Ontology Based Knowledge System for Ceramic Multi-Layer Components

Sahar Ben Hassine[1], Rainer Stark[1]

[1]Department of Industrial Information Technology - Technical University of Berlin (TU Berlin), Germany

Abstract

This paper introduces the project "Ceramic multi-layer development through redesign of ontology-based knowledge systems" (Know-Now), which aims to develop an adaptable ontology for ceramic multi-layer technology. The primary objective is to link material-related and technological data automatically with simulations in order to solve engineering problems. The presented work is based on the principles of information science, in particular ontologies, to address the challenge of systematic data acquisition and knowledge retrieval in the manufacturing process. By applying ontologies, a structured and self-explanatory framework is created for various data related to powder preparation, grinding, milling, calcining, tape casting, stamping, stacking, laminating and sintering. The approach developed in the project not only enables efficient data refinement, but also serves as a basis for knowledge-based decision-making to optimize the manufacturing process of multi-layer ceramic components.

Keywords

Ontology, Knowledge, Semantic data integration

1. Introduction

The rapid development of new technologies, particularly in information technology, places high demands on the sustainable development of high-performance ceramic materials. In order to meet environmental requirements and at the same time satisfy ongoing industrial demands, experiments and simulations are carried out that generate considerable amounts of data. The efficient development of new materials could be enhanced by careful organization of this data. However, we face unavoidable challenges regarding the collection, integration and sharing of this generated data. Information Logistics (IL), as a new discipline, focuses on sharing and transferring information objects specifically to engineering entry points, engineering process activities and model-based engineering execution points. Thereby, IL addresses the inevitable challenges of capturing, integrating and sharing this generated data and helps to drive advances in advanced ceramics more effectively [1].

To avoid data loss and lack of linkage, the application of a standardized approach is crucial. The digitization of materials offers a highly beneficial way to organize the generated data and ensure its interoperability [2]. However, the practical implementation of material digitization requires more than just collecting and storing data. Information from different sources, be

Joint Proceedings of RCIS 2024 Workshops and Research Projects Track, May 14-17, 2024, Guimarães, Portugal
sahar.ben.hassine@tu-berlin.de (S. Ben Hassine); rainer.stark@tu-berlin.de (R. Stark)
0000-0003-0509-204X (S. Ben Hassine); 0000-0002-2599-0130 (R. Stark)

it through experiments or simulations, needs to be formalized and combined in a common language to make it understandable and applicable for both humans and machines.

In this context, information technology plays a central role, specially the use of ontologies. The interoperability of information, models, software and data is crucial to enable an integrated approach to material design and product improvement.

Within the ongoing KNOW-NOW project, which extends until the end of 2024, an ontology is being developed that will make it possible to standardize existing data in materials development and ceramic multi-layer technology. This ontology plays a decisive role in making the data transparent and usable by third parties. Metadata and semantics ensure a clear representation and the data can be used for the development and production of ceramic products in research and industry. The project's overarching goal of creating digital tools for the rapid industrial application of experimental data is supported by the ontology created. This enables the linking of real technological data and material properties with simulation models for predicting sintering behavior, particularly with regard to deformation and cracking.

2. Objectives and Expected tangible Outputs

The KNOW-NOW project pursues the primary goal of developing a pioneering approach to bundle implicit and distributed expert knowledge in ceramic multi-layer development and make it accessible for interdisciplinary applications [3]. The focus is on the strategic use of ontologies as a fundamental structure for the systematic integration and continuous updating of knowledge in a structured form. In pursuit of this goal, a fit-gap analysis is conducted to identify the ontology components that need to be added to the ontology already provided by the Material Digital platform [4]. All partners define mandatory and optional metadata. The required extensions to the existing EMMO[1] ontology are created in a suitable ontology editor tool and provided in OWL (Web Ontology Language) format.

The project aims to establish an ontology as a comprehensive "body of knowledge" and to define its role as a formal representation of knowledge and concepts in the context of the development of ceramic multi-layers. All experimental, technological and simulated data are to be semantically linked in order to create a digital image of material behavior (see Figure 1). The ontology serves as a common vocabulary for defining terms, concepts and relationships within the domain, using ontology languages such as RDF(Resource Description Framework), OWL, or RDFS (Resource Description Framework Schema).

