PERSPECTIVE

ADVANCED ENGINEERING MATERIALS
www.aem-journal.com

Concepts for a Semantically Accessible Materials Data Space: Overview over Specific Implementations in Materials Science

Bernd Bayerlein, Jörg Waitelonis, Henk Birkholz, Matthias Jung, Markus Schilling, Philipp v. Hartrott, Marian Bruns, Jörg Schaarschmidt, Kristian Beilke, Marcel Mutz, Vincent Nebel, Veit Königer, Lisa Beran, Tobias Kraus, Akhilesh Vyas, Lars Vogt, Moritz Blum, Basil Ell, Ya-Fan Chen, Tina Waurischk, Akhil Thomas, Ali Riza Durmaz, Sahar Ben Hassine, Carina Fresemann, Gordian Dziwis, Hossein Beygi Nasrabadi, Thomas Hanke, Melissa Telong, Stephan Pirskawetz, Mohamed Kamal, Thomas Bjarsch, Ursula Pähler, Peter Hofmann, Mena Leemhuis, Özgür L. Özçep, Lars-Peter Meyer, Birgit Skrotzki, Jörg Neugebauer, Wolfgang Wenzel, Harald Sack, Chris Eberl, Pedro Dolabella Portella, Tilmann Hickel, Lutz Mädler, and Peter Gumbsch*

This work is dedicated to the promising young scientists who, with their fresh perspectives and relentless curiosity, represent the future of digital transformation within the MaterialDigital Initiative

This article describes advancements in the ongoing digital transformation in materials science and engineering. It is driven by domain-specific successes and the development of specialized digital data spaces. There is an evident and increasing need for standardization across various subdomains to support science data exchange across entities. The MaterialDigital Initiative, funded by the German Federal Ministry of Education and Research, takes on a key role in this context, fostering collaborative efforts to establish a unified materials data space. The implementation of digital workflows and Semantic Web technologies, such as ontologies and knowledge graphs, facilitates the semantic integration of heterogeneous data and tools at multiple scales. Central to this effort is the prototyping of a knowledge graph that employs application ontologies tailored to specific data domains, thereby enhancing semantic interoperability. The collaborative approach of the Initiative's community provides significant support infrastructure for understanding and implementing standardized data structures, enhancing the efficiency of data-driven processes in materials development and discovery. Insights and methodologies developed via the MaterialDigital Initiative emphasize the transformative potential of ontology-based approaches in materials science, paving the way toward simplified integration into a unified, consolidated data space of high value.

1. Introduction

Materials science and engineering is a knowledge area with a broad interdisciplinary basis in the natural sciences and engineering disciplines,[1] making it highly diverse. This diversity reflects into the current process of digitalization of materials science and engineering. Each community and subdomain establishes its own materials data structures with its own domain-specific taxonomy and concepts.[2–4] Semantic Web[5] technologies promise to solve some of these challenges, as they enable the integration of heterogeneous knowledge, data, and tools through ontologies[6–8] and crossdomain knowledge graphs.[9–12] Nevertheless, implementing these ontologies and knowledge graphs requires the integration of knowledge across various subdomains.[13,14] Effective knowledge integration requires a deep understanding of the subdomains involved. This understanding helps

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adem.202401092.

DOI: 10.1002/adem.202401092

B. Bayerlein, M. Schilling, T. Waurischk, H. Beygi Nasrabadi, M. Telong, S. Pirskawetz, B. Skrotzki, T. Hickel
Bundesanstalt für Materialforschung und-prüfung (BAM)
Unter den Eichen 87, 12205 Berlin, Germany
E-mail: bernd.bayerlein@bam.de

J. Waitelonis, H. Sack
Leibniz Institute for Information Infrastructure (FIZ Karlsruhe)
Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

identify suitable points for integration, such as places where taxonomies from different domains can be connected. Another example is the creation of bridging concepts, properties, or equivalency assertions that connect subdomains at an appropriate level of abstraction.

At its core, materials science and engineering is based on relationships between raw material composition–processing steps–materials structure–materials properties–device performance.[15,16] Each pair of research areas highlighted above can be analyzed on multiple scales, experimentally or using simulation methods.[17] Crossarea analysis is particularly challenging sometimes due to the magnitude of the datasets but more often due to the diffuse character and lack of specificity of relevant data. In the chain of materials data relationships above, an often encountered deficit is the lack of structural data which often hinders the establishment of clear relationships to explicit processing paths and resulting properties.[18,19] Current advances in digital image recognition of microstructures[20,21] as well as in analytical methods and digital data handling[8,22] hold promise to close this gap. Well-annotated structured data together with the advancing digital data handling are vital prerequisites for the systematic information transfer along this chain of research areas and pivotal to enable new strategies for accelerated material development and consequently improved potential.[23,24]

Data management within the materials science community traditionally relies on independent databases. Several material acceleration platforms aim to integrate multiple databases and approaches to enhance effectiveness and productivity in developing materials solutions.[25–27] Connecting the subdomains, their databases, and data spaces would provide the materials science community access to much broader data on different scales.[19,28] For example, component design for reliability or lifetime expectancy could be significantly improved if the local materials structures resulting from processing were known.[21,29] Achieving these positive effects requires harmonizing data formats across different databases and thereby going beyond hitherto-implemented domain-specific workflow solutions like refs. [30,31]. This harmonization is a key issue in ensuring high level of interoperability among both public and private data custodians.[32,33] Of course, it is crucial to maintain sovereignty over data repositories, ensuring that the intellectual property and data confidentiality are protected even in networks of material data repositories.

The present article outlines how the MaterialDigital Initiative addresses these challenges. Funded by the German Ministry of

H. Birkholz, L. Mädler
Leibniz-Institute for Materials Engineering - IWT
Badgasteiner Str. 3, 28359 Bremen, Germany

M. Jung, P. v. Hartrott, A. Thomas, A. R. Durmaz, T. Hanke, C. Eberl, P. D. Portella, P. Gumbsch
Fraunhofer Institute for Mechanics of Materials (IWM)
Wöhlerstr. 11, 79108 Freiburg, Germany

M. Bruns, J. Neugebauer
Max Planck Institute for Sustainable Materials
Max-Planck-Straße 1, 40237 Düsseldorf, Germany

J. Schaarschmidt, W. Wenzel, P. Gumbsch
Karlsruhe Institute of Technology (KIT)
Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

K. Beilke
Fraunhofer Institute for Applied and Integrated Security (AISEC)
Lichtenbergstraße 11, 85748 Garching, Germany

M. Mutz, V. Nebel
August-Wilhelm Scheer Institute for Digital Products and Processes gGmbH
Uni Campus Nord D 5 1, 66123 Saarbrücken, Germany

V. Königer
Materials Research Institute (IMFAA)
Aalen University
73430 Aalen, Germany

L. Beran, T. Kraus
INM — Leibniz-Institute for New Materials gGmbH
Campus D2 2, 66123 Saarbrücken, Germany

T. Kraus
Colloid and Interface Chemistry
Saarland University
66123 Saarbrücken, Germany

A. Vyas, L. Vogt
TIB Leibniz Information Centre for Science and Technology
Welfengarten 1B, 30167 Hannover, Germany

M. Blum, B. Ell
CITEC
Bielefeld University
Inspiration 1, 33619 Bielefeld, Germany

B. Ell
Department of Informatics
University of Oslo
Gaustadalléen 23B, 0373 Oslo, Norway

Y.-F. Chen
Otto Schott Institute of Materials Research
Friedrich Schiller University Jena
Löbdergraben 32, 07743 Jena, Germany

S. Ben Hassine, C. Fresemann
Industrial Information Technology
Technische Universität Berlin (TU Berlin)
Pascalstr. 8-9, 10587 Berlin, Germany

G. Dziwis, L.-P. Meyer
Institut für Angewandte Informatik (InfAI)
Goerdelerring 9, 04109 Leipzig, Germany

M. Kamal
Institute of Digital and Autonomous Construction (IDAC)
Hamburg University of Technology
Blohmstraße 15, 21079 Hamburg, Germany

T. Bjarsch
Fraunhofer Institute for Casting, Composite and Processing Technology IGCV
Am Technologiezentrum 2, 86159 Augsburg, Germany

U. Pähler, P. Hofmann
Centre for Applied Research, Applied Computer Science & Bionics - Technology Campus Freyung
Deggendorf Institute of Technology (DIT)
Dieter-Görlitz-Platz 1, 94469 Deggendorf, Germany

M. Leemhuis
Institute of Information Systems
University of Lübeck
Ratzeburger Allee 160, 23562 Lübeck, Germany

Ö. L. Özçep
Institute of Humanities-Centered Artificial Intelligence
University of Hamburg
Warburgstraße 28, 20354 Hamburg, Germany

L. Mädler
Department of Production Engineering
University of Bremen
Badgasteiner Straße 3, 28359 Bremen, Germany

Education and Research, this initiative consists of a platform coordinating the activities of the different partners involved and proposing harmonized solutions and tools for different tasks.[34,35] In the first phase of the MaterialDigital Initiative, 13 projects were funded, aiming at the development of innovative digital methods for solving current engineering problems. The diversity of materials classes and themes covered by these projects is shown in **Table 1**. This initiative reflects the immanent diversity of materials science and engineering in its different aspects. Through a collaborative approach, the initiative has gathered different techniques proposed in the individual projects and developed convergent solutions applying methods of the Semantic Web.

In this article, the concept of a data space in materials science and engineering is first introduced in Section 2. Subsequently, the specific features of MaterialDigital Initiative are detailed in Section 3. Building on this foundation, Section 4 delves into the development of application ontologies (AO), the ontological framework, and an integration strategy for prototype knowledge graphs, and their creation. This is the central theme of the paper. Section 5 showcases exemplary analysis tools from the MaterialDigital projects that benefit from leveraging ontologies. Strategies for the ongoing enhancement of proposed tools and a comparison with similar (inter)national initiatives are outlined in Section 6 and 7. The sustainability of the MaterialDigital Initiative, vital for its applicability and acceptance within the materials science and engineering community, is also addressed here. The article concludes in Section 8, summarizing key findings and providing future directions.

In a differently scoped overview article, the development of workflows during the first phase of the MaterialDigital Initiative is described in more detail in ref. [36], where the focus is on the incremental refinement of workflow concepts and the underlying infrastructure that supports the FAIR principles. The article also highlights specific challenges in workflow standardization and their impact on interoperability and reusability across multiple MaterialDigital projects.

2. Data Spaces in Materials Science and Engineering

Digital transformation involves all scientific disciplines and is advanced by the MaterialDigital Initiative in the field of materials science and engineering. The significance of data-centric approaches and methods, including artificial intelligence,[37,38] is especially apparent in the production, acquisition, and management of vast data collections and their integration into digitally interconnected repositories (see Section 2.1). Shared data spaces[39] necessitate a delicate balance between data sovereignty and transparency, facilitated by the application of appropriate technologies and frameworks for data sharing and utilization (see Section 2.2). This approach enhances data space ecosystems and creates revenue opportunities (see Section 2.3). However, adherence to the FAIR principles[32] is essential for realizing these benefits. For instance, ensuring the reliability and crossdomain reusability of data is facilitated by comprehensive metadata collection (see Section 2.4).

2.1. Data Production, Data Transfer, and Data Repositories

Researchers, engineers, and professionals are familiar with the form and structure of original data in their domain of application; however, automation methods supporting continuous data production and refinement are often less understood. Original data from experimental measurements (see **Figure 1**) frequently arrives in undocumented, weakly defined, or origin-specific forms. For instance, data might be presented as a comma-separated values (CSV) file but could consist of multiple CSVs with various type-value preludes. While domain experts may find

Table 1. Overview of Material Digital initiative projects of first funding period.

Project	Short project description	Partners	References
DigiBatMat	Digital platform for battery material data, knowledge, and their linking.	5	[179,180]
DIGIT RUBBER	Digitalization in rubber processing on the example of extrusion.	7	[181,182]
DiProMag	Digitalization of a process chain for the production, characterization, and prototypical application of magnetocaloric alloys.	2	[183,184]
DiStAl	Digital strategy for the development of new, hot cracking-resistant Al powder alloys for selective laser melting.	2	[185]
GlasDigital	Data-driven workflow for accelerated glass development.	4	[186,187]
iBain	Intelligent data-guided process design for fatigue-resistant steel components using the example of bainitic microstructure.	3	[188,189]
KNOW-NOW	Ceramic multilayer development through redesign of ontology-based knowledge systems.	4	[190,191]
KupferDigital	Data ecosystem for digital materials research on the basis of ontology-based digital representations of copper and its alloys.	7	[192,193]
LeBeDigital	Lifecycle of concrete - ontology development for the concrete production process chain.	3	[194,195]
ODE_AM	Ontologies for the decentralized acquisition of multiscale static and cyclic characteristics of additively manufactured steel structures from experiment and simulation.	4	[196,197]
SensoTwin	Sensor-integrated digital twin for high-performance fiber composite applications.	2	[198,199]
SmaDi	Digitalization of smart materials and their manufacturing processes.	7	[200,201]
StahlDigital	Ontology-based interoperable workflows for the development and optimization of steel materials for component use: from sheet metal production to crash safety.	3	[202,203]

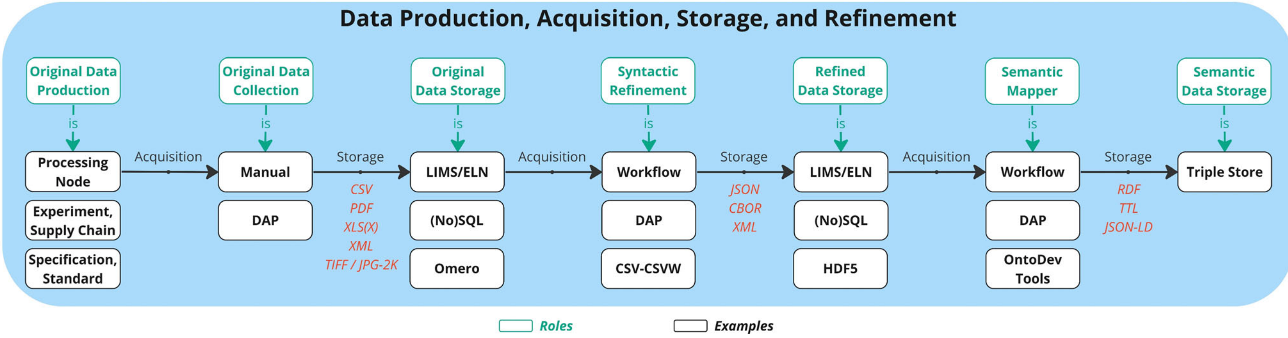

Figure 1. Schematic overview of a data transformation flow from original data production to semantic data storage within a triple store. Data originates from a pmd:ProcessingNode via experiments or supply chains and is acquired either manually or through an automated DAP using heterogeneous structures in diverse formats such as CSV, PDF, and XLS (X) (Microsoft Excel spreadsheet), including image formats such as TIFF or JPG-2K for imaging techniques. The data transformation flow can use ELNs or LIMS, as well as various databases and repositories such as Omero[176] for image handling. Syntactic refinements are carried out in formats such as JSON, CBOR, and XML using automated workflows and tools. This is followed by semantic mapping, where the refined data is integrated into a triple store using formats such as the RDF, serialized via Turtle (TTL)[177] or JSON linked data (JSON-LD).[178] The triple store is optimized for the storage and retrieval of semantic data in triples in a graph-based structure (for more details see Figure 4).

it straightforward to transform original data produced by specific equipment, the same is rarely true for crossdomain original data. Generic transformation tools often struggle with the diversity of original data from multiple domains, which typically lacks a strict definition.

