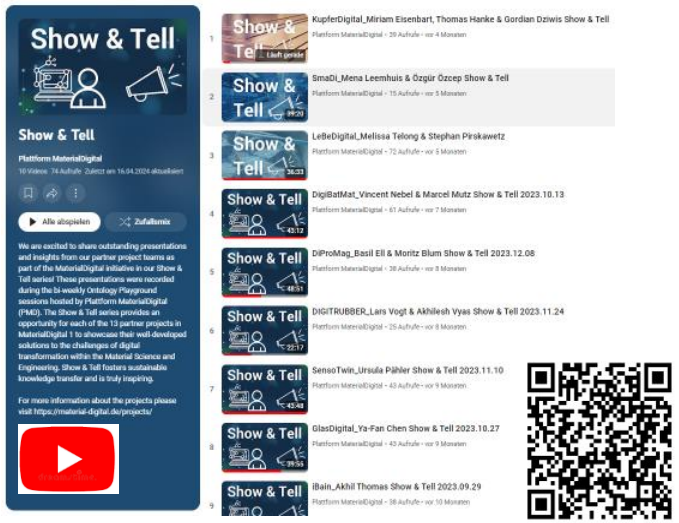
**Key activities**

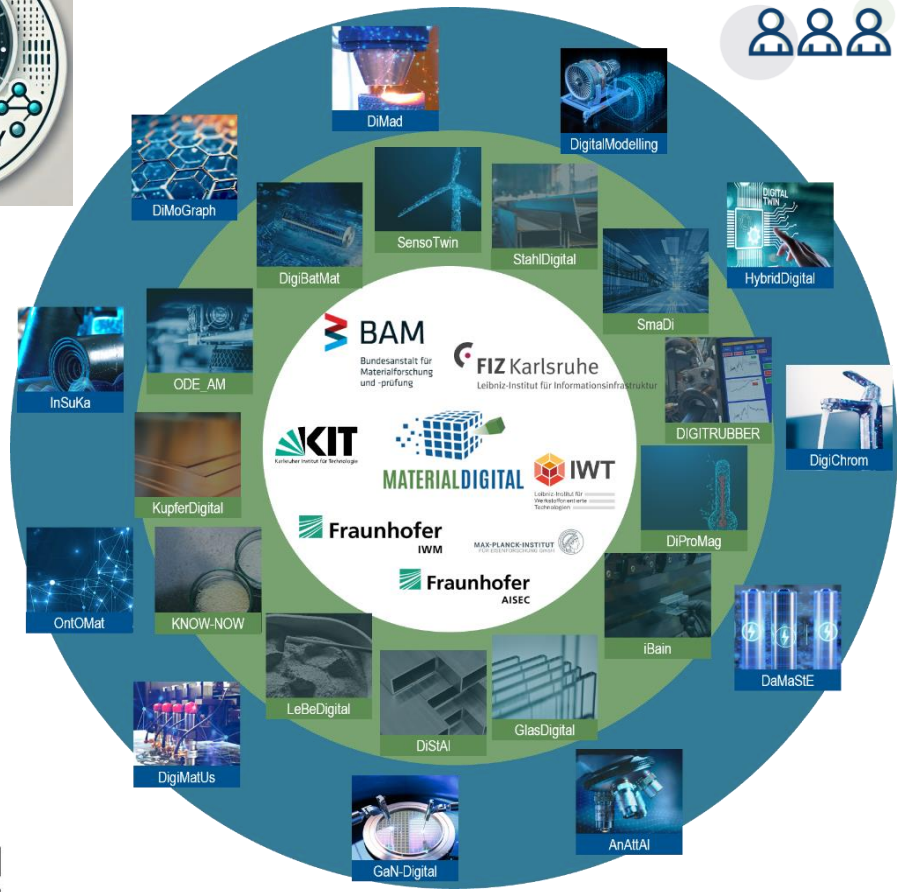
- Knowledge transfer
- Experience exchange
- Modeling challenges – User insights
- Collaborative PMDco enhancement