In order to ensure effective mapping of the ontology with meta data, research data and simulation data, digital tools should be developed for the use and maintenance of the knowledge formalized in the ontology. The focus is on developing a data pipeline that enables seamless raw data integration with ontology. The aim is to create a uniform understanding and realize

[1]The Elementary Multiperspective Material Ontology (EMMO), https://emmo-repo.github.io/

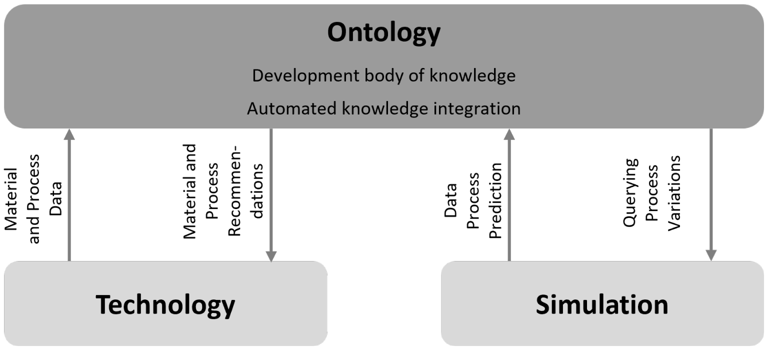


Figure 1: Ontology intersection with the multi-layer ceramic process

a semantic link between different data types. The integration of metadata into the data pipeline is another goal in order to add contextual information to the research data and to enable comprehensive documentation and interpretation. The creation of mechanisms for semantic linking between ontology, research data and metadata is crucial to enable improved interoperability of the data.

The data pipeline aims to realize a comprehensive and high-quality data linkage. This forms the basis for precise analyses and well-founded decisions in materials development and multi-layer technology. It is designed to systematically and efficiently manage the data captured by the researchers during the course of the conducted experiments.

A user interface is being developed to manage the ontology and the data pipeline. This interface will make it possible to link the ontology with experimental data and perform simulation experiments based on stored process and material parameters.

The user interface will include various functions. Firstly, it will enable the seamless integration of experimental data into the ontology through a structured workflow, creating a comprehensive knowledge base. Furthermore, it will be possible to define and save relevant process and material parameters. Users will be able to enter experimental data into the ontology, with the user interface ensuring that it conforms to the defined metadata standards.

A key aspect of the user interface is the configuration and execution of simulation experiments. The stored process and material parameters are used for this. The results of these simulations can be visualized in the user interface and compared with the experimental data.

In addition, the user interface provides tools for managing the ontology, including the option of adding new ontology components or modifying existing ones. The development of this user interface enables efficient management and utilization of knowledge in the context of ceramic multi-layer development.

3. Current project results

3.1. Ontology of ceramic multi-layer component development

The ontology development process is, in practice, an iterative process that is constantly repeated in order to continuously improve or expand the ontology. This process can be described as open-ended, as knowledge itself is subject to constant change [5]. The proposed approach to ontology development in this project begins with defining the overall goal for the use of the ontology. This goal was defined in a workshop together with all participants. The ontology should link simulations with test data and process parameters in such a way that the material properties, process characteristics and component behavior can be estimated in advance. On this basis, two user stories were formulated, based on the approach of Lucassen et al. [6]:

- As a material scientist, I want to use sintering simulation with a specific part design, sintering process characteristics and physical parameters of the material system to predict the bending after the sintering process.

- As a material scientist, I want to use the sintering simulation to optimize the part design with respect to the multi-layer material and the parameters of the sintering process.