As a first step toward homogenized or standardized data production, syntactic refinement typically simplifies subsequent data transformation steps by introducing better structured and well-established serialization formats. The creative use of the “CSV format” over the past decades has left today’s scientists with some systemic issues that are unchangeable with that “format”. A prominent and more structured serialization format available today is IETF STD 90, the JavaScript Object Notation (JSON),[40] which combines several valuable features: it is “human readable” while also providing simple predefined types. These types are weakly defined, for example, a value is interpreted as a number until the first nonnumeric character is parsed, after which it is treated as a string. However, JSON has limitations: it requires processing the entire document even if only a portion of it needs to be accessed, and the set of types it supports is fixed without the ability to extend them.

IETF STD 94, the Concise Binary Object Representation (CBOR),[41] is a subset of JSON that features well-defined primitive and complex types, along with an extensibility mechanism. For all types, CBOR eliminates the need for a type-length-value mechanism (a family of encoding schemes established decades ago that enables the definition of highly specific binary data formats but also results in noninteroperability between them). In nested data structures (data serialized in a tree structure within a data item), only leaf values are encoded with an actual length. CBOR’s layout results in a small footprint in terms of computing power, energy consumption, memory size, message size, and effort required for message integrity checking. While JSON excels with its lower entry threshold, quicker initial results, and ease of editing with a larger number of editors, it quickly reaches its limits when processing bulk data containing tens of thousands of data points—especially in scenarios involving recurring high-frequency experiments or production lines. Typically, as the number of reproducible experiments or the level of industry interoperability increases, CBOR becomes the more attractive choice.

Digital transformation and its corresponding refinement and homogenization of data include a significant amount of original data that results from computer simulations today. Depending on the class of software tools used, a large variety of file formats are employed. Materials data is increasingly predicted using ab initio codes based on density functional theory, often utilizing extensible markup language (XML).[42] The primary distinction from JSON is the use of tags in XML to represent the (meta)data structure. Another efficient format for storing large amounts of data is the hierarchical data Format, version 5,[43] which is gaining popularity in the ab initio community, particularly in connection with Python interfaces. Other software solutions generate original data that imply a strong demand for visualization of complex structures. Formats like those used in the visualization toolkit have become standard within the community.

Original data exhibits diversity in both its representation and formats, as well as in the means by which it is acquired from its origin for data transformation. Data transfer mechanisms range from interfaces using analog signals, to general-purpose input/output, to field busses (such as CAN-bus or PROFI-bus), to various industrial control systems (e.g., from the S7 protocol based on IETF STD 35 (RFC1006,[44] from 1987), to the LightWeight Machine-2-Machine protocol based on RFC7252[45] (from 2014, with the most recent update via RFC9175[46] from 2022), and even to cloud and edge-computing protocols employing (RESTful Application Programming Interface) (REST APIs). The first challenge in enabling data transformation lies in the diversity of data transfer methods supported by equipment. Both syntactic refinement and support for specific data conveyance protocols require a significant degree of customization and equipment-specific maintenance. Today, bridging this gap comes with considerable costs. To address this challenge effectively, a two-pronged approach is necessary: 1) enhancing the representation and formats, as well as the protocols supported by the equipment; and 2) improving available data acquisition tools to accommodate a wider range of specific and perhaps even deprecated formats and protocols, achieved through augmenting configuration options and expanding the repertoire of predefined transformation methods and configurations. Ultimately, the most promising approach to overcome this challenge involves the adoption of state-of-the-art, scalable data formats and conveyance protocols.

Separate storage systems for original data, syntactic refined data, and ultimately semantic refined data help to continuously improving refinement procedures, thereby improving the basis for scientific outputs and their reproducibility. The lesser the need for data transformation between different domains of applications and types of equipment, the lesser the data must be transformed and homogenized, reducing the costs involved in storing and managing data to ensure good scientific practice and reproducibility. The roles delineated in Figure 1 within the data flow simplify the challenge of bridging these gaps. These roles, which can be adopted by both new and existing data management systems, typically require the use of structured data formats, for example, using JSON with the option to be more precise with CBOR. These roles can be consolidated, and fewer will be necessary as original data converges toward a shared set of vocabularies or comes already in a modern representation that can directly processed semantically. A similar effect applies to the number of specialized or customized data conveyance protocols required to support effective data management. The fewer the protocols or the more unified they are across different types of equipment, the less effort and cost are needed to enable and maintain data conveyance from a data origin to a semantic data storage. Hence, this article illustrates candidates for convergence, such as the data formats JSON and CBOR.

2.2. Sovereignty and Transparency

Next to establishing common standards, a central challenge in creating a shared data space lies in reconciling competing interests regarding the findability and accessibility of proprietary data, which are central pillars of the FAIR data paradigm. Here, the principles of sovereignty and transparency are of utmost importance. Sovereignty, in this context, refers to the autonomy of data providers in controlling their data. They have the authority to determine who can access their data and under what conditions. However, the usefulness of the data space relies on the findability of specific datasets. Transparency, therefore, entails making the existence of data and its providers known within the data space without necessarily revealing the specifics of the data. Information about the existence of specific datasets can be valuable, yet the data provider may wish to disclose it only to a specific subset of participants within the data space.

The architecture of the data space plays a crucial role in balancing these principles. It should provide a unified infrastructure for navigating and accessing data while respecting the sovereignty of data providers and ensuring transparency for users. Data providers are responsible for deciding which data can be shared in which context and for implementing access policies. However, the underlying architecture must provide the necessary access control mechanisms to maintain data sovereignty. These mechanisms can operate at various levels. At an organizational level, access to specific entities in the data space can be limited, with links established only to trusted parties. In the case of the Platform MaterialDigital (PMD), this is reflected in individual instances of PMD-Satellites (PMD-S) that establish a secure WireGuard[47] mesh network with a central PMD-S as a trust anchor. At an individual level, fine-grained access control can be implemented to identify specific users and manage specific groups or roles with authentication and authorization infrastructure. In this regard, the PMD also relies on state-of-the-art components and employs KeyCloak[48] for this purpose in the reference architecture.

This reference architecture is outlined in the PMD Deployment guide[49] and serves as an example of a system that embodies these principles. It uses container technology to ensure consistency and compatibility across different systems and implements fine-grained access control at both the organizational and individual levels. Specifically, Docker Compose[50] is used as a simple mean to provide complex application stacks as multicontainer Docker applications. Perspectively, it is intended to support other container technologies, such as the widely used Kubernetes platform.[51] Specific applications of these implementations are discussed in Section 3, such as the PMD Demonstrator (see Section 3.2), the OntoDocker (see Section 3.6), and the Discovery Service (see Section 3.7).

In conclusion, creating a shared data space requires a delicate balance between data sovereignty and transparency. Applying appropriate technologies and implementing robust access control mechanisms enables the creation of a flexible and scalable infrastructure that respects these competing needs.

2.3. Higher Revenue from Data Transfer and Sharing

Access to large quantities of high-quality data and improved data transfer are crucial for developing and implementing competitive automation and artificial intelligence solutions that result in higher revenues.[52,53] Creating understandable, supportive, and accessible solutions and business models is crucial. This ensures that others can benefit alongside major data-driven projects that possess the necessary expertise and resources. Despite the high initial costs associated with data acquisition, integration, and sharing, it is essential to facilitate access for all stakeholders.[54] Data spaces contribute to this area by promoting trustworthy, federated digital ecosystems, and prioritizing data sovereignty and transparency among other benefits (see Section 2.2). Interoperability is emphasized here as a fundamental technical prerequisite, enabling the integration of a wide variety of data sources into one's research and development work, leading to more efficient processes and higher-quality products.[39]

The integration of semantic data, its provision, and its application depend on the terminology used. The Terminological Box (TBox) within the ontological framework ensures that data mapping and semantic integration are consistently understood, regardless of the data provider. This makes terminology synchronization and updating an important focus. A possible iterative curation process for semantic data storage systems is shown in **Figure 2**. The terminology for the description of domain data and its semantic integration into the triple store results from the TBox and the Assertion Box (ABox). GitHub mechanisms and synchronization with terminology services

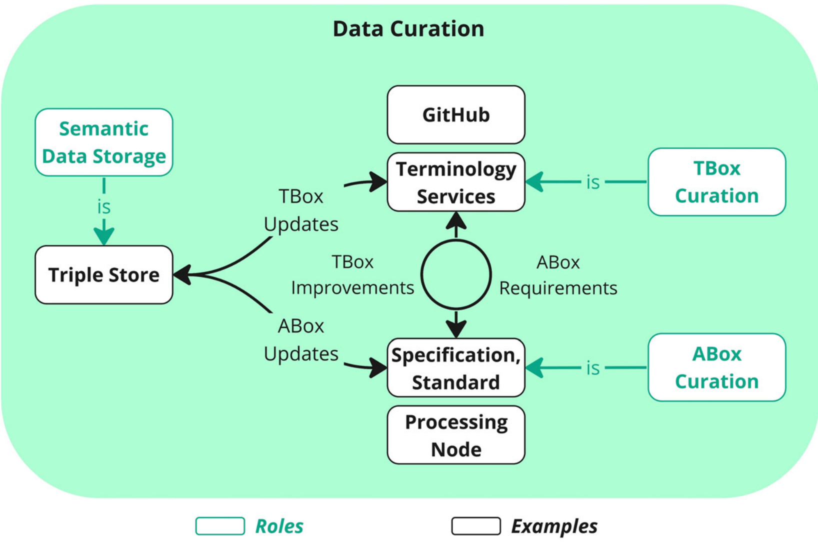

Figure 2. This schematic diagram illustrates the iterative data curation process for semantic data storage systems. The triple store is continuously updated through cyclic interactions between the TBox and the ABox. TBox curation, facilitated by GitHub and related terminology services, informs updates and improvements that impact the requirements for ABox updates. ABox curation leverages terminology from specifications, standards, and processing nodes. This dynamic, community-driven process is essential for enriching the semantic data storage, enhancing data accuracy and relevance, and ensuring adaptability to new information.

facilitate community-oriented TBox curation and harmonization. ABox curation ensures that the TBox adapts to new requirements, such as changes in standards and specifications, as well as new processing nodes, such as new laboratory equipment. Regarding the consistent description of datasets, existing metadata catalogs with standardized vocabularies are used to ensure interoperability across domain boundaries. Thus, effective terminology curation contributes to enhancing the interoperability of data spaces, which not only standardizes communication and understanding across system and organizational boundaries but also plays a crucial role in the success of data-driven environments.

Furthermore, the terminology controls the retrieval of semantically integrated data from the triple store for various applications (see **Figure 3**). By formulating SPARQL (SPARQL Protocol and resource description framework [RDF] Query Language) queries[55] according to the used terminology, the structured RDF data triples[56] stored in a knowledge graph of a data space can be systematically retrieved (see **Figure 4**). The subsequent processing and analysis of the retrieved data generate new insights, thereby enriching the knowledge graph. This also illustrates the process of data sharing within collaborative activities, where reliable data reusability is a key factor.

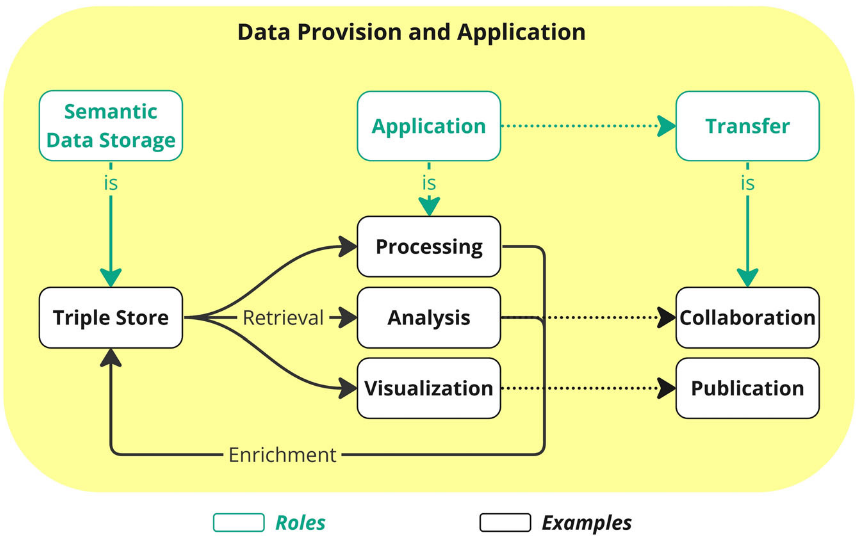

Figure 3. The schematic diagram shows semantic data stored in a triple store being used for specific applications. The retrieved data can undergo processing, analysis, and visualization. These applications enrich the triple store and contribute to cooperation and publication activities through their transfer.