**MS Teams Meeting**

- Easy Access
- No Obligation

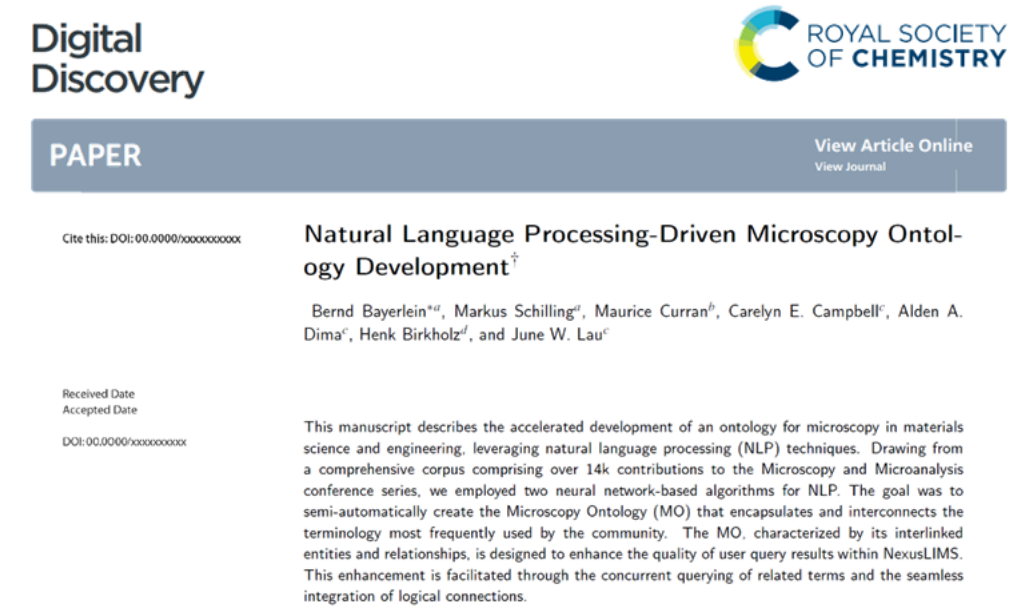
*Ontology Playground Information List*


B. Bayerlein, PMD General Assembly @ BAM, 09/18/2024, Berlin, Germany

# NLP-Driven Microscopy Ontology Development



Digital Discovery

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal

Cite this: DOI: 00.0000/xxxxxxxxxx

## Natural Language Processing-Driven Microscopy Ontology Development†

Bernd Bayerlein*[a], Markus Schilling[a], Maurice Curran[b], Carelyn E. Campbell[c], Alden A. Dima[c], Henk Birkholz[d], and June W. Lau[c]

Received Date
Accepted Date

DOI: 00.0000/xxxxxxxxxx

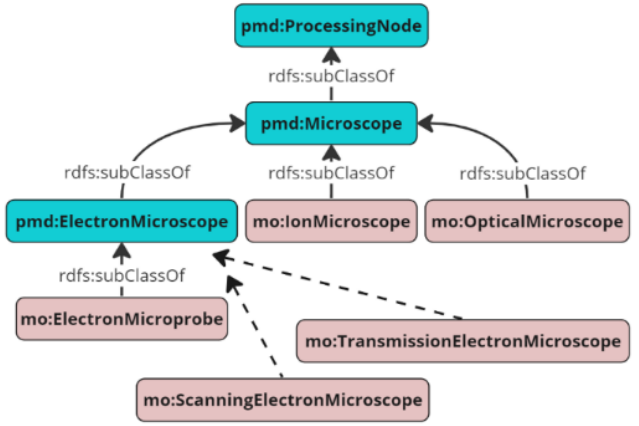
This manuscript describes the accelerated development of an ontology for microscopy in materials science and engineering, leveraging natural language processing (NLP) techniques. Drawing from a comprehensive corpus comprising over 14k contributions to the Microscopy and Microanalysis conference series, we employed two neural network-based algorithms for NLP. The goal was to semi-automatically create the Microscopy Ontology (MO) that encapsulates and interconnects the terminology most frequently used by the community. The MO, characterized by its interlinked entities and relationships, is designed to enhance the quality of user query results within NexusLIMS. This enhancement is facilitated through the concurrent querying of related terms and the seamless integration of logical connections.