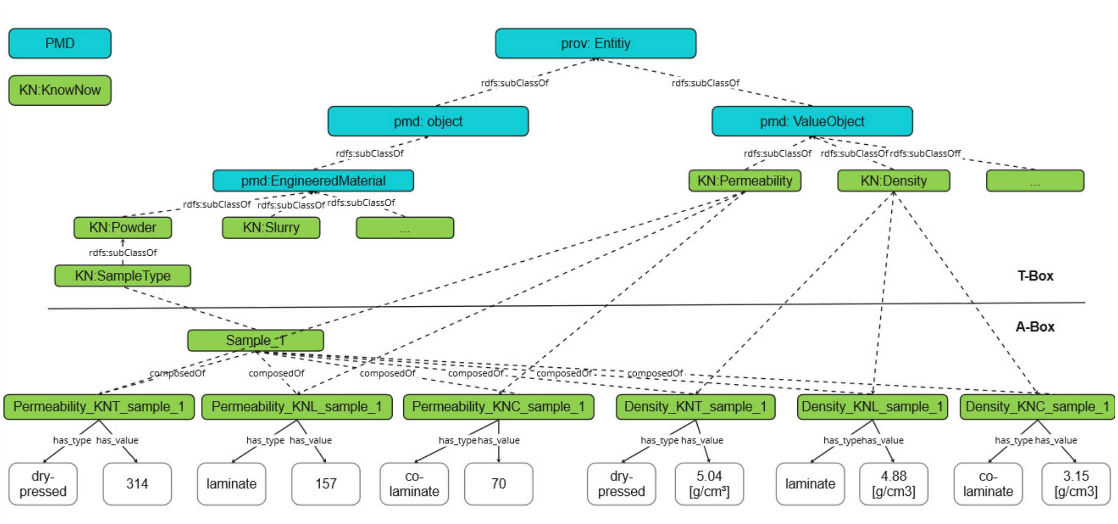


Figure 2: Excerpt of Know-Now Ontology

An analysis of the multi-layer ceramic manufacturing process and identification of the relevant data flows was the next step. A map of the process flow was created, showing specific data sources and sinks along the entire process. This analysis formed the basis for determining the relevant data required for the defined objective.

The Protégé [2] ontology editor was used as the central tool for the implementation and further development of the Know-Now ontology [3] in the project. Classes, properties and relationships were defined to form the structure of the ontology 2. The continuous integration of Protégé into the development process made it possible to adapt the ontology to new findings or changing requirements. In addition, Python-based libraries such as rdflib [7] were used for semantic data processing within EMMO Ontology and the PMDco [4] Ontology.

Figure 2 presents the KN ontology, which illustrates how the relationship between permeability and density is represented. The upper section shows T-box statements defining classes and relations, while the lower part details concrete entities and their links in A-box entries. Classes from the PMDco ontology are symbolized by blue balloons, whereas more specific classes such as density and powder are derived for the domain ontology and shown as green balloons. In the context of ceramics, a powder with a specific composition is classified as an EngineeredMaterial. The SampleType subclass of powder includes various sample generations. For example, a laminated sample, which is characterized by different properties, is linked to subclasses of ValueObject.The investigation on a specific ferrite powder, named sample 1, shows that dry pressed, co-laminated and laminated samples have different densities and permeability values. The integration of these data into the KN ontology clarifies important relationships and allows valuable conclusions to be drawn for the analysis of material samples in different experiments.

3.2. Data Pipeline

In an academic context, the data pipeline may include the following steps to process heterogeneous raw data on a local server, convert them to a machine-readable format, generate RDF triples, link them to a domain ontology, and generate a Turtle file format (ttl.) as output. Finally, the linked data is stored in a database such as Apache Fuseki [5] Server to perform reasoning using SPARQL [6] queries. The data pipeline consists of three main steps: Pre-Processing, Mapping, and Storage (see Figure 3).

1. **Data storage:** The data is stored in structured templates, with each file containing experimental parameters and results in defined sections. A hierarchical directory structure allows data to be organized in an orderly format based on the experimental methods used. This standardization of both the templates and the directory structure is crucial for efficient pre-processing and automation.

2. **Pre-Processing:** The data pipeline starts by extracting the raw heterogeneous data from a local server. This can be various file formats such as CSV (Comma-Separated Values), Excel, text files or database exports. The data is extracted from the sources and

[2] https://protege.stanford.edu/
[3] Know-Now Ontology: https://github.com/materialdigital/materialdigital1_ontology_collection/tree/main/KNOW-NOW
[4] https://github.com/materialdigital/core-ontology
[5] https://jena.apache.org/documentation/fuseki2/
[6] https://www.w3.org/TR/sparql11-query/