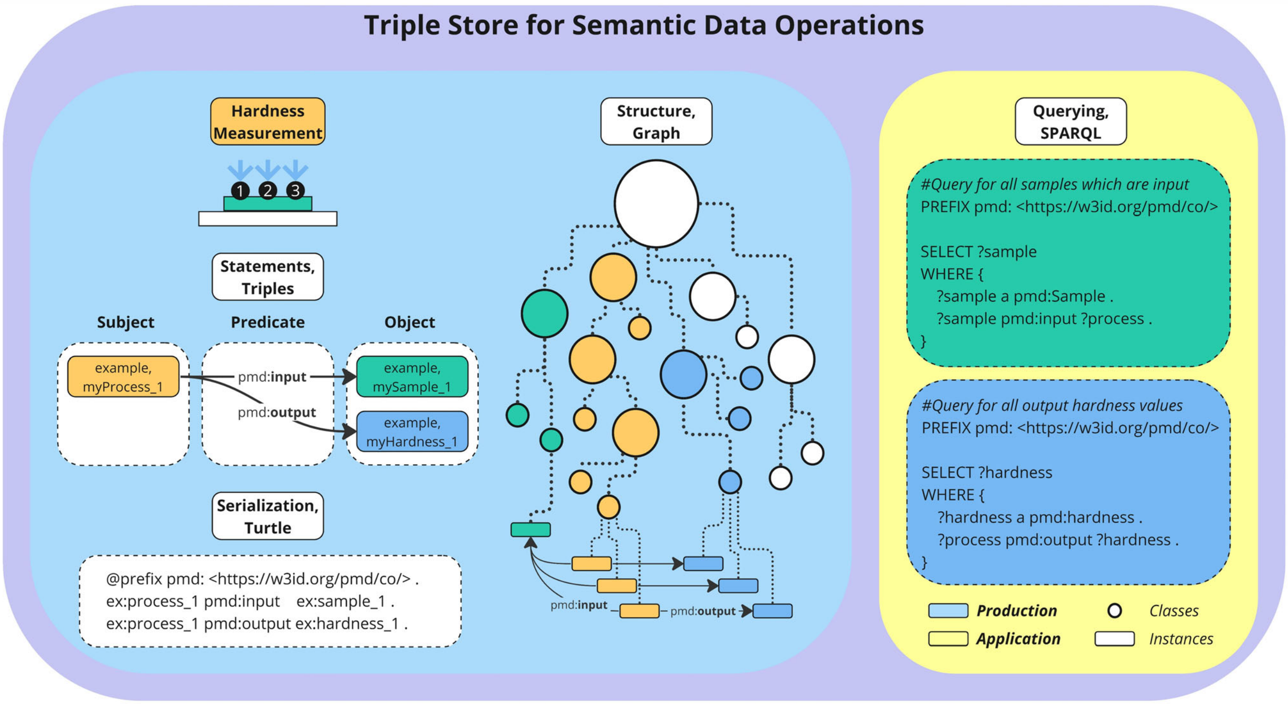

Figure 4. Schematic representation of semantic data operations in a triple store, illustrated through a simplified Brinell hardness measurement. The diagram shows a single sample undergoing the test process three times, with the sample being the pmd:input and the three hardness values as pmd:output of the process. These relationships are expressed as subject–predicate–object triples, serialized in Turtle syntax, enriching the graph's structure with new instances. The figure also highlights how tailored SPARQL queries facilitate detailed retrieval of information about the sample and corresponding hardness values.

2.4. Reproducibility for Reliable Reusability

To optimally and reliably reuse materials science data and thus realize various benefits, such as promoting scientific exchange and economic cooperation, saving resources, and accelerating material discovery and development, this data as well as its production and processing must be reproducible. This requires the comprehensive collection and availability of all contextual information crucial for the beneficial application and dissemination of this data. This includes metadata, as well as the provenance, required for the reproduction of original or processed data. Moreover, the inherent potential of this metadata can have motivating effect, leading to new findings at a later stage. For example, analyses of experiments and experimental setups could provide insights into the measurement precision of resulting data on materials and processes or enhance their presentation through more accessible visualizations.[57,58]

The detailed acquisition is carried out by the original data producer, who, in addition to the easily accessible metadata already contained in machine outputs, should also integrate additional, often hidden information such as parameters used in algorithms and scripts, as well as inaccessible information provided by proprietary formats, for instance. Digital tools such as electronic lab notebooks (ELN)[59] and laboratory information management systems (LIMS)[60] are helpful here, as they enable the digital acquisition and linking of data along their chain of origin. The challenge is to combine these tools into an automated, reproducible workflow that makes all intermediate results systematically available at any given moment. Ontological frameworks, especially the highly specified AO that extend the higher ontologies of the framework, offer the possibility of semantically capturing every procedural and experimental nuance, thus fully promoting data reproducibility.[10]

Only when materials and processes data and their metadata are acquired and stored in their entirety, and when information on their uncertainty is available, is the reliable reuse of this data guaranteed, fundamentally supporting data-driven approaches.

3. MaterialDigital: Specific Features

In the MaterialDigital Initiative, partners within project consortia work on solutions tailored toward digital challenges within specific areas of materials science and engineering (see Table 1). Their aim is to develop flexible and reliable methods for the production, integration, and exchange of materials and processes data in their data spaces and beyond. The practical application of these methods is tested and implemented using various demonstrators.

A core goal is to promote interoperability among these material data spaces created in the individual partner projects that are based on specialized languages and concepts. The semantic representation of processes, their sequences, and the resulting data structures is realized via AOs. These ontologies are developed decentrally and are primarily used for semantic data integration. An overarching ontological framework supports interoperability between AOs (see Section 4.1.2) and promotes the development of dynamic, continuously growing knowledge graphs (see Section 4.3).

The MaterialDigital Initiative also targets the development of consistent frameworks for workflow implementations and executions.[36] This ensures that data production and processing is user-friendly, sufficiently flexible, largely automatable, and reproducible. Part of this effort is the integration of various software tools into these frameworks and the registration of developed workflows in a central repository. The connection and integration of these workflows to the ontology concepts of MaterialDigital is a major topic. This includes the workflows that are part of the data flow for semantic storage in Figure 1. Workflows that access data via ontologies and workflows that utilize ontologies to semantically compose scientific protocols. Defining and describing experiences and requirements for such kinds of workflow infrastructure solutions form the technical basis for reliable data integration and exchange.

The PMD actively supports the objectives of the MaterialDigital projects by creating and providing prototypical solutions and concepts for widespread use across different materials science and engineering sectors. In this role, the PMD is instrumental in fostering, coordinating, collecting, and sustainably implementing digitalization measures. Through regular networking, like exchange rounds, workshops, and hackathons, within the initiative and the larger materials science and engineering community, the PMD facilitates the exchange of experiences and feedback and promotes knowledge transfer. While workflows are primarily discussed in regular meetings of the open workflow group, the "Ontology Playground" is an important space for collaborative work on ontology design and use (see Section 3.1).

To demonstrate possible approaches and methods of data refinement, that is, from the acquisition to the integration and the retrieval of data, several demonstrators based on Jupyter Notebooks[61] for hands-on interactions are offered. These illustrate the use of tools and frameworks developed in PMD's focus areas, such as semantic interoperability, workflows, and architecture, aiming to enhance understanding of digital infrastructure use and lower technical hurdles for implementation. For automatic data refinement, the PMD provides a set of components for a modular data acquisition pipeline (DAP). These components can serve as the basis for independent microservices or as code modules in support of existing data refinement procedures.

The PMD demonstrator (see Section 3.2), the Orowan demonstrator (see Section 3.3), and the ELN demonstrator (see Section 3.4) are particularly noteworthy and are each described in more detail in following subsections. Moreover, composable microservice building blocks, that is, the DAP (see Section 3.5) and the knowledge graph platform OntoDocker (see Section 3.6) are presented. The intention is to sensitize the community toward digital solutions and to demonstrate their technical benefits. Elevating these digital efforts within MaterialDigital at a national level could encourage broader activities over time. The developments and services provided by the PMD are detailed below.

3.1. Ontology Playground

The Ontology Playground is an open and continuous virtual space for interacting with the materials science and engineering community. In this collaborative and engaging environment, typically involving about 20 participants including domain and ontology experts, there is an ongoing effort to build a shared understanding of semantic standards and technologies. This involves working with materials and processes data intended for sharing in the designated data space. Through collaborative ontology modeling, participants explore high-level ontologies and identify key, broadly applicable materials science and engineering concepts and modeling patterns. This approach in the interactive environment also facilitates the concurrent identification and resolution of various issues, such as missing concepts, incorrect definitions, and requests for changes to existing concepts directly within the Ontology Playground. This dynamic feedback mechanism allows for continuous enhancement of the corresponding ontologies. Such collective learning activities are essential in setting requirements for designing application-specific ontologies, guided by competency questions that shape the broader ontological framework. For example, one such competency question could be, "How do properties of a material depend on its composition and production?" The Ontology Playground plays a crucial role in presenting and debating developments and uses related to the PMD Core Ontology (PMDco) (for details see Section 4.1.2), significantly contributing to the enhancement of the PMDco and related ontologies through joint application cases, feedback, and suggestions.[14] Presentations from the Show & Tell event series from MaterialDigital projects are archived and made available.[62] Useful tools for ontology development and application, which play a harmonizing role in digitalization efforts, are showcased here as well.

3.2. PMD Demonstrator

The aspects showcased with this demonstrator[63] focus on the functionalities of a specifically deployed PMD-S,[49] as the central deployment of the PMD infrastructure tools. A PMD-S is a specifically deployed environment, hosting services such as workflow environments, triple stores, and IT infrastructure components like the wireguard-based[64] PMD-mesh, which connects PMD-S instances. The demonstrator consists of Jupyter Notebooks and workflows using the two workflow tools supported by the PMD:pyiron[65] and SimStack.[66] It handles experimental data, as for example from tensile tests conducted on samples cut from a cold-rolled S355 steel sheet,[67] using a programmatic approach. A first notebook showcases a more general (meta)data exploration. It employs the semantic description of the datasets to collect information. Details gathered include the material system, test device, samples' cut orientation relative to the rolling direction, standards followed during the measurement, and the uniform resource identifier (URI) pointing to the raw data. An analysis workflow automatically collects a set of relevant metadata. It calculates different quantities of interest, such as the Young's modulus, E, and the offset-yield strength, $R_{p0.2}$, from the stress–strain data. The demonstrator is undergoing continuous development, aiming to offer a whole suite of demonstrators that will showcase the benefits of a digital infrastructure for coupling simulation and experiment, loading data into a triple store (for details on triple stores, see Figure 4) using automated workflows running computations on remote resources, and more. Recent versions can be obtained from the openly accessible GitHub repository.[63]

3.3. Orowan Demonstrator

The Orowan demonstrator[8,68] details the beneficial application of Semantic Web technologies for the integration and correlation of diverse materials science datasets. In a Jupyter Notebook script the workflow steps required to correlate mechanical[69] and microstructural data[70] from tensile tests and dark-field transmission electron microscopy over different aging times, using a 2000 series age-hardenable aluminum alloy as an example, are demonstrated. The two publicly available, unstructured datasets are transformed into the RDF format using an ontological framework based on the PMDco (see Section 4.1.2). Exemplifying good practice, this method shows how this data within a local knowledge graph can be leveraged for materials analysis through SPARQL queries, and how the calculated intermediate results extend the knowledge graph, showcasing the benefits of semantic data integration.

3.4. ELN Demonstrator

ELNs fulfill the digital documentation function of scientific work and provide a convenient method to link data acquisition with subsequent stages such as processing, analysis, and visualization. To demonstrate these capabilities, a knowledge graph is created from tensile test data collected using the free and open-source ELN, eLabFTW.[71] The contextual information and the results are semantically linked to the tensile test ontology (TTO) developed by the PMD (see Section 4.2.1). Scripts facilitate the mapping and transformation of this (meta)data into a semantically cohesive and interoperable RDF dataset. This approach of connecting ELN-recorded data with ontology-based semantics streamlines experimental processes and enhances data traceability, leading to a semantically integrated digital experimental process applicable to various tests and experiments. It supports data acquisition, analysis, processing, and reuse and educates students on data organization and semantic technologies, contributing to their training.[72]

3.5. Data Acquisition Pipeline

In scenarios where data is continuously produced, such as on factory floors, in quality assurance, or through stationary experiments used for materials science and engineering research, automated tools that operate more independently than Jupyter Notebooks with graphical user interfaces (GUIs) are beneficial. In some of these scenarios, direct network connections between a researcher's workflow tool and the data origin are prohibited due to security policies. These policies may include network isolation or separation, zero trust architectures with finely grained authentication schemes, or the confidentiality of the data involved. The DAP[73] addresses some of the requirements arising from these scenarios, mainly the acquisition, refinement, and storage of data, as illustrated in Figure 1. By decoupling data acquisition from the initial syntactical transformation, the pipeline facilitates compliance with the security policies of the environments in which it operates. This principle of decoupling duties through roles (e.g., the original data production role) and placing these duties in separate building blocks reflects the main architectural design of the modular DAP. Decoupling and modularity also serve another purpose related to data refinement needs: In scenarios where data is relatively unstructured and heterogeneous, or comes in peculiar device-specific formats, early and sometimes drastic syntactical refinement steps are necessary. However, in cases where data already adheres to newer encodings or utilizes well-known or standardized labels or semantic representations, the refinement process can skip several basic transformation steps and focus on mapping or re-encoding into more efficient formats. As a result, the modular aspects of the pipeline also support customization for different deployment scenarios and a range of data quality available.

3.6. OntoDocker

OntoDocker[74] is PMD app that provides one or more open-source triple stores (see Figure 4) as a knowledge graph platform, bundling them with a few basic features out of the box: 1) Authentication: Access control to a SPARQL endpoint can be implemented in different ways. The OntoDocker app includes single-sign-on (SSO) authentication, fueled by a proof-of-concept realm and userbase provided by the PMD SSO service. It is also possible to use OntoDocker with a separate, independent userbase under a different authority. Out of the box, it serves both as a usage example and as initial protection from direct internet access; 2) Visualization: For every triple store supported, OntoDocker includes the open-source visualization tool WebVOWL.[75] While WebVOWL focuses on the representation of classes and relationships (the "semantic schema") rather than on the assertion representation (instantiated knowledge graphs), it provides a simple and accessible graphical overview of the triple-store contents, out of the box; and 3) Basic Management: OntoDocker offers a minimal set of management features necessary for operation within a project context, particularly where instance administration is performed by nonexpert users. A straightforward backup functionality allows responsible users to save instance payload data and configuration data to an external file storage. The basic management interface enables the assignment of roles to users, specifically determining whether they have read-only or read-write permission on the data.