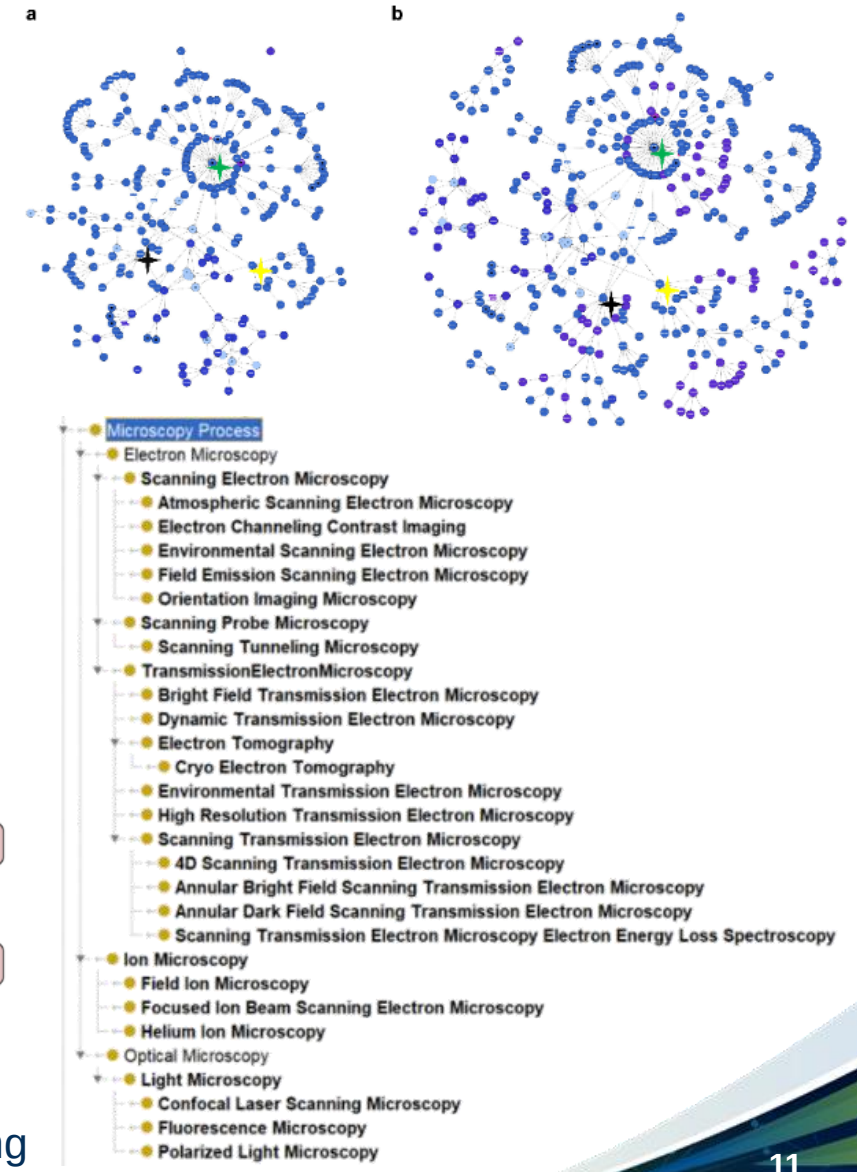
BAM

NIST

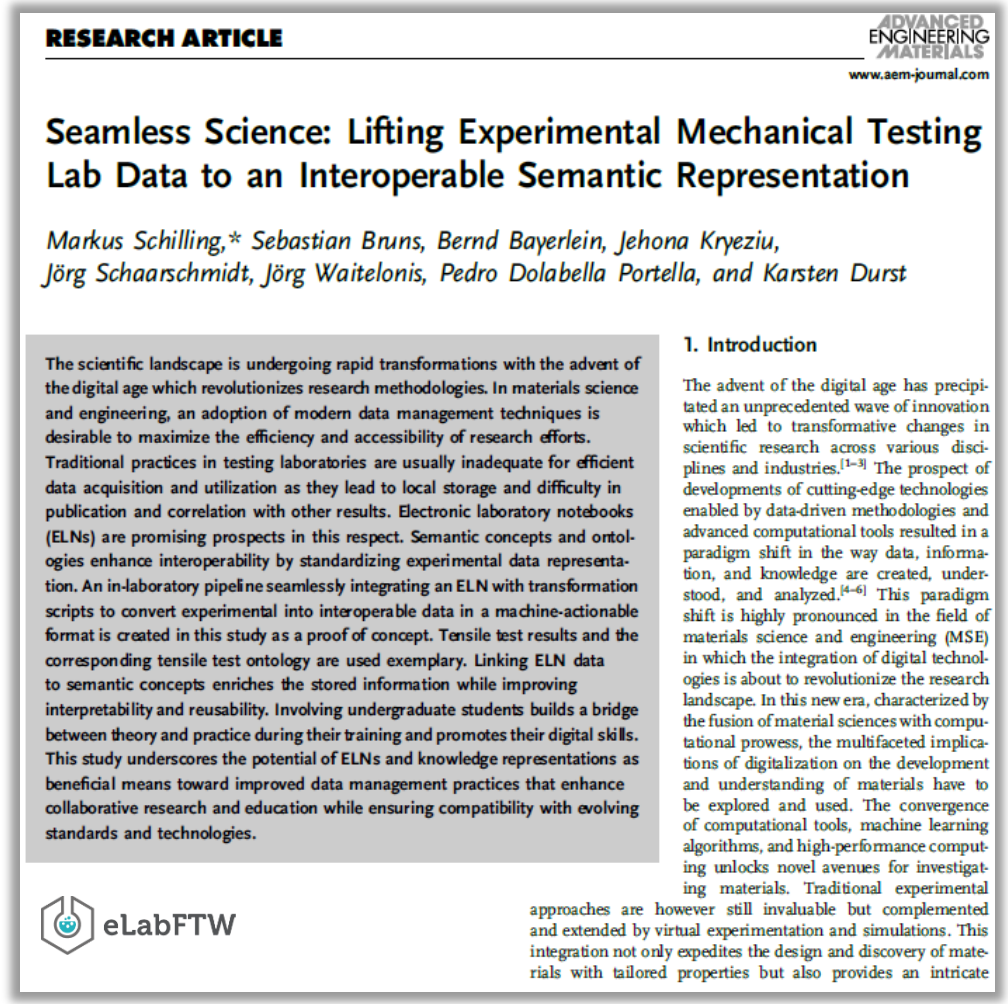
IWT
Leibniz-Institut für Werkstofforientierte Technologien