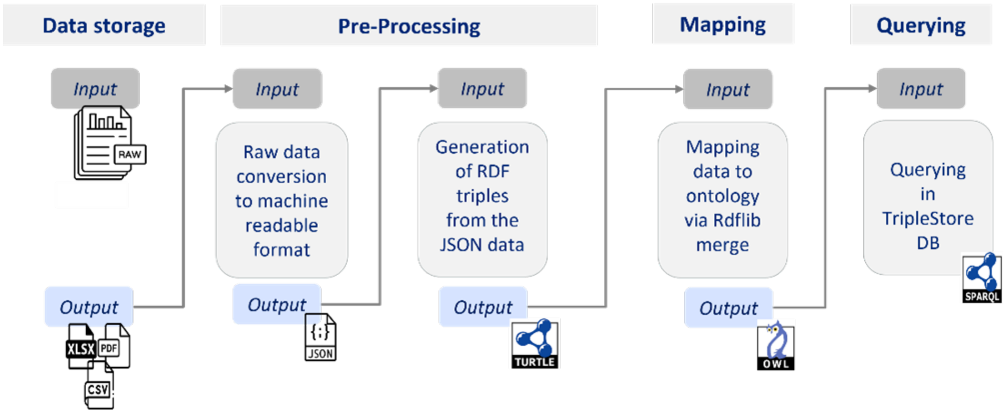


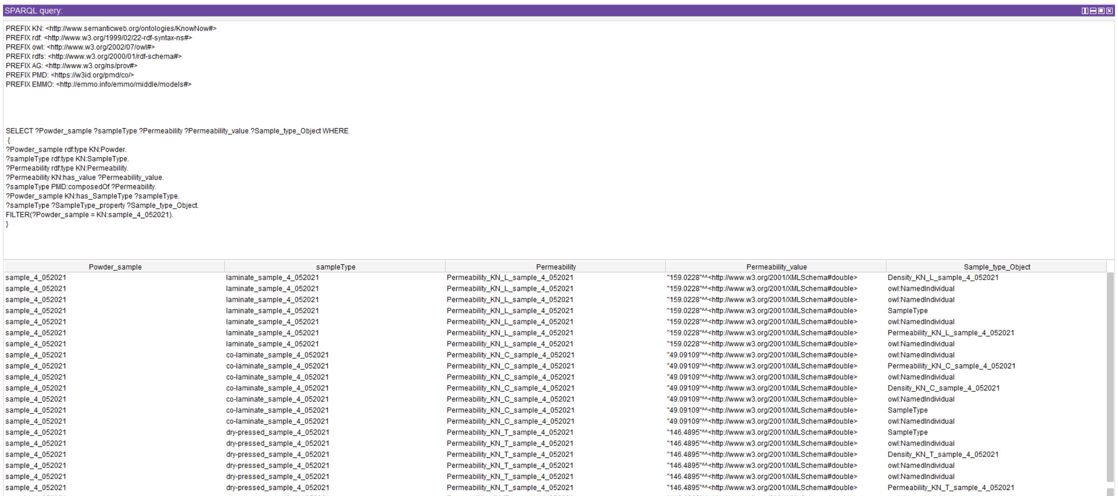
Figure 3: Data Pipeline

made available for further processing. After extraction, the data pipeline goes through a pre-processing step. Here, the raw data is prepared and put into a consistent format. This may include cleaning up errors or inconsistencies, converting data to the correct format, or removing duplicates. The pre-processed data is converted to a machine-readable format such as JSON. The process of converting raw data into a machine-readable format often involves parsing, data wrangling, or data cleansing techniques to put the data into a structured form. The conversion makes it easier to process the data and prepare it for RDF triples generation. Finally, the converted data is structured according to an ontology and converted into RDF triples to create a semantic data model. RDF triples consist of subject-predicate-object, where the subject represents a resource, the predicate represents a property of that resource, and the object specifies the value of that property. Generating RDF triples from the data allows the information to be represented in a structured and semantic form that is easier to process and interpret.

3. **Mapping:** In the mapping step, the generated RDF triples are linked to a domain ontology. In this process, data attributes are mapped to corresponding concepts and properties in the ontology to create an extended semantic context. Mapping categorizes the generated RDF triples according to the classes and properties defined in the ontology. This step gives semantic meaning to the data and enables uniform and consistent data integration. The processed data is generated as output in Turtle format, a readable text form for describing RDF data. The Turtle format facilitates the readability and exchange of the linked data.

4. **Storage & Reasoning:** The generated linked data is stored in a specialized Triplestore database such as Apache Fuseki Server. Using a Triplestore database such as Fuseki allows for fast and flexible querying of data and ensures scalability as the volume of data increases. Apache Jena Fuseki Server, as a well-known SPARQL endpoint system, is specifically designed for storing and querying RDF data. After the data is stored in the Triplestore database, SPARQL queries can be used to draw conclusions from the linked data.