3.7. Discovery Service

The PMD architecture supports data management, visualization, and accessibility. The Discovery Service, with its underlying concepts,[76] plays an important role in this regard enabling users to efficiently find and use data. A fundamental component of any system that provides data is a catalogue detailing the available content. The Discovery Service within the PMD architecture is designed to fulfill this role. Currently, it allows users to graphically navigate through semantically structured data with the goal of identifying datasets of interest. This service requires data providers to submit a subset of their structured data, enriched with metadata (license, format, authors, etc.), which is then indexed and made searchable for users. Consequently, data providers have the autonomy to decide which information about their available data can be discovered.

4. Building a Materials Science Knowledge Graph

In the previous section, we detailed among others PMD's concepts and components essential for building a materials science knowledge graph. This section introduces an integration strategy that outlines how these elements interlock to form such a graph (see Section 4.1). Thereby, an ontological framework defines the structure and relationships, which is further extended and informed by domain-specific AOs for semantic data integration (see Section 4.2). A prototype knowledge graph in which all available AO are integrated as named graphs and which enables a variety of search queries is described (see Section 4.3). Section 4 concludes with a summary and the most important insights (see Section 4.4).

4.1. Integration Strategy

The integration strategy highlights how essential concepts and components for building a materials science knowledge graph work together (see **Figure 5**). Critical to this process are active community engagement and the combination of suitable technologies.

4.1.1. Community Involvement and Competence Development

The Ontology Playground (detailed in 3.1) has fostered a collaborative space for domain and ontology experts to dynamically exchange approaches and solutions on digitalization within the MaterialDigital projects. This ongoing forum enables the community to share its latest developments, facilitating technical knowledge transfer and consensus on essential data representation standards and technical specifications. This collective effort is foundational for building a knowledge graph and achieving interoperable data spaces.

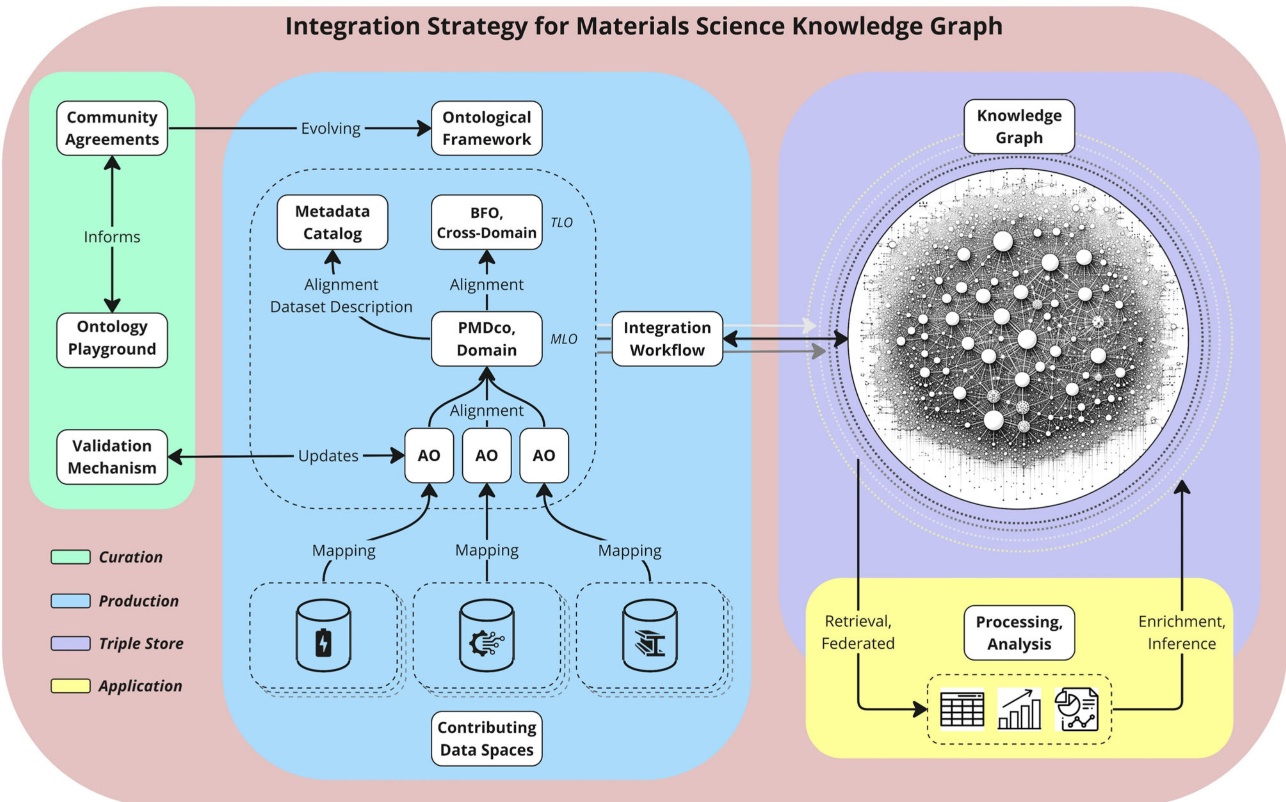

Figure 5. This schematic representation depicts the integration strategy for a materials science knowledge graph with inputs from the MaterialDigital project's data spaces. The strategy involves the use of data space-specific AO, which are refined through Git-based curation and validation mechanisms, to extend the ontological framework. The Ontology Playground is a collaborative environment for reaching consensus on employing MLO (PMDco), TLO (BFO), as well as standard metadata catalog vocabularies for dataset description. This consensus ensures that materials science datasets are standardized and interoperable. Integration workflows then incorporate RDF triples, produced from dataset mappings, into the knowledge graph, which is further enriched through processing and analytical operations on RDF data. SPARQL queries facilitate the retrieval of this enriched data from the triple store, symbolizing a dynamic, evolving structure of connected knowledge.

4.1.2. Establishing an Ontological Framework with the PMDco

At the core of the knowledge graph lies a scalable ontological framework, anchored by the PMD Core Ontology (PMDco).[77] The PMDco, a midlevel ontology (MLO), defines key concepts and relationships in materials science and engineering and is continually refined through collaboration.[14] It forms a semantic bridge to the detailed concepts within the AO of the MaterialDigital projects, consolidating fragmented solutions, enhancing interoperability, and setting guidelines for their design. This promotes a unified and interoperable materials data space.

The use of abstract concepts from standardized top-level ontologies (TLO), such as the basic formal ontology (BFO),[78] ensures crossdomain interoperability over the long term. Moreover, generic RDF vocabularies and metadata schemata standardize the description of data and ontological resources.

4.1.3. Domain-Independent Metadata Catalog

Crossdomain RDF metadata vocabularies and schemata facilitate detailed descriptions of datasets and ontological resources, increasing interoperability. Incorporating generic concepts such as publishers, authors, licenses, etc. enhances resource findability and interoperability across domains and defines usage terms. The Data Catalog Vocabulary (DCAT),[79] combining established vocabularies such as the Simple Knowledge Organization System[80] and the Dublin Core Metadata Initiative,[81] among others, is particularly useful for describing datasets and ontological resources within MaterialDigital and associated initiatives. Notably, DCAT can be extended by an application profile (DCAT) which is a recognized standard for the publication of open data in the European public sector. The AO from MaterialDigital's initial funding phase have been systematically organized into a coherent collection, integrating key DCAT catalog entities.[82]

4.1.4. Semantic Data Integration and Knowledge Graph Enrichment Using Workflows

Data integration workflows are of central importance in the construction of knowledge graphs, automating the process of data acquisition, refinement, and storage. PMD's demonstrators (see Sections 3.2–3.4) highlight the utility of workflows in semantic data integration, making complex processes accessible to non-experts. Data integration workflows, combined with ontological frameworks, standardize the conversion of data annotations to RDF format, thus ensuring uniformity across materials and processes data. Although initial setup might pose challenges, ongoing standardization efforts involving specific workflows and ontological frameworks are gradually migrating these hurdles and reducing associated costs. Semantically representing workflows enhances their transferability and interoperability, with standards such as the PROV-O[83] and its extension, the OPMW-PROV Ontology,[84] providing structured workflow representation. Data integration workflows also contribute to RDF applications and graph operations, with SPARQL query results enriching the knowledge graph.[8] Furthermore, computational workflows in materials science and engineering contribute semantic data to the knowledge graph and will be the focus of another article.[36]

4.1.5. Evaluation of Ontological Frameworks in MaterialDigital Projects

Evaluating individual ontological framework solutions within the MaterialDigital projects is a central element of our integration strategy, involving user feedback, to refine and enhance our approach. An iterative process fine tunes boundary conditions and constraints, aiming to enhance crossdomain interoperability, make integration more efficient, and improve user experience.

As part of this iterative evaluation process, the TBoxes of the provided AO developed during the initial phase of MaterialDigital were evaluated in terms of structure and content using an automated script.[85] The ontologies were first loaded, followed by an analysis of their content for completeness of metadata and adherence to technical requirements. This included applying filters for key concepts from the PMDco, quantifying the extension of key concepts via new subclasses (see Section 4.2.5), and analyzing import statements to extract foundational information such as metadata on ontology creators, contributors, licenses, and namespaces used. The comprehensive results of the evaluation can be found in ref. [86]. Overall, the automated script-driven analysis demonstrates strong potential for increasing efficiency in processing and evaluating AO for consistency and interoperability. However, further work is required to fully exploit this potential and improve the overall quality of the knowledge graph.

Table 2 summarizes the higher-level ontologies employed by the projects.

Most projects have adopted the PMDco v2.0, grounded in PROV-O.[83] However, some projects chose earlier versions of PMDco or developed mappings or bridge axioms to PMDco, especially when integrating with other frameworks such as the BFO[78] and the elementary multiperspective material ontology(EMMO).[87]

Table 2. Overview of higher-level ontologies used by MaterialDigital phase 1 projects, where x° indicates use of PMDco v1, x* denotes mapping to PMDco, and x′ indicates bridge axioms.

Project Name	PMDco	PROV-O	BFO/OBO	EMMO
DigiBatMat	x	x	–	–
DIGIT RUBBER	–	–	x	–
DiProMag	x	x	–	–
GlasDigital	x°	–	x	–
iBain	x*	–	–	x
KNOW-NOW	x	x	–	x
KupferDigital	x	x	x	x
LeBeDigital	x	x	–	–
ODE_AM	–	–	x	–
SensoTwin	x	x	–	–
SmaDi	x′	–	–	–
StahlDigital	x°	–	–	x

The gradual convergence toward a shared ontological framework and consequently a collective understanding relies to a large extent on social interactions such as the exchange of experiences. In the first phase it became apparent that the choice of the framework depends on personal preferences, pre-existing works, as well as strategic decisions.

4.2. Application Ontology Development

Beyond the AO developed by the PMD, the MaterialDigital Initiative has produced various AO specifically tailored for use in project-specific data spaces (described in detail in Section 5). These ontologies are elaborated upon in the subsequent two subsections, 4.2.1 and 4.2.2, respectively. Addressing the needs of less experienced ontology developers, projects have sought appropriate collaborative tools (see Section 4.2.3). Initial attempts at AI-supported ontology design and application have also been explored (see Section 4.2.4). Furthermore, it is shown how the ontological framework of PMDco v2.0.7 is being extended through the integration of compatible AO (see Section 4.2.5).

4.2.1. PMD Application Ontologies

PMD's available prototype AO, all based on PMDco v2.0.7, are highlighted below. These ontologies are intended to test and refine the ontological framework and provide guidance and inspiration for the MaterialDigital projects: 1) Tensile Test Ontology (TTO): This standard-compliant ontology provides a structured vocabulary tailored for capturing tensile test datasets. It facilitates interoperable, transparent, reliable, and reproducible data representation, aligning with the tensile tests of metals at room temperature standard (ISO 6892-1).[88] It includes classes for integrating tensile test (meta)data and semantic relationships for logical inference, enhancing data comparability, and test reproducibility;[89,90] 2) Precipitate Geometry Ontology (PGO): The PGO focuses on semantic integration of microstructural data from dark-field transmission electron microscopy image analysis.[8] It extends the PMDco class structure for the representation of precipitate data;[91] and 3) Microscopy Ontology (MO): Derived from a comprehensive text analysis of a large corpus of conference proceedings (see Section 4.2.4), the MO contains classes specific to microscopy and microanalysis. These classes allow for detailed description of microscopy processes, equipment, and parameters supporting reasoning for data completeness and new information derivation.[92,93]