- NLP-approach based on huge text corpus of Microscopy & Microanalysis Conference Proceedings
- Extracted key terms used by community categorized according to PMDco structure
- Results in Microscopy Ontology with ~150 new classes and ~60 abbreviations



# ELN Data Acquisition Demonstrator



RESEARCH ARTICLE

ADVANCED ENGINEERING MATERIALS

www.aem-journal.com

## Seamless Science: Lifting Experimental Mechanical Testing Lab Data to an Interoperable Semantic Representation

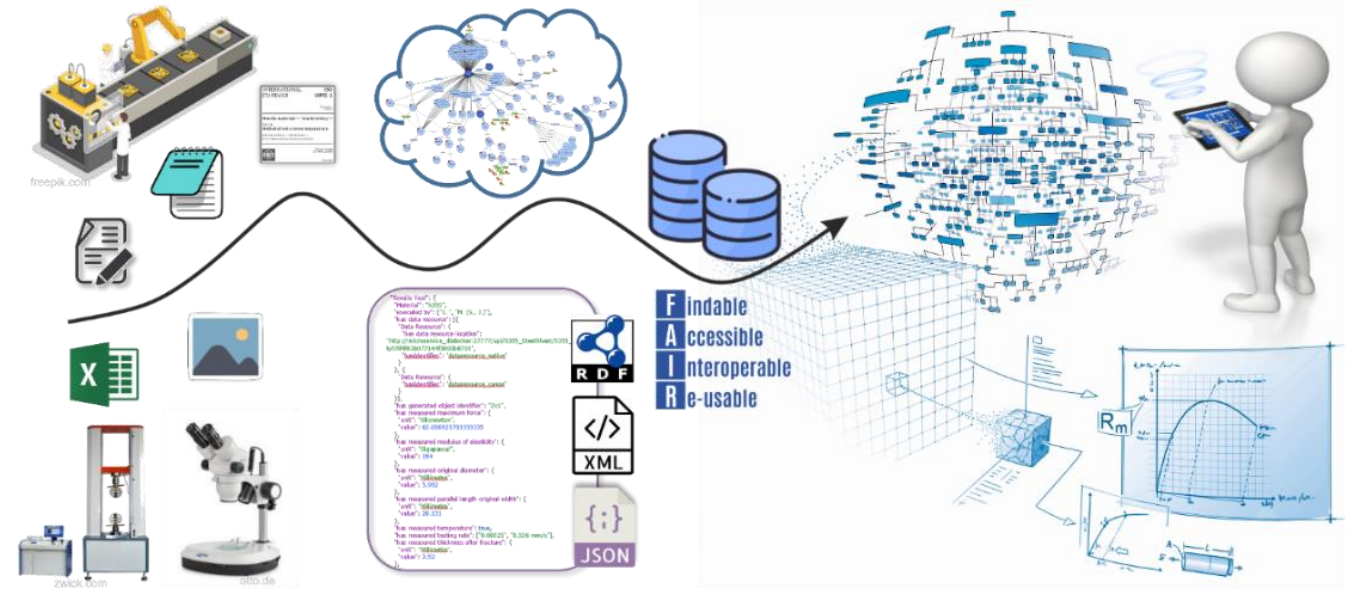
*Markus Schilling,* Sebastian Bruns, Bernd Bayerlein, Jehona Kryeziu, Jörg Schaarschmidt, Jörg Waitelonis, Pedro Dolabella Portella, and Karsten Durst*