During the demonstration of functionality, the data pipeline streamlines tasks by enabling laboratory technicians to input data, triggering automated processes without necessitating programming expertise from researchers. SPARQL queries provide access to interconnected data in the triple store. Figure 4 demonstrates a SPARQL query retrieving permeability data for a specific powder sample (sample_4_052021), aiding a component developer's investigation into variations in results. The retrieved data highlights differences in sample types (laminate, co-laminate, dry-pressed) and their corresponding density values, emphasizing the influence of manufacturing methods on properties and enabling informed decisions by non-experts in component simulations.



SPARQL query:

```
PREFIX KN: <http://www.semanticweb.org/ontologies/KnowNow#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX AG: <http://www.w3.org/ns/prov#>
PREFIX PMD: <https://w3id.org/pmd/co/>
PREFIX EMMO: <http://emmo.info/emmo/middle/models#>

SELECT ?Powder_sample ?sampleType ?Permeability ?Permeability_value ?Sample_type_Object WHERE
 {
?Powder_sample rdf:type KN:Powder.
?sampleType rdf:type KN:SampleType.
?Permeability rdf:type KN:Permeability.
?Permeability KN:has_value ?Permeability_value.
?sampleType PMD:composedOf ?Permeability.
?Powder_sample KN:has_SampleType ?sampleType.
?sampleType ?SampleType_property ?Sample_type_Object.
FILTER(?Powder_sample = KN:sample_4_052021).
}
```

Powder_sample	sampleType	Permeability	Permeability_value	Sample_type_Object
sample_4_052021	laminate_sample_4_052021	Permeability_KN_L_sample_4_052021	"159.0228"^^<http://www.w3.org/2001/XMLSchema#double>	Density_KN_L_sample_4_052021
sample_4_052021	laminate_sample_4_052021	Permeability_KN_L_sample_4_052021	"159.0228"^^<http://www.w3.org/2001/XMLSchema#double>	owl:NamedIndividual
sample_4_052021	laminate_sample_4_052021	Permeability_KN_L_sample_4_052021	"159.0228"^^<http://www.w3.org/2001/XMLSchema#double>	owl:NamedIndividual
sample_4_052021	laminate_sample_4_052021	Permeability_KN_L_sample_4_052021	"159.0228"^^<http://www.w3.org/2001/XMLSchema#double>	SampleType
sample_4_052021	laminate_sample_4_052021	Permeability_KN_L_sample_4_052021	"159.0228"^^<http://www.w3.org/2001/XMLSchema#double>	owl:NamedIndividual
sample_4_052021	laminate_sample_4_052021	Permeability_KN_L_sample_4_052021	"159.0228"^^<http://www.w3.org/2001/XMLSchema#double>	Permeability_KN_L_sample_4_052021
sample_4_052021	laminate_sample_4_052021	Permeability_KN_L_sample_4_052021	"159.0228"^^<http://www.w3.org/2001/XMLSchema#double>	owl:NamedIndividual
sample_4_052021	co-laminate_sample_4_052021	Permeability_KN_C_sample_4_052021	"49.09109"^^<http://www.w3.org/2001/XMLSchema#double>	owl:NamedIndividual
sample_4_052021	co-laminate_sample_4_052021	Permeability_KN_C_sample_4_052021	"49.09109"^^<http://www.w3.org/2001/XMLSchema#double>	Permeability_KN_C_sample_4_052021
sample_4_052021	co-laminate_sample_4_052021	Permeability_KN_C_sample_4_052021	"49.09109"^^<http://www.w3.org/2001/XMLSchema#double>	owl:NamedIndividual
sample_4_052021	co-laminate_sample_4_052021	Permeability_KN_C_sample_4_052021	"49.09109"^^<http://www.w3.org/2001/XMLSchema#double>	Density_KN_C_sample_4_052021
sample_4_052021	co-laminate_sample_4_052021	Permeability_KN_C_sample_4_052021	"49.09109"^^<http://www.w3.org/2001/XMLSchema#double>	owl:NamedIndividual
sample_4_052021	co-laminate_sample_4_052021	Permeability_KN_C_sample_4_052021	"49.09109"^^<http://www.w3.org/2001/XMLSchema#double>	SampleType
sample_4_052021	co-laminate_sample_4_052021	Permeability_KN_C_sample_4_052021	"49.09109"^^<http://www.w3.org/2001/XMLSchema#double>	owl:NamedIndividual
sample_4_052021	dry-pressed_sample_4_052021	Permeability_KN_T_sample_4_052021	"146.4895"^^<http://www.w3.org/2001/XMLSchema#double>	SampleType
sample_4_052021	dry-pressed_sample_4_052021	Permeability_KN_T_sample_4_052021	"146.4895"^^<http://www.w3.org/2001/XMLSchema#double>	owl:NamedIndividual
sample_4_052021	dry-pressed_sample_4_052021	Permeability_KN_T_sample_4_052021	"146.4895"^^<http://www.w3.org/2001/XMLSchema#double>	Density_KN_T_sample_4_052021
sample_4_052021	dry-pressed_sample_4_052021	Permeability_KN_T_sample_4_052021	"146.4895"^^<http://www.w3.org/2001/XMLSchema#double>	owl:NamedIndividual
sample_4_052021	dry-pressed_sample_4_052021	Permeability_KN_T_sample_4_052021	"146.4895"^^<http://www.w3.org/2001/XMLSchema#double>	owl:NamedIndividual
sample_4_052021	dry-pressed_sample_4_052021	Permeability_KN_T_sample_4_052021	"146.4895"^^<http://www.w3.org/2001/XMLSchema#double>	Permeability_KN_T_sample_4_052021