4.2.2. MaterialDigital Application Ontologies

An overview of various AO developed by the projects, which are accessible in a shared repository,[82] is provided below: 1) DigiBatMat: This is developed the battery production and characterization ontology (BPCO), which describes electrode and cell production processes and different characterization methods used in battery research and pilot-scale production.[94] The ontology is modeled on the lithium-ion electrode and cell production pilot line at the Battery LabFactory at the Technical University of Braunschweig. It includes all the production steps toward lithium-ion battery electrodes and all assembly steps to a final lithium-ion cell. Furthermore, it links data from items, such as materials, to each processing step. The BPCO also models a wide range of characterization methods, including particle sizing, scanning electron microscopy, and impedance spectroscopy, underpinning the data model for the projects research data management platform;[95] 2) DIGIT RUBBER: This created the DigitRubber Ontology (DRO), designed to annotate the relational database of the Deutsches Institut für Kautschuktechnologie (DIK). The DRO functions primarily as a controlled vocabulary, providing human-readable definitions without incorporating axioms or reasoning capabilities, focusing instead on offering a structured descriptive framework that enhances data accessibility and clarity within DIK's extensive rubber technology database;[96] 3) DiProMag: DiProMag Magnetocaloric Ontology (DMCO), focusing on magnetocaloric alloys, is developed. The DMCO semantically models the entire process chain, including production, characterization, and prototypical application. Furthermore, the ontology allows to capture the goals and ambitions of the researchers that design and carry out experiments; 4) GlasDigital: The glass ontology for robot-assisted synthesis of oxide glasses is designed. This ontology considers various aspects, such as raw materials, melting processes, glass properties, literature background, both traditional and advanced data visualization tools (multidimensional phase diagrams), and composition–structure–property modeling. It integrates terminology from the SciGlass database, which became open source in 2019[97,98] and that of the robot-assisted glass melting plant; 5) iBain: An ontology is developed that supports multiscale modeling of bainitic steels with a focus on manufacturing, computational modeling, and characterization. Specifically, this ontology enables modeling the interaction of microstructure and fatigue properties. It facilitates the interoperability of data from phase field simulations and microscopic microstructure assessments, integrating them for future data-driven modeling efforts. The project adopted EMMO beta4.0 as the TLO, supported by domain-specific ontologies like the microstructure domain ontology[99] and CHAMEO,[100] which provides essential concepts for microstructure modeling and characterization. Additionally, it features equivalence mappings with PMDco v2.0, enabling federated queries with other PMD projects, aligning high-level concepts such as chameo:Sample $\xrightarrow{\text{owl:equivalent}}$ pmd:Sample, emmo:ManufacturedMaterial $\xrightarrow{\text{owl:equivalent}}$ pmd:EngineeredMaterial, and emmo:Quantity $\xrightarrow{\text{owl:equivalent}}$ pmd:ValueObject; 6) KNOW-NOW: KNOW-NOW ontology (KN) is developed to facilitate the acquisition and structuring of technological and material-related data across the stages of ceramic multilayer component development. The KN ontology integrates experimental and simulated data along with simulation models, enhancing cross-scale simulation of ceramic firing processes. It serves three purposes, structuring data for review, contrasting simulated and experimental data, and substituting physical experiments with simulations; 7) KupferDigital: This is aimed at the digitalization of the life cycle of copper alloys, this project has created several ontologies and knowledge graphs using various ontological frameworks. These efforts address key areas such as processing, postprocessing, characterization, lifecycle, and recycling of copper. Central to the project are the Kupfer*Digital* Ontology, Copper Key Ontology (CKO), and Mechanical Testing Ontology (MTO). The MTO[101] compiles knowledge graphs for various mechanical testing processes, such as tensile, hardness (Brinell and Vickers), stress relaxation, and fatigue testing, conforming to ISO standards and incorporating an extensive mechanical testing terminology.[102] The CKO[103] provides a semantic representation of data from the Copper Key platform,[104] capturing essential features of copper alloys, such as industrial standards, material designations, types, and chemical compositions, positioning the CKO as a potential connector along the entire value chain; 8) LeBeDigital developed the Concrete Production and Testing Ontology (CPTO) aiming to digitally encompass the entire concrete production chain, from mix design to curing and testing. The CPTO forms the foundation of a set of knowledge graph templates designed to describe various process steps such as mixing, hardening, and compressive strength testing;[105] 9) ODE_AM created specialized ontologies to organize data and information from raw materials to final material quality within additive manufacturing domains. These include the Additive Manufacturing AO, the Composite Extrusion Modeling Ontology, the manufacturing AO, the powder bed fusion AO), and the wire arc additive manufacturing ontology. The design of this suite primarily utilizes the BFO as the TLO and the common core ontologies (CCO)[106] as intermediate ones, strategically chosen for efficient data categorization and retrieval in a web-app-based tool;[107] 10) SensoTwin is developed the SensoTwin AO (STAO), concentrating on fiber-reinforced composites tailored to the specific needs of the project. The STAO encompasses various facets such as the materials used, production processes, testing procedures, parameters and results, simulation techniques, and the geometry of components, designed to capture the relationships between these elements; 11) SmaDi is developed an ontology including only bridge axioms to the PMDco, since the employed ontology-based data access (OBDA) approach[108] restricts the ontology to OWL2QL, not allowing for an import; and 12) StahlDigital is focused on experiments, simulations, and workflows in the domain of processing steel sheets for automotive applications. The Steel Process Ontology is automatically generated from spreadsheets, where domain experts provide input and knowledge. Automated updates are published online in a machine and human-readable form.[109]

4.2.3. Development Process and Collaborative Approaches

The development of AO typically follows a structured process, beginning with defining system boundaries and formulating competency questions the ontology aims to address. Domain-specific terminology is identified through interviews with domain experts, including lab managers, technical staff, engineers, and researchers, supplemented by insights from relevant regulations and standards. These terms represent the essential concepts and relationships within the project specific data space, which are then integrated into the AO.

To facilitate this process, visual design and ontology creation tools are used to help domain experts and ontology developers collaboratively conceptualize, refine, and model ontologies. Concept boards[110,111] support expert interactions as part of design activities and thereby the creation of initial models, while the Ontopanel tool[112] provides intuitive, visual modeling of ontologies, bridging the gap between domain experts and technical developers. Once initial vocabularies and models are established, tools, for example, an open-source software editor such as Protégé,[113,114] are used to formalize the ontology, ensuring it adheres to semantic standards.

Highlighted below are collaborative techniques and mechanisms from the MaterialDigital projects: 1) DiProMag introduced an innovative ontology engineering method utilizing OTTR (Reasonable Ontology Templates) templates, leveraging prestructured data common in materials science. This method involves designing and verifying templates, followed by data integration through template instantiation. OTTR templates act as an abstraction layer above RDF, simplifying ontology development for domain experts by hiding RDF's complexities. This approach improves communication between experts, leading to more consistent ontologies, and a flexible engineering process where design decisions can be easily adjusted;[115,116] 2) KupferDigital presents OntoFlow, a collaborative, user-friendly ontology development tool that enables active participation from domain experts, including those with limited technical background. OntoFlow automates the ontology design, implementation, serialization, documentation, and testing process through a modular pipeline of interconnected containers;[117] 3) LeBeDigital developed an approach using Ontopanel[112] to create related templates based on the CPTO. These templates form the data spaces knowledge graph covering various concrete producing and testing methods and processes. They feature unique identifiers for mapping to specific tests or assessments and can be merged during mapping by linking them through a shared entity. This method simplifies the addition and modification of entities;[118] and 4) StahlDigital collects domain knowledge via online spreadsheets for efficient terminology capturing and facilitating collaborative development. An automated pipeline checks, translates, and publishes the input as RDF on GitHub,[119] which ensures long-term storage and version control over ontological terms. W3ID enables content-type based access according to the linked-data protocol.[109]

4.2.4. AI Approaches

Artificial intelligence (AI) approaches, particularly text processing techniques, offer promising solutions to support and expedite ontology engineering processes. These techniques improve language understanding and automate the categorization of key terms. Notable applications from the initiative include the following: 1) PMD: This utilizes natural language processing techniques to speed up the MO development.[92,93] Neural network-based algorithms analyzed a corpus of over 14 k contributions from the "Microscopy and Microanalysis" conference series, identifying key microscopy terms used by the community. These terms were categorized and linked using the PMDco framework, aiming to enhance data management systems by enabling the retrieval of related terms and generating new logical connections through reasoning; and 2) iBain: This explores at Fraunhofer IWM the combination of symbolic modeling, such as ontologies, with data-driven methods to connect and utilize both structured and unstructured data sources jointly. With the growing availability of unstructured scholarly articles,[120] automated methods are needed to extract structured information conforming to existing ontologies. The MaterioMiner dataset,[121] coupled with the Materials Mechanics Ontology,[122] was published to link text entities from literature corpus with ontological classes, facilitating data integration. The dataset, notable for its detailed annotation, supports the training of specialized language models for named-entity recognition, foundation models, and potentially even neurosemantic language models, aiming to populate materials databases and enable new insights.

Looking ahead, AI-based approaches hold significant potential to accelerate and enhance ontology development and application. For instance, large language models can assist with several key tasks in ontology engineering, including (constrained) ontology learning,[123] ontology mapping,[124] and ontology population.[125] In ontology learning, language models and semantic search methods can retrieve relevant, diverse texts from a corpus to help identify unbiased, objective, and representative concepts, classes, properties, and axioms. Another promising avenue is enabling natural language querying of RDF graph repositories, making it easier for less experienced users to interact with these systems compared to complex SPARQL queries. However, challenges remain concerning the robustness of general language models, particularly their inconsistent performance in domain-specific settings, which remains a limiting factor.[126] Nonetheless, the emergence of domain-specialized large language models,[127] along with multiagent systems equipped with access to domain-specific tools and functions,[128] offers promising solutions to these challenges in materials science. Machine-processable semantic shapes, such as graph patterns and SHACL shapes, should be leveraged for AI training. However, human expertise will remain essential in knowledge engineering for the foreseeable future.

4.2.5. Extension of the Ontological Framework via Application Ontologies

Utilizing a shared higher-level ontology, such as the PMDco v2.0.7, offers the advantage of allowing AO built on it to specify and extend general concepts in materials science and engineering. This process leads to the creation of a coherent structure of integrated concepts, classes, properties, and relationships. Integrating these domain-specific AO into the ontological framework makes the knowledge implemented in various projects across different areas of materials science and engineering and its created data spaces interoperable.

The systematic structuring facilitated by PMD's midlevel concepts enables the consolidation of specific extensions through compatible AO (see **Figure 6**). The detailed extension involves adding new subclasses predominantly under pmd:Process, pmd:ProcessingNode, pmd:Object, and pmd:ValueObject, which is specified in **Table 3**.

However, it should be noted that this table only includes all newly integrated subclasses of selected PMD concepts and a detailed evaluation, in particular the curation of these, is still pending. The numbers given are therefore preliminary

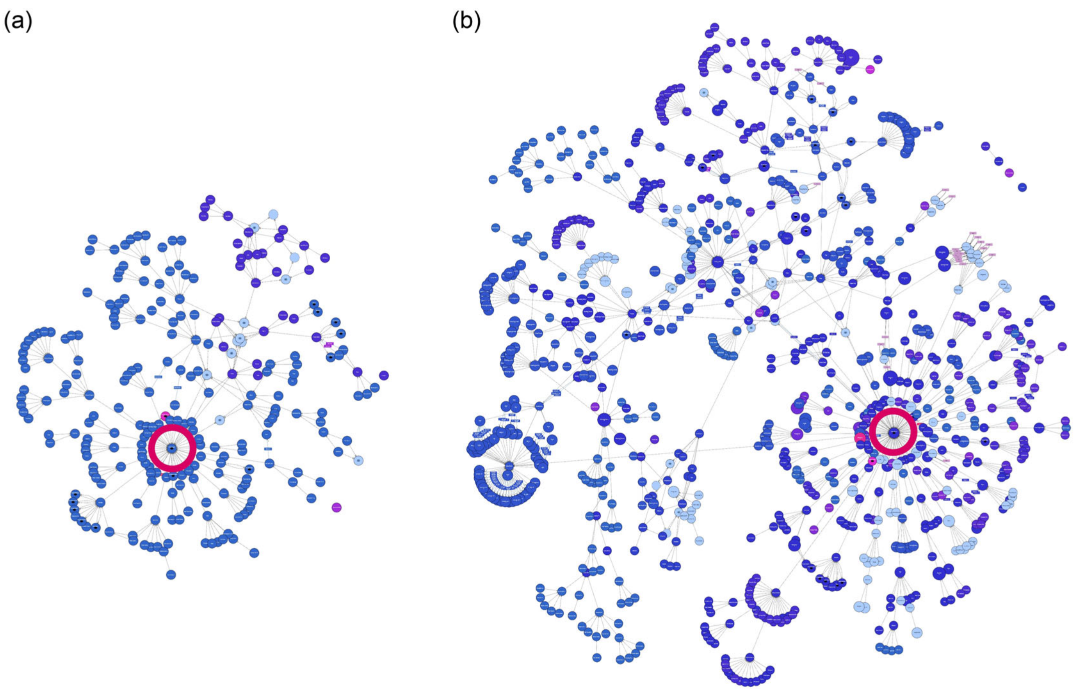

Figure 6. Comparison of the TBox structure of PMDco v2.0.7 a) prior and b) following its integration with various AO, including TTO, MO, PGO, MTO, CPTO, and STAO. The expansion depicted underscores how the framework extends through ontology integration. The pmd:ValueObject highlighted in red signifies a key element of this growth. The visualizations have been generated with WebVOWL.[75]

Table 3. Detailed extension of PMDco v2.0.7 key concepts via AO.

Project Name	Application Ontology	Process	Processing Node	Object	Value Object
PMD	TTO	1	2	–	44
PMD	MO	56	37	11	38
DigiBatMat	BPCO	–	4	–	–
DiProMag	DMCO	15	21	4	55
KNOW-NOW	KN	17	6	7	51
KupferDigital	MTO	55	28	–	196
LeBeDigital	CPTO	13	9	47	42
SensoTwin	STAO	52	141	16	82
Sum	8	209	248	85	508

estimates. Nevertheless, it is clear that with each AO, a large number of classes extend the PMDco class tree.

Several positive outcomes can already be noted for the ontological framework. For instance, the integration allows for a more refined development of the midlevel, including the addition of essential but previously missing concepts that are identifiable by different AOs. Furthermore, this process facilitates the validation of whether the midlayer concepts are sufficiently well defined, thus providing a measure of concept accuracy. The curation of the newly introduced classes offers valuable insights for both the PMDco and the corresponding AO. A comprehensive presentation of these results is planned.

4.3. MaterialDigital Knowledge Graph

The following describes the shared, prototypical knowledge graph,[129] which incorporates the ontological framework based on the PMDco v2.0.7 along with compatible AO and alternative frameworks from the MaterialDigital projects (see Section 4.3.1). It also details how this knowledge graph synergistically integrates with other knowledge graphs (see Section 4.3.2). Further, it outlines the necessary steps toward achieving far-reaching interoperability and cooperation (see Section 4.3.3).

4.3.1. Prototypical MaterialDigital Knowledge Graph

An initial shared knowledge graph has been created using the Open version of OpenLink Virtuoso.[130] Alongside the framework of the PMDco, all available AO from the projects[82] have been integrated as named graphs via script. This knowledge graph provides a functional basis for future expansions and optimizations, currently enabling SPARQL queries through the triple store. It is important to note that, as of now, only RDF structures can be discovered according to their underlying TBoxes, since no ABoxes with data have yet been integrated. The gradual integration of specific data spaces from MaterialDigital projects is planned to consolidate this further.

Predefined queries have been set up to facilitate beginners' use of SPARQL, enabling them to easily navigate through stored graphs. Example queries include Count Triples, Show PMDco Classes, Show All Graphs (Ontologies), Class Lookup, Show ProcessingNodes w. Inferencing, and Federated Query to Matwerk KG. An example of a flexible class query is illustrated in **Figure 7**.

As Virtuoso acts as an OWL reasoner, inference operations such as rdfs:subClassOf and rdfs:subPropertyOf are supported, as utilized by the Show ProcessingNodes w. Inferencing. Especially for mappings to PMDco concepts, the relationships owl:sameAs, owl:equivalentClass, and owl:equivalentProperty are useful. For long-term applications, reasoning can be invaluable for automatically validating inconsistencies, thus supporting data quality control. Moreover, there is potential to enhance Virtuoso's user interface by integrating tools such as LodLive,[131] Triply,[132] and SHMARQL.[133]

4.3.2. Alignment with Other Knowledge Graphs

Aligning with other existing knowledge graphs is an intended objective to make data spaces across domain boundaries interoperable and searchable. A preliminary demonstration is provided by the predefined Federated Query to MatWerk KG. This query accesses the NFDI-MatWerk knowledge graph,[134] searching for pmd:ValueObject, and returns 1652 entries derived from an extensive dataset. This illustrates the value added by linked data and the importance of standardized, semantically interoperable ontological frameworks.

4.3.3. Future Direction

With the initial phase of MaterialDigital concluded, the PMDco v2.0.7 has been established as a cornerstone for an evolving ontological framework. It is used across shared knowledge graphs, various demonstrators, and projects, establishing itself as a user-friendly and robust MLO for materials science and engineering.

From this point forward, the focus will shift toward the collaborative advancement of the PMDco, responding to the industry's demand for a standardized TLO and leveraging insights from the prior phase. As a result, the PMDco v3.0 is set to be developed in alignment with the BFO v2.0.[78]

The future structure of the ontology will adopt a modular approach, segmenting into distinct modules such as Material, Material Qualities, Material Manufacturing, Material Characterization, Data Transformation, Logistics, and Devices. This modularization aims to focus development efforts. The inherent functionalities of BFO will also facilitate the gradual integration of reasoning capabilities into PMDco v3.0, utilizing defined roles and functions.

Semantic schemata and shapes will play a crucial role in its development process. Notably, ontology design patterns will address common challenges in ontology modeling and are foundational building blocks for conceptual models.[135,136] Providing these patterns in a well-documented form in the PMDco GitHub repository will enable their systematic reuse in ontology development, enhancing reproducibility and structural consistency. This repository[77] actively supports iterative enhancements, with users currently able to raise issues and contribute feedback, alongside participating in the Ontology Playground

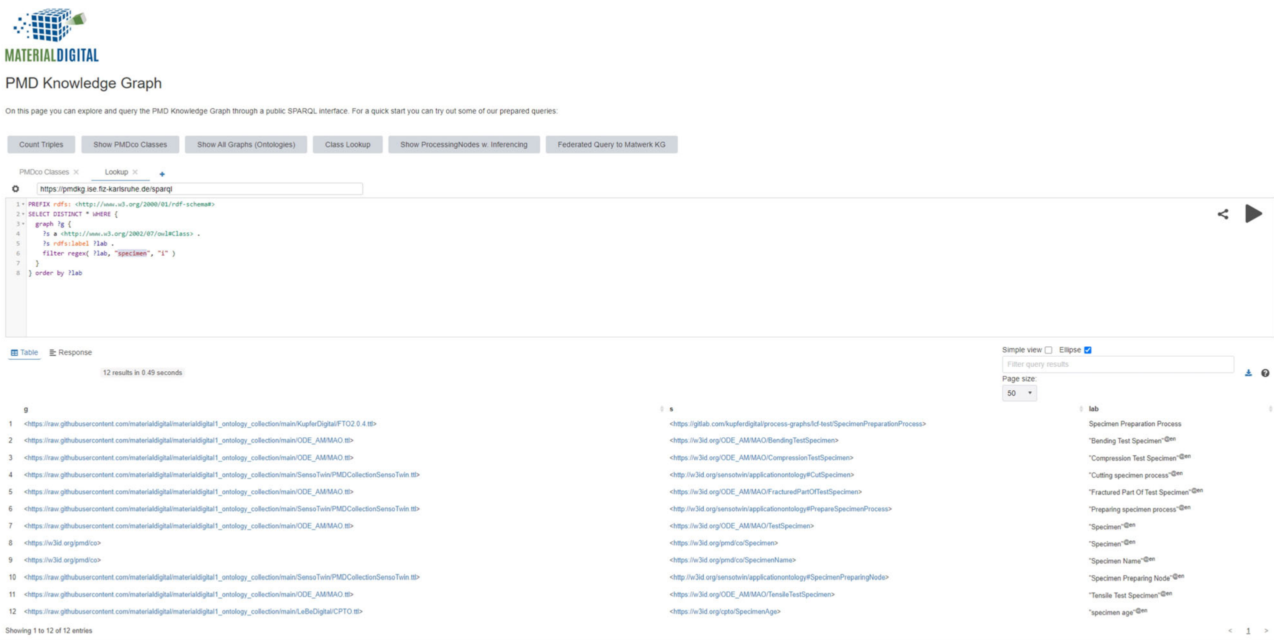

Figure 7. Shared knowledge graph SPARQL query on the specific pmd:Specimen class using predefined Class Lookup.

(see Section 3.1). Furthermore, the provision of shape-specific SPARQL queries will facilitate systematic expansion of the ontological framework around BFO v2.0-PMDco v3.0, which also benefits reasoning rules. This strategy increases interoperability between the framework extensions, that is, the AO, and also makes the materials science knowledge graph more consistent and the ontological axioms transferable. Moreover, data structures become more uniform, information rich, and accessible for SPARQL queries. Identifying patterns as generically as possible for materials science and engineering strikes a balance between generalizability and realistic constraints, paving the way for future AO in MaterialDigital to be interoperable across various domains, including those based on BFO, as seen in intersecting fields like biology and medicine.[137] Looking ahead, current measures to comply with the FAIR principles should be further developed,[138] considering existing groundwork,[139,140] and implemented into standardized metrics and mechanisms for systematic integration into the MaterialDigital projects.

4.4. Summary and Key Insights

In summary, this work demonstrates the successful integration of initial project data space contributions into a shared and scalable knowledge graph prototype, showcasing practical applications. The future enhancement of the PMDco and the shared ontological framework required for semantic interoperability will be further driven by the development and adoption of specific AO, promoting systematic integration into knowledge graphs. Social interactions, facilitated by forums such as the Ontology Playground, play a crucial role in fostering a shared understanding, building consensus, and facilitating the creation of a collaborative data environment. The use of innovative and collaborative tools for ontology development and application advances digitalization in materials science and engineering. Moreover, to ensure long-term success, the community should focus on developing training tools and guides that serve as a foundation for workshops and social tech events, effectively introducing non-experts to the topic.

The key insights from this section are as follows: 1) Technical skills need to be enhanced and consolidated for digitalization requirements; 2) Ontology design patterns and shorter evaluation cycles will improve ontological framework; 3) Stricter rules and feedback mechanisms in repositories will reduce integration efforts; 4) The PMDco v3.0 will adopt the BFO as TLO for cross-domain interoperability; 5) Automation, including AI methods, will accelerate ontology engineering; 6) Flexible interfaces satisfy user requirements and simplify data integration; and 7) Metrics and case studies should be established to assess level of semantic interoperability.

5. Materials Research Benefiting from Ontologies

This section highlights how various projects of the MaterialDigital Initiative have combined ontologies with data on materials and processes, along with workflows, to enable a wide range of applications in their data spaces. These approaches facilitate dynamic operations for data storage and querying, ranging from materials and processes design to service life solutions.

5.1. Optimizing Battery Materials Production Through Data Interlinking

The DigiBatMat project has developed a digital platform dedicated to battery materials and their production processes, utilizing the BPCO as the data model. This collaborative effort involved partners from four distinct areas of battery research, along with specialists in knowledge graphs and digital systems, creating a shared digital structure that interlinks material, production, and characterization data. The ontology-driven platform facilitates data exchange and enables efficient analysis of correlations between material properties and production parameters. This is critical for understanding the interplay between these factors. The platform has been validated through specific research on three use cases: 1) optimizing mixing parameters for conductive additives; 2) comparing properties of active materials NCA and NCM811 under high current rates; and 3) optimizing the electrochemical performance of lithium-titanium-oxide (LTO)/lithium-ion-manganese-oxide (LMO) cells. The performance of Li-ion cells, influenced by material properties and production parameters, shows how even minor modifications can significantly affect energy, power density, aging, and safety. The digital platform, enriched with comprehensive data, supports efficient analysis, enabling the identification of correlations between material properties and performance.

5.2. Enhancing Glass Materials Design with Ontology-Based Knowledge Management

The GlasDigital project employs its glass ontology within a decentralized web-based Semantic Knowledge Base (SKB), developed using Fast Ontodocker.[141] This SKB meticulously stores all parameters of the glass synthesis process alongside SciGlass data, enabling searches for both experimental and simulated properties. Integrating raw data from the glass synthesis process with the ontology, the SKB enhances the evaluation and management of process parameters across the entire process chain. The SKB also integrates tools for property prediction using machine learning algorithms and for extracting tabular information[142] from existing patents. These tools, along with visualization aids such as ternary phase diagrams to visualize the glass composition and time series plots to visualize time-dependent parameters, empower glass experts to monitor each step of the process and make informed decisions to refine it. Consequently, with the support of this ontology-based SKB, glass experts are equipped to design new glass materials with optimally tailored properties.

5.3. Advancing Predictive Modeling in Ceramics Using Triple Store

The KNOW-NOW project has developed an innovative demonstrator for predictive modeling of electrical coil performance, incorporating both geometry and material parameters.[143] In this context, data is systematically linked by annotating and storing it in an ontology connected to a triple store. The KN ontology serves as a structured knowledge base, precisely annotating relevant information on geometric and material-related parameters, as well as experimental data. Using SPARQL queries, data stored ontologically can be efficiently retrieved, linked, and integrated into a workflow. This workflow automates the process and provides users with access to ontologically structured data, allowing flexible parameter adjustments through a user-friendly interface. This automated workflow seamlessly integrates the data with COMSOL Multiphysics simulation, ensuring holistic and efficient data management. After the simulation, results are fed back into the ontology and triple store, creating a consistent database for future analysis, development, and predictive modeling. A data pipeline supports any user to easily upload data to a server, where an automated process realizes the storage linked to the KN ontology.

5.4. Ontological Framework for Optimizing 3D Printing Processes

The ODE_AM project uses an ontological framework as a semantic backbone, integrating diverse data sources, from powder characteristics to machine parameters, with part qualities such as compression strength and fatigue behavior. This framework covers the entire manufacturing process, from raw materials to the characteristics of manufactured parts. A web application, built with Python backend, features specialized importers for mapping data from sources like spreadsheets from tensile testing machines to the classes within the ontological framework. Initially, necessary data are parsed from raw data files and formatted into triples according to the guidelines of BFO and CCO. This framework supports two primary use cases: the creation of digital part record, which captures characteristics and manufacturing parameters for each part, and a material/process data dashboard that uses SPARQL queries to explore correlations between processing factors such as part orientations and resulting material properties. This configuration is invaluable for selecting optimal parameter windows and part designs by demonstrating the variability of material characteristics.

5.5. Estimating Service Life of Wind Turbine Blades Using Ontological Data Integration

The SensoTwin project employs the STAO as a central interface in a workflow covering the entire lifecycle of glass fiber-reinforced polymer wind turbine rotor blades, enabling module interoperability and interchangeability. This multiscale approach spans from base material to process to component. Integrating data from experiments, simulations, and calculations into a knowledge graph ensures ready access to appropriate inputs for various simulations and calculations. The use of Owlready2 module[144] facilitates managing the STAO, aiding in storing and retrieving information from tests, simulations, and calculations. This approach leverages ontology data, combined with weather conditions, to calculate the remaining service life of the turbine blades.

5.6. Smart Materials Design Facilitated by Ontology-Based Data Access

The SmaDi project employs an OBDA approach[108] to facilitate the retrieval of relevant information, such as hyperelastic material models and specific parameters, along with the context of how this data was collected for smart materials from diverse datasets. This approach uses an ontology and SPARQL queries[145] and has been extended to ontology-based data and model access, integrating data access with model-based workflow steps.[146–148]

This approach enhances smart materials design by enabling the determination of specific, indirectly stored requirements in the material selection process.

5.7. Summary and Key Insights

In summary, the application of ontologies in the MaterialDigital Initiative has demonstrated that semantic data integration from various sources enables efficient and flexible data management across different data spaces, thus fundamentally supporting materials research and development. Functionally linking digital workflows of experimental processes and simulations to ontological frameworks simplifies ongoing integration and data exchange, thereby enabling knowledge graphs to expand in depth and complexity.

The key insights from this section are as follows: 1) Ontology-based data management facilitates systematic correlation of materials parameters with production and process data; 2) Ontologies enable the combination of data from both experimental and computational workflows; and 3) Workflows benefit from ontology-based data for processing inputs and outputs, as well as for simulation steps.

6. Current Approaches Elsewhere: Contact Points, Interfaces, Synergies

The MaterialDigital Initiative proactively collaborates with national and international initiatives to prevent parallel developments and seek joint solutions in the digital transformation of materials science and engineering. However, the full potential for collaboration is yet to be realized. This section aims to provide an overview of current initiatives and their approaches, identifying common points of contact, interfaces, and synergies to amplify cooperation in digitalization.

6.1. National and European Initiatives

The German National Research Data Infrastructure (NFDI)[149] unites 26 consortia from interdisciplinary institutions collaborating on accessing, connecting, and sustainably harnessing scientific research data in line with the FAIR principles.

FAIRmat,[150] an NFDI consortium, is building the largest data infrastructure in computational materials science with the Novel Materials Discovery (NOMAD) Laboratory[151] and promotes data integration and use through management and analysis tools. It focuses particularly on experimental and theoretical condensed matter physics and chemistry. The FAIR-DI e.V. association,[152] with its European member institutions, aims to permanently establish NOMAD as a significant resource for big data from materials science, engineering, and astronomy.

The consortium NFDI-MatWerk[153,154] specializes in developing a digital data infrastructure for materials science and engineering that encompasses complex relationships between materials data and makes them technologically accessible for broad use, incorporating a materials ontology and a graph database infrastructure.

As a national organization, NFDI collaborates with the European Open Science Cloud (EOSC),[155] which aims to create a FAIR data and services web for science in Europe, where data, tools, and services can be published, made discoverable, and reused. Common interfaces and metadata standards should enable interoperable digital objects to be used and exploited by academic, private, and public users to the fullest extent possible.

Gaia-X,[156] a multinational initiative with industry engagement, supports open data principles with a broader focus and seeks a secure, federated European data infrastructure to strengthen the competitiveness of European companies. This involves developing strategies, rules and specifications aligned with European values. Gaia-X and EOSC share compatible visions of federated, secure data spaces and ecosystems that promote digital sovereignty and innovation in Europe, which is beneficial for extending the EOSC to both the public and private sectors. Related to the objectives of Gaia-X are Catena-X and Manufacturing-X. Catena-X,[157] as an initial implementation project, focuses on collaborative digitalization along supply lines in the automotive industry to optimize business processes through a data-driven value chain. Manufacturing-X,[158] currently being defined and implemented by the Plattform Industrie 4.0,[159] aims to achieve a fully digitally networked industry with the capability to confidently and collaboratively use data across the entire manufacturing and supply chain. In conjunction with Gaia-X and Catena-X, basic building blocks and prototypes for a collective Industry 4.0 data space are being created.

The FAIR Data Spaces project,[160] a collaboration between NFDI and Gaia-X, is developing a shared, cloud-based data space for industry and research, implementing Gaia-X's standards and technical requirements for sovereign data exchange. The International Data Spaces Association (IDSA)[161] and Gaia-X pursue complementary goals to promote secure and sovereign data spaces in Europe, with IDSA focusing on creating a standardized framework for secure data exchange and Gaia-X on strengthening transparency and trust in digital ecosystems.

In France, the DIADEM program,[162] led by CNRS and CEA, aims to set up integrated platforms for material innovation. It supports the introduction of innovative materials using AI, focusing on developing sustainable high-performance materials to strengthen French industry and economic growth, while supporting demonstration projects and research initiatives with an international focus.

6.2. Non-European and Global Initiatives

Beyond Europe, the National Institute for Materials Research (NIMS) in Japan promotes the development and selection of new materials through the MatNavi database,[163] which contains extensive information on polymers, inorganic, and metallic materials, and electronic structures, as well as tools for composite material design and property prediction.

In the U.S., the Materials Genome Initiative (MGI),[164,165] a federal government-wide initiative since 2011, fosters the rapid and cost-effective discovery, development, and deployment of new materials by creating a collaborative framework integrating advanced modeling techniques, computational tools, and experimental data, embracing Semantic Web technologies prevalent in materials science and engineering since MGI's inception.

The research data alliance (RDA)[166] is building a global social and technical bridge for the open use and reuse of data, supporting relevant groups such as the established RDA/CODATA Materials Data, Infrastructure & Interoperability IG,[167] the Harmonised terminologies and schemas for FAIR data in materials science and related domains, which is pending endorsement,[168] and the related Materials Research Data Alliance (MaRDA),[169] as cornerstones of international materials science collaboration.

7. Integration Between the MaterialDigital Initiative and Others

The MaterialDigital Initiative is pioneering the development and establishment of materials data spaces that encompass entire value chains. These data spaces exemplify and facilitate the joint use, comparison, and transferability of data across different domains, benefiting the community. This approach promotes data-centric methodologies, including AI applications, and supports the creation of interoperable data spaces and knowledge graphs for systematic access to materials data. Furthermore, the MaterialDigital Initiative aims to remain compatible with European infrastructures while adhering to FAIR principles and ensuring data security, transparency, and sovereignty.

A key to the successful integration of the MaterialDigital Initiative with others is adherence to and promotion of FAIR principles in data management and tool development. Special emphasis is placed on ensuring the reproducibility of materials and processes data through comprehensive metadata acquisition and representation (see Section 2.4). Semantic interoperability is achieved through a shared, accepted, and compatible ontological framework that incorporates standardized, industry-compatible TLOs (see Section 4.1.2) and utilizes uniform metadata standards (see Section 4.1.3). Materials science data prepared for integration into European infrastructures must be appropriately formatted, employing widely accepted and generic metadata vocabularies for dataset descriptions, and must fully capture the specificity of the data space and its generated data structure using the ontological framework, which is extended by AO.

To realize this, ongoing efforts within MaterialDigital and NFDI-MatWerk to develop a compatible ontological framework will be sustained and strengthened. Adopting the Basic Formal Ontology (BFO) marks a strategic move towards an industry-accepted and cross-domain description of materials and processes data (see Section 4.3.3). Additionally, active participation in the newly established materials science working group of the Industrial Ontologies Foundry[170] aims to solidify the Initiative's presence there. Existing contacts with the European Materials Modelling Council,[171] which supports the integration of material modeling and digitalization to make product development more flexible and sustainable, and with OntoCommons,[172] a project to standardize the documentation of data in all areas related to materials and manufacturing, are also maintained.

The continuous development of the community-driven Ontology Playground is crucial for fostering mutual understanding and encouraging participation. The implementation of automated mechanisms for publishing TBoxes on MaterialDigital's Git Platform, along with the use of terminology services[173] and ontology repositories,[174] will aid in harmonizing the curation of TBoxes and vocabularies. Efforts to align with European infrastructure initiatives like GAIA-X and EOSC include adopting globally accepted metadata standards and vocabularies for describing datasets (see Section 4.1.3).

To increase adoption and raise awareness within the international community, the MaterialDigital Initiative will continue to develop, make available, and present both understandable and customizable demonstrators at international conferences and in journals (see Sections 3.2 and 3.3, respectively). Organizing a symposium on Digital Materials: Experiments, Simulation Workflows, Ontologies and Interoperability at EUROMAT 2023, in collaboration with FEMS,[175] a nonprofit federation of European material societies across 22 countries, contributed to this effort and provided a forum for scientific exchange and acceptance.

The development of ELN templates, as already used in demonstrators (see Section 3.4), enables practical implementation in everyday laboratory work. This simplifies the application of technical solutions using an ontological framework, which enables the transformation of own data into interoperable RDF datasets.

Overall, this strategic direction aims to promote the sustainable integration of the MaterialDigital Initiative with other initiatives, thereby improving how materials science and related fields work and operate.

8. Conclusion

In conclusion, the digitalization of materials science and engineering mainly advances through the dedicated efforts of its subdomains and the creation of their specific digital data spaces. Nevertheless, there is an increased need for overarching, more standardized structures to ensure synergies and interoperability, as well as sustainable and international progress beyond domain boundaries. A cornerstone of the MaterialDigital Initiative is the continuous, collaborative interaction and exchange on semantic interoperability in the Ontology Playground, which significantly promotes a common understanding and consequently a shared vision of an interoperable, densely interconnected data landscape.

The ontological framework, which is essential for the functioning of a shared knowledge graph and the access to broad data across different scales, is informed, extended, and enhanced by AO developed for specific subdomain data spaces. The insights from the MaterialDigital Initiative and its prototype construction of a shared knowledge graph have demonstrated the necessity of integrating the BFO as a standardized and industry-compatible top level. The PMDco will use the BFO in its new version to ensure crossdomain interoperability within materials science and engineering. In doing so, the development of AO will be more harmonized through ontological design shapes and patterns, while stricter software constraints and shorter feedback mechanisms will enhance solutions and minimize integration efforts. The ontological framework created in this way will enable the uniform and consensual semantic description of materials and processes data. At the same time, data catalogs and their generic, standardized vocabulary, which likewise constitute the ontological framework, will make it possible to identify and specify dataset resources in accordance with other infrastructure initiatives.

The ontology-based approach has the potential to interconnect data from a wide variety of sources in a standardized way and to flexibly correlate them to fundamentally support data-driven approaches by optimizing data management and significantly transform materials research and development.

In the long run, digitalization in materials science and engineering may only succeed if solutions being developed are accepted by the community and consequently integrated into daily laboratory and research routines. Key to this success is the development and availability of innovative, collaborative tools and interfaces for data integration, featuring user-friendly GUIs and solutions that provide tangible benefits for users. Success will also depend on whether the newly skilled personnel and the technical expertise they have acquired can be consolidated so that the coming demands of digitalization can be met.

In this respect, the application of the approaches developed in MaterialDigital across other international initiatives—each addressing specific scenarios for the integration of heterogeneous data into a common framework—can be significantly strengthened through comparative analysis. By coordinating efforts, sharing knowledge, and adopting recognized digital standards alongside increased commitment to the FAIR principles, these initiatives can collaboratively advance digitalization in materials science and engineering, leading to more sustainable and innovative use of materials on a large scale. While promising, this transformation still offers considerable room for improvement and further exploration, particularly through detailed evaluation and ongoing refinement.

Acknowledgements

The authors thank the German Federal Ministry of Education and Research (BMBF) for financial support of the project Innovation-Platform MaterialDigital through project funding grant nos. 13XP5094E (BAM), 13XP5094F (FIZ), 13XP5094C (MPIE), 13XP5094A (KIT), 13XP5094D (IWT), and 13XP5094B (IWM).

Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Bernd Bayerlein: conceptualization (equal); methodology (equal); supervision (equal); visualization (equal); writing—original draft (lead); writing—review & editing (lead). **Jörg Waitelonis**: methodology (equal); visualization (supporting); writing—review & editing (supporting). **Henk Birkholz**: conceptualization (equal); visualization (equal); writing—original draft (supporting); writing—review & editing (supporting). **Matthias Jung**: methodology (equal). **Markus Schilling**: writing—original draft (supporting). **Philipp v. Hartrott**: writing—original draft (supporting); writing—review & editing (supporting). **Marian Bruns**: writing—original draft (supporting). **Jörg Schaarschmidt**: writing—original draft (supporting). **Kristian Beilke**: writing—original draft (supporting). **Marcel Mutz**: writing—original draft (supporting); writing—review & editing (supporting). **Vincent Nebel**: writing—original draft (supporting); writing—review & editing (supporting). **Veit Königer**: writing—original draft (supporting); writing—review & editing (supporting). **Lisa Beran**: writing—original draft (supporting); writing—review & editing (supporting). **Tobias Kraus**: writing—original draft (supporting); writing—review & editing (supporting). **Akhilesh Vyas**: writing—original draft (supporting); writing—review & editing (supporting). **Lars Vogt**: writing—original draft (supporting); writing—review & editing (supporting). **Moritz Blum**: writing—original draft (supporting); writing—review & editing (supporting). **Basil Ell**: writing—original draft (supporting); writing—review & editing (supporting). **Ya-Fan Chen**: writing—original draft (supporting); writing—review & editing (supporting). **Tina Waurischk**: writing—original draft (supporting); writing—review & editing (supporting). **Akhil Thomas**: writing—original draft (supporting); writing—review & editing (supporting). **Ali Riza Durmaz**: writing—original draft (supporting); writing—review & editing (supporting). **Sahar Ben Hassine**: writing—original draft (supporting); writing—review & editing (supporting). **Carina Fresemann**: writing—original draft (supporting); writing—review & editing (supporting). **Gordian Dziwis**: writing—original draft (supporting); writing—review & editing (supporting). **Hossein Beygi Nasrabadi**: writing—original draft (supporting); writing—review & editing (supporting). **Thomas Hanke**: writing—original draft (supporting); writing—review & editing (supporting). **Melissa Telong**: writing—original draft (supporting); writing—review & editing (supporting). **Stephan Pirskawetz**: writing—original draft (supporting); writing—review & editing (supporting). **Mohamed Kamal**: writing—original draft (supporting); writing—review & editing (supporting). **Thomas Bjarsch**: writing—original draft (supporting); writing—review & editing (supporting). **Ursula Pähler**: writing—original draft (supporting); writing—review & editing (supporting). **Peter Hofmann**: writing—original draft (supporting); writing—review & editing (supporting). **Mena Leemhuis**: writing—original draft (supporting); writing—review & editing (supporting). **Özgür L. Özçep**: writing—original draft (supporting); writing—review & editing (supporting). **Lars-Peter Meyer**: writing—original draft (supporting); writing—review & editing (supporting). **Birgit Skrotzki**: conceptualization (supporting); writing—review & editing (supporting). **Jörg Neugebauer**: conceptualization (supporting). **Wolfgang Wenzel**: conceptualization (supporting). **Harald Sack**: conceptualization (equal). **Chris Eberl**: conceptualization (equal). **Pedro Dolabella Portella**: conceptualization (equal); supervision (equal); writing—original draft (supporting); writing—review & editing (supporting). **Tilmann Hickel**: conceptualization (equal); writing—original draft (supporting); writing—review & editing (supporting). **Lutz Mädler**: conceptualization (equal); supervision (equal); visualization (equal); writing—review & editing (supporting). **Peter Gumbsch**: conceptualization (equal); writing—review & editing (supporting).

Keywords

data spaces, knowledge graphs, knowledge representations, materials science and engineering, Platform Materialdigital, semantic data integrations, semantic interoperabilities

Received: May 6, 2024
Revised: October 21, 2024
Published online:

[1] M. F. Ashby, H. Shercliff, D. Cebon, *Materials: Engineering, Science, Processing and Design*, 4th ed., Butterworth-Heinemann, Oxford, UK **2018**.
[2] L. Lin, P. S. Kollipara, Y. Zheng, *Mater. Today* **2019**, *28*, 49.

[3] M. Alam, F. Dittmann, M. Niebel, J. S. Lehmann, D. Dessí, J. F. Morgado, P. v. Hartrott, C. Eberl, P. Gumbsch, H. Sack, *White Paper* **2020**, pp. 1–10, https://api.semanticscholar.org/CorpusID:220902452 (accessed: February 2024).
[4] J. Kimmig, S. Zechel, U. S. Schubert, *Adv. Mater.* **2021**, *33*, 2004940.
[5] T. Berners-Lee, J. Hendler, O. Lassila, *Sci. Am.* **2001**, *284*, 34.
[6] T. R. Gruber, *Knowl. Acquis.* **1993**, *5*, 199.
[7] B. G. Pelkie, L. D. Pozzo, *Digital Discovery* **2023**, *2*, 544.
[8] B. Bayerlein, M. Schilling, P. v. Hartrott, J. Waitelonis, *Sci. Data* **2024**, *11*, 434.
[9] A. Hogan, E. Blomqvist, M. Cochez, C. D'amato, G. D. Melo, C. Gutierrez, S. Kirrane, J. E. L. Gayo, R. Navigli, S. Neumaier, A.-C. N. Ngomo, A. Polleres, S. M. Rashid, A. Rula, L. Schmelzeisen, J. Sequeda, S. Staab, A. Zimmermann, *ACM Comput. Surv.* **2021**, *54*, 1.
[10] B. Bayerlein, T. Hanke, T. Muth, J. Riedel, M. Schilling, C. Schweizer, B. Skrotzki, A. Todor, B. Moreno Torres, J. F. Unger, C. Völker, J. Olbricht, *Adv. Eng. Mater.* **2022**, *24*, 2101176.
[11] A. Valdestilhas, B. Bayerlein, B. Moreno Torres, Z. Ghezal Ahmad Jan, T. Muth, *Adv. Intell. Syst.* **2023**, *5*, 2300051.
[12] V. Venugopal, E. Olivetti, *Sci. Data* **2024**, *11*, 1.
[13] X. Zhang, C. Zhao, X. Wang, *Comput. Ind.* **2015**, *73*, 8.
[14] B. Bayerlein, M. Schilling, H. Birkholz, M. Jung, J. Waitelonis, L. Mädler, H. Sack, *Mater. Des.* **2024**, *237*, 112603.
[15] D. Helm, A. Butz, D. Raabe, P. Gumbsch, *JOM* **2011**, *63*, 26.
[16] L. E. Murr, *Handbook of Materials Structures, Properties, Processing and Performance*, 1st ed., Springer, Cham **2015**.
[17] S. R. Kalidindi, M. De Graef, *Annu. Rev.* **2015**, *45*, 171.
[18] A. Prakash, S. Sandfeld, *Pract. Metall.* **2018**, *55*, 493.
[19] L. Himanen, A. Geurts, A. S. Foster, P. Rinke, *Adv. Sci.* **2020**, *7*, 1900808.
[20] A. R. Durmaz, M. Müller, B. Lei, A. Thomas, D. Britz, E. A. Holm, C. Eberl, F. Mücklich, P. Gumbsch, *Nat. Commun.* **2021**, *12*, 6272.
[21] A. Thomas, A. R. Durmaz, M. Alam, P. Gumbsch, H. Sack, C. Eberl, *Sci. Rep.* **2021**, *13*, 12562.
[22] G. A. J. Zia, T. Hanke, B. Skrotzki, C. Völker, B. Bayerlein, *Integr. Mater. Manuf. Innovation* **2024**, *13*, 257.
[23] A. Agrawal, A. Choudhary, *APL Mater.* **2016**, *4*, 053208.
[24] S. Bauer, P. Benner, T. Bereau, V. Blum, M. Boley, C. Carbogno, C. R. A. Catlow, G. Dehm, S. Eibl, R. Ernstorfer, A. Fekete, L. Foppa, P. Fratzl, C. Freysoldt, B. Gault, L. M. Ghiringhelli, S. K. Giri, A. Gladyshev, P. Goyal, J. Hattrick-Simpers, L. Kabalan, P. Karpov, M. S. Khorrami, C. T. Koch, S. Kokott, T. Kosch, I. Kowalec, K. Kremer, A. Leitherer, Y. Li, et al., *Modell. Simul. Mater. Sci. Eng.* **2024**, *32*, 063301.
[25] M. M. Flores-Leonar, L. M. Mejía-Mendoza, A. Aguilar-Granda, B. Sanchez-Lengeling, H. Tribukait, C. Amador-Bedolla, A. Aspuru-Guzik, *Curr. Opin. Green Sustainable Chem.* **2020**, *25*, 100370.
[26] E. Stach, B. DeCost, A. G. Kusne, J. Hattrick-Simpers, K. A. Brown, K. G. Reyes, J. Schrier, S. Billinge, T. Buonassisi, I. Foster, C. P. Gomes, J. M. Gregoire, A. Mehta, J. Montoya, E. Olivetti, C. Park, E. Rotenberg, S. K. Saikin, S. Smullin, V. Stanev, B. Maruyama, *Matter* **2021**, *4*, 2702.
[27] M. J. Statt, B. A. Rohr, D. Guevarra, S. K. Suram, J. M. Gregoire, *Digital Discovery* **2024**, *3*, 238.
[28] C. P. Gomes, B. Selman, J. M. Gregoire, *MRS Bull.* **2019**, *44*, 538.
[29] P. Gumbsch, in *Der Stoff, aus dem Innovationen sind*, 1.WING-Konferenz, Weimar, 29–31.10.2003, Bundesministerium für Bildung und Forschung (BMBF) **2003**, pp. 70–81.
[30] A. R. Durmaz, N. Hadzic, T. Straub, C. Eberl, P. Gumbsch, *Exp. Mech.* **2021**, *61*, 1489.
[31] N. T. Garabedian, P. J. Schreiber, N. Brandt, P. Zschumme, I. L. Blatter, A. Dollmann, C. Haug, D. Kümmel, Y. Li, F. Meyer, C. E. Morstein, J. S. Rau, M. Weber, J. Schneider, P. Gumbsch, M. Selzer, C. Greiner, *Sci. Data* **2022**, *9*, 315.
[32] M. Wilkinson, M. Dumontier, I. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. Bonino da Silva Santos, P. Bourne, J. Bouwman, A. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C. T. Evelo, R. Finkers, A. Gonzalez-Beltran, A. J. G. Gray, P. Groth, C. Goble, J. S. Grethe, J. Heringa, P. A. C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, et al., *Sci. Data* **2016**, *3*, 160018.
[33] L. C. Brinson, L. M. Bartolo, B. Ben, D. Elbert, I. Foster, A. Strachan, P. W. Voorhees, *MRS Bull.* **2024**, *49*, 12.
[34] Bundesministerium für Bildung und Forschung (BMBF) and VDI Technologiezentrum GmbH, MaterialDigital - Digitalisierung der Materialforschung, https://www.bmbf.de/SharedDocs/Publikationen/de/bmbf/5/31701_MaterialDigital.pdf?__blob=publicationFilev=5 (accessed: March 2022).
[35] MaterialDigital Initiative, https://www.materialdigital.de/ (accessed: April 2024).
[36] S. Bekemeier, C. Rêgo, H. Mai, U. Saikia, O. Waseda, M. Apel, F. Arendt, A. Aschemann, B. Bayerlein, R. Courant, G. Dziwis, F. Fuchs, U. Giese, K. Junghanns, M. Kamal, L. Koschmieder, S. Leineweber, M. Luger, M. Lukas, J. Maas, J. Mertens, B. Mieller, L. Overmeyer, N. Pirch, J. Reimann, S. Schröck, P. Schulze, J. Schuster, A. Seidel, O. Shchyglo, M. Sierka, F. Silze, S. Stier, M. Tegeler, J. Unger, M. Weber, T. Hickel, J. Schaarschmidt, *Adv. Eng. Mater.* **2024**, 202402149.
[37] V. Gupta, W.-K. Liao, A. Choudhary, A. Agrawal, *MRS Commun.* **2023**, *13*, 754.
[38] I. Papadimitriou, I. Gialampoukidis, S. Vrochidis, I. Kompatsiaris, *Comput. Mater. Sci.* **2024**, *235*, 112793.
[39] B. Otto, M. ten Hompel, S. Wrobel, *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage, Computer Science*, 1st ed., Springer, Cham **2022**.
[40] T. Bray, The JavaScript Object Notation (JSON) Data Interchange Format, RFC 8259, https://www.rfc-editor.org/info/rfc8259 (accessed: February 2024).
[41] C. Bormann, P. E. Hoffman, Concise Binary Object Representation (CBOR), RFC 8949, https://www.rfc-editor.org/info/rfc8949 (accessed: February 2024).
[42] T. Bray, J. Paoli, C. M. Sperberg-McQueen, E. Maler, F. Yergeau, *Extensible Markup Language (XML) 1.0* (5th ed.), https://www.w3.org/TR/REC-xml/ (accessed: February 2024).
[43] GitHub - hdf5, https://github.com/HDFGroup/hdf5 (accessed: May 2024)
[44] M. T. Rose, D. E. Cass, ISO Transport Service on Top of the TCP Version: 3, RFC 1006, https://www.rfc-editor.org/info/rfc1006 (accessed: February 2024).
[45] Z. Shelby, K. Hartke, C. Bormann, The Constrained Application Protocol (CoAP), RFC 7252, https://www.rfc-editor.org/info/rfc7252 (accessed: February 2024).
[46] C. Amsüss, J. P. Mattsson, G. Selander, Constrained Application Protocol (CoAP): Echo, Request-Tag, and Token Processing, RFC 9175, https://www.rfc-editor.org/info/rfc9175 (accessed: February 2024).
[47] Wireguard - Fast, Modern, Secure VPN Tunnel, https://www.wireguard.com/ (accessed: May 2024).
[48] KEYCLOAK - Open Source Identity and Access Management, https://www.keycloak.org/ (accessed: May 2024).
[49] GitHub - PMD Deployment Guide, https://materialdigital.github.io/pmd-server/ (accessed: February 2024).
[50] Docker Compose, https://docs.docker.com/compose/ (accessed: May 2024)
[51] kubernetes - Production-Grade Container Orchestration, https://kubernetes.io/ (accessed: May 2024).

[52] S. R. Kalidindi, *Int. Mater. Rev.* **2015**, *60*, 150.
[53] C. Ward, J. Scott, L. C. Brinson, G. Galli, S. R. Kalidindi, A. Mehta, B. Meredig, J. Petters, B. Puchala, Z. Trautt, V. Venkatesh, *Met. Mater. Soc.* **2017**, 100, https://www.nsf.gov/attachments/297691/public/Building-a-Materials-Data-Infrastructure-Study.pdf.
[54] E. Curry, S. Scerri, T. Tuikka, *Data Spaces: Design, Deployment and Future Directions, Computer Science*, 1st ed., Springer, Cham **2022**.
[55] S. Harris, A. Seaborne, E. Prud'hommeaux, SPARQL 1.1 Query Language **2013**, https://www.w3.org/TR/sparql11-query/ (accessed: March 2024).
[56] RDF Working Group, Resource Description Framework (RDF) **2014**, https://www.w3.org/RDF/ (accessed: March 2024).
[57] Q. Duan, X. Wang, N. Song, *Learned Publ.* **2022**, *35*, 7.
[58] L. M. Ghiringhelli, C. Baldauf, T. Bereau, S. Brockhauser, C. Carbogno, J. Chamanara, S. Cozzini, S. Curtarolo, C. Draxl, S. Dwaraknath, A. Fekete, J. Kermode, C. T. Koch, M. Kühbach, A. N. Ladines, P. Lambrix, M.-O. Himmer, S. V. Levchenko, M. Oliveira, A. Michalchuk, R. E. Miller, B. Onat, P. Pavone, G. Pizzi, B. Regler, G.-M. Rignanese, J. Schaarschmidt, M. Scheidgen, A. Schneidewind, T. Sheveleva, C. Su, D. Usvyat, O. Valsson, C. Wöll, M. Scheffler, *Sci. Data* **2023**, *10*, 626.
[59] S. G. Higgins, A. A. Nogiwa-Valdez, M. M. Stevens, *Nat. Protoc.* **2022**, *17*, 179.
[60] J. A. Taillon, T. F. Bina, R. Plante, M. W. Newrock, G. R. Greene, J. W. Lau, *Microsc. Microanal.* **2021**, *27*, 511.
[61] T. Kluyver, B. Ragan-Kelley, F. Pérez, B. Granger, M. Bussonnier, J. Frederic, K. Kelley, J. Hamrick, J. Grout, S. Corlay, P. Ivanov, D. Avila, S. Abdalla, C. Willing, *Jupyter Notebooks - A Publishing Format for Reproducible Computational Workflows*, IOS Press, Amsterdam **2016**.
[62] Plattform MaterialDigital, Show & Tell Series on YouTube channel of Plattform MaterialDigital, https://www.youtube.com/watch?v=lrVbGGq9ZEQlist=UULFAwf5QXQ6Oa4NPaL3bXFvAA (accessed: January 2024)
[63] GitHub - PMD Demonstrator, https://github.com/materialdigital/PMD-pyiron_SPARQL_demonstrator (accessed: February 2024).
[64] J. A. Donenfeld, in *Network and Distributed System Security Symp.*, San Diego, CA, February **2017**, pp. 1–12.
[65] J. Janssen, S. Surendralal, Y. Lysogorskiy, M. Todorova, T. Hickel, R. Drautz, J. Neugebauer, *Comput. Mater. Sci.* **2019**, *163*, 24.
[66] C. R. C. Rêgo, J. Schaarschmidt, T. Schlöder, M. Penaloza-Amion, S. Bag, T. Neumann, T. Strunk, W. Wenzel, *Front. Mater.* **2022**, *9*, 877597.
[67] M. Schilling, S. Glaubitz, K. Matzak, B. Rehmer, B. Skrotzki, *Full Dataset of Several Mechanical Tests on an S355 Steel Sheet as Reference Data for Digital Representations*, Zenodo **2022**.
[68] B. Bayerlein, M. Schilling, P. v. Hartrott, J. Waitelonis, GitHub - Orowan Demonstrator **2023**, https://github.com/materialdigital/demo-orowan (accessed: March 2024).
[69] P. v. Hartrott, B. Skrotzki, *Room Temperature and