The scientific landscape is undergoing rapid transformations with the advent of the digital age which revolutionizes research methodologies. In materials science and engineering, an adoption of modern data management techniques is desirable to maximize the efficiency and accessibility of research efforts. Traditional practices in testing laboratories are usually inadequate for efficient data acquisition and utilization as they lead to local storage and difficulty in publication and correlation with other results. Electronic laboratory notebooks (ELNs) are promising prospects in this respect. Semantic concepts and ontologies enhance interoperability by standardizing experimental data representation. An in-laboratory pipeline seamlessly integrating an ELN with transformation scripts to convert experimental into interoperable data in a machine-actionable format is created in this study as a proof of concept. Tensile test results and the corresponding tensile test ontology are used exemplary. Linking ELN data to semantic concepts enriches the stored information while improving interpretability and reusability. Involving undergraduate students builds a bridge between theory and practice during their training and promotes their digital skills. This study underscores the potential of ELNs and knowledge representations as beneficial means toward improved data management practices that enhance collaborative research and education while ensuring compatibility with evolving standards and technologies.

eLabFTW

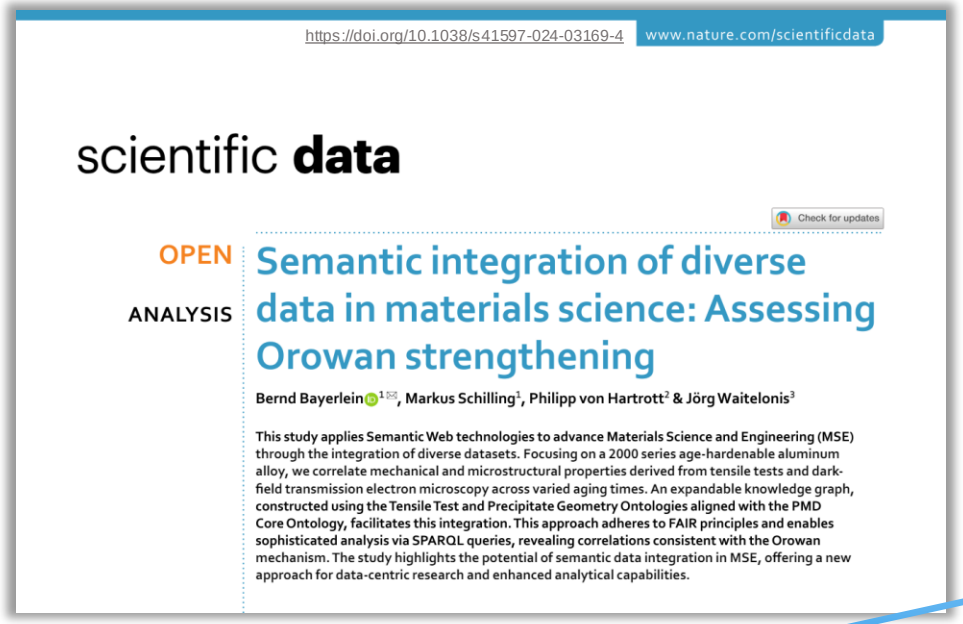
### 1. Introduction

The advent of the digital age has precipitated an unprecedented wave of innovation which led to transformative changes in scientific research across various disciplines and industries.[1–3] The prospect of developments of cutting-edge technologies enabled by data-driven methodologies and advanced computational tools resulted in a paradigm shift in the way data, information, and knowledge are created, understood, and analyzed.[4–6] This paradigm shift is highly pronounced in the field of materials science and engineering (MSE) in which the integration of digital technologies is about to revolutionize the research landscape. In this new era, characterized by the fusion of material sciences with computational prowess, the multifaceted implications of digitalization on the development and understanding of materials have to be explored and used. The convergence of computational tools, machine learning algorithms, and high-performance computing unlocks novel avenues for investigating materials. Traditional experimental approaches are however still invaluable but complemented and extended by virtual experimentation and simulations. This integration not only expedites the design and discovery of materials with tailored properties but also provides an intricate




- Used by students during internship at TU Darmstadt
- Data acquisition pipeline supported by ontological framework
- Script-based automatic RDF transformation
- Growing graph over years with varying parameters

# Harmonized Diverse Datasets: Orowan Demonstrator



https://doi.org/10.1038/s41597-024-03169-4 www.nature.com/scientificdata

scientific data

Check for updates

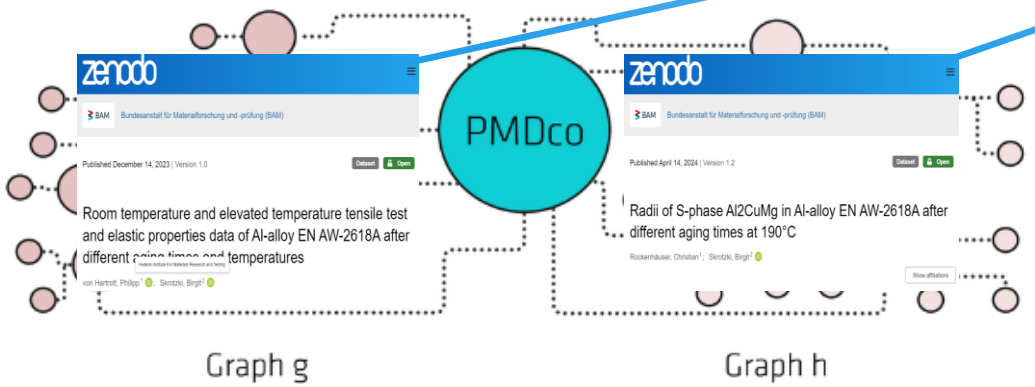
OPEN

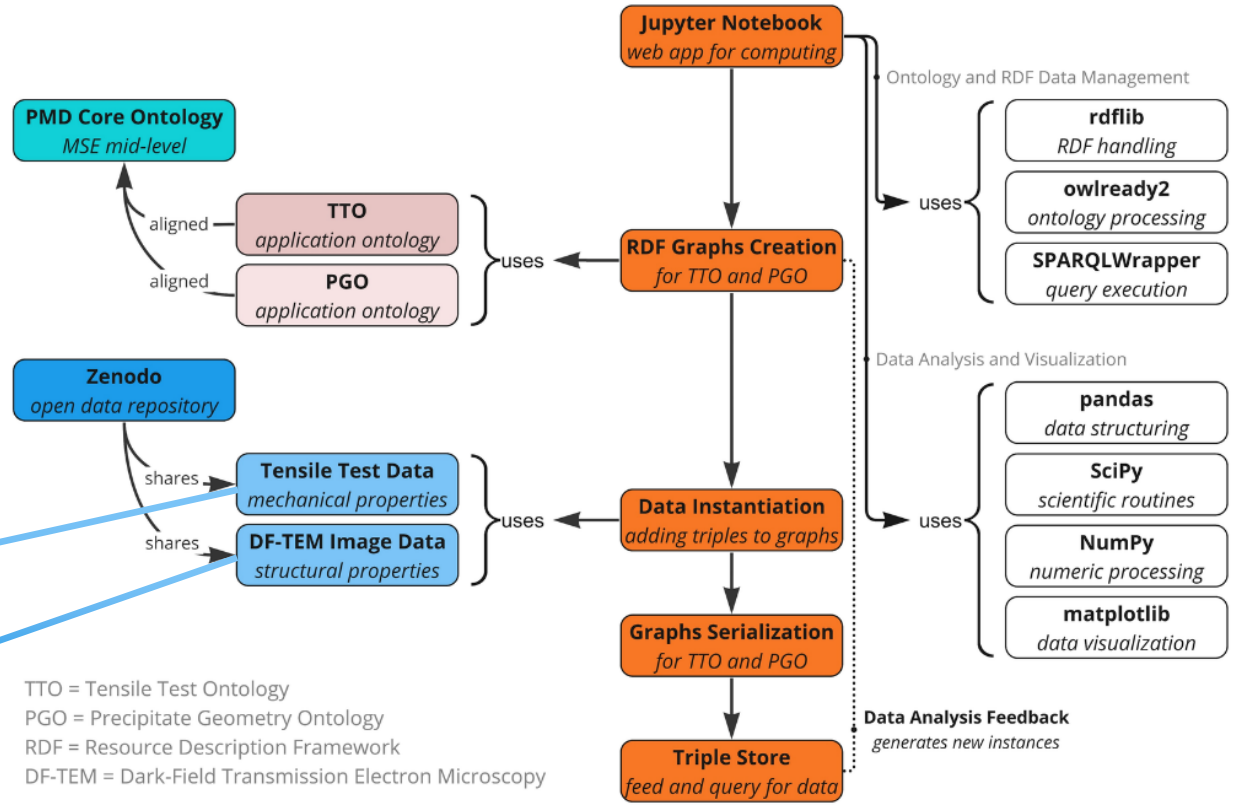
ANALYSIS

## Semantic integration of diverse data in materials science: Assessing Orowan strengthening

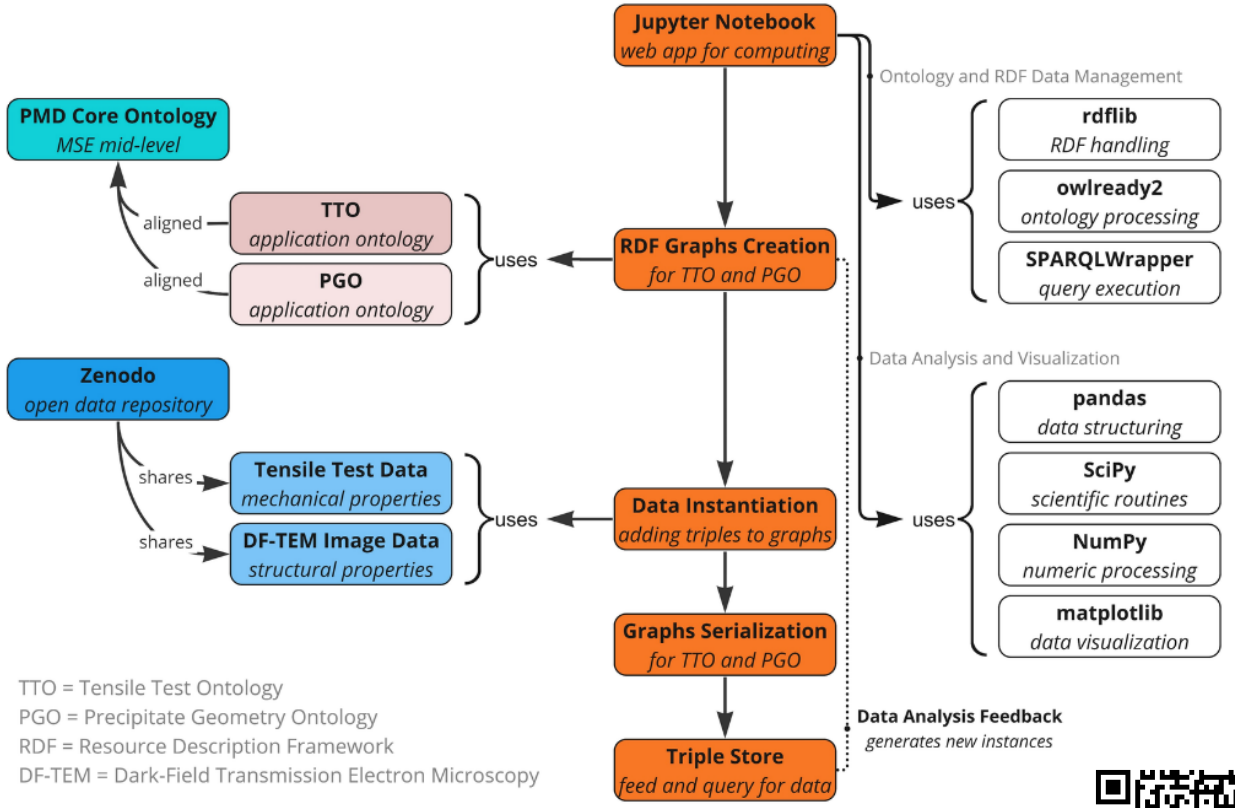
Bernd Bayerlein[1✉], Markus Schilling[1], Philipp von Hartrott[2] & Jörg Waitelonis[3]

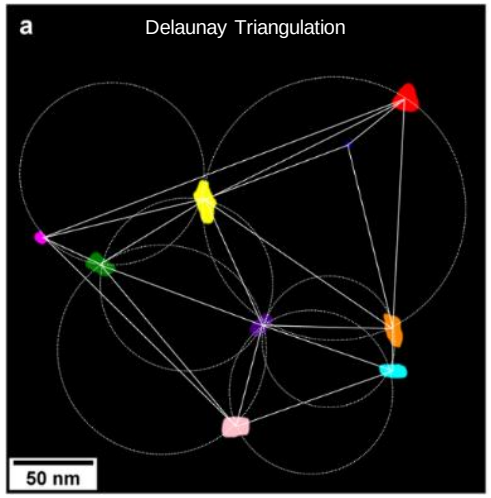
This study applies Semantic Web technologies to advance Materials Science and Engineering (MSE) through the integration of diverse datasets. Focusing on a 2000 series age-hardenable aluminum alloy, we correlate mechanical and microstructural properties derived from tensile tests and dark-field transmission electron microscopy across varied aging times. An expandable knowledge graph, constructed using the Tensile Test and Precipitate Geometry Ontologies aligned with the PMD Core Ontology, facilitates this integration. This approach adheres to FAIR principles and enables sophisticated analysis via SPARQL queries, revealing correlations consistent with the Orowan mechanism. The study highlights the potential of semantic data integration in MSE, offering a new approach for data-centric research and enhanced analytical capabilities.

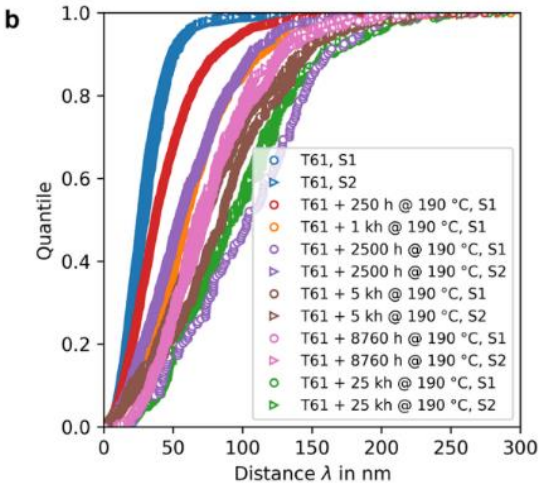


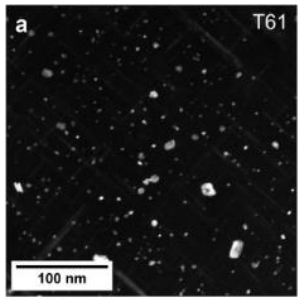



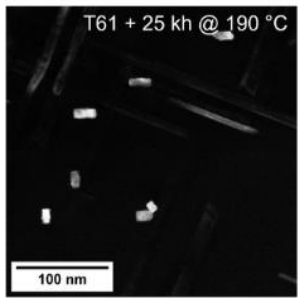
# Harmonized Diverse Datasets: Orowan Demonstrator

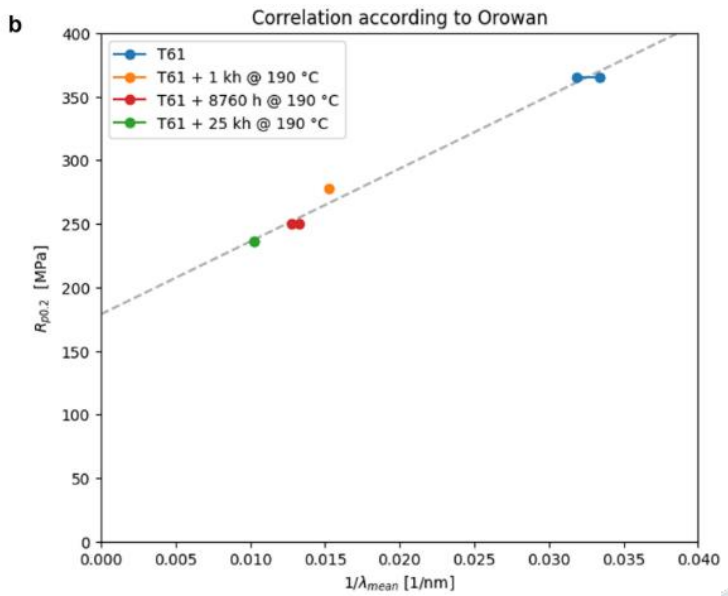



- Semantic data integration from diverse data sources
- Demonstrate knowledge graph operations
- Proof of concept for semantic interoperability
- Prototypical representation of MSE knowledge in RDF/OWL

# PMD Project Team

BAM

Fraunhofer IWM

IWT
Leibniz-Institut für Werkstofforientierte Technologien

FIZ Karlsruhe
Leibniz-Institut für Informationsinfrastruktur

MAX-PLANCK-INSTITUT FÜR NACHHALTIGE MATERIALIEN GMBH

Fraunhofer AISEC

KIT
Karlsruher Institut für Technologie

## Contact and join us!

info@material-digital.de

PMD Core Ontology (PMDco)