Figure 4: SPARQL Query for Extracting Data from the Triple Store

4. Relevance to information science

The research conducted as part of the KNOW-NOW project is of considerable importance for information science. The challenges addressed in the course of the project reflect the far-reaching requirements and developments in the field of information science. In view of the rapid progress in information technology, we are faced with the complex task of developing high-performance ceramic materials in a sustainable manner. The generation of extensive data through experiments and simulations requires careful organization to meet environmental and industrial needs. Standardized approaches are crucial to avoid data loss and missing links. The digitization of materials offers an effective solution for organizing and ensuring the interoperability of this generated data.

However, the practical implementation of material digitization requires more than just data collection. Information from different sources must be formalized and combined in a common language to ensure comprehensibility for humans and machines. Ontologies play a central role here, especially in information technology, to ensure the interoperability of information, simulation models and data. The data pipeline, which enables seamless integration of the raw data with the ontology, also plays a central role in the project. This pipeline supports the efficient management of the data collected during the experiments and forms the basis for precise analyses and well-founded decisions in material development and multi-layer technology.

5. Conclusion

In this paper we have presented the goals and current results of the ongoing Know-Now project, which aims to meet the requirements of information and industrial technology through the development of high-performance ceramics. A key element of this effort is the introduction of an ontology-based system that combines experimental, technological and simulated data into a common vocabulary through the use of a specialized data pipeline. The present results of this paper focus on the development of the ontology, which bundles implicit expert knowledge, and the data pipeline, which enables efficient management and semantic linking. The ontology-based system plays a crucial role in ensuring an integrated and comprehensive approach to data processing within the Know-Now project.

Acknowledgments

The Know-Now project (Ontology based knowledge systems for ceramic multi-layer components) project has been funded by the BMBF (13XP5123B).

References

[1] R. Stark, Virtual Product Creation in Industry, Springer, 2022.
[2] B. Bayerlein, T. Hanke, T. Muth, J. Riedel, M. Schilling, C. Schweizer, B. Skrotzki, A. Todor, B. Moreno Torres, J. F. Unger, et al., A perspective on digital knowledge representation in materials science and engineering, Advanced Engineering Materials 24 (2022) 2101176.
[3] R. Stark, C. Fresemann, S. Ben Hassine, Ontologien als datenquelle für prädiktive simulation, WiEGep News (2022).
[4] B. Bayerlein, M. Schilling, H. Birkholz, M. Jung, J. Waitelonis, L. Mädler, H. Sack, PMD core ontology: Achieving semantic interoperability in materials science, Materials & Design 237 (2024) 112603. doi:10.1016/j.matdes.2023.112603.
[5] F. Ocker, C. J. Paredis, B. Vogel-Heuser, Applying knowledge bases to make factories smarter, at-Automatisierungstechnik 67 (2019) 504–517.
[6] G. Lucassen, F. Dalpiaz, J. M. E. van der Werf, S. Brinkkemper, Improving agile requirements: the quality user story framework and tool, Requirements engineering 21 (2016) 383–403.
[7] C. Boettiger, rdflib: A high level wrapper around the redland package for common rdf applications (2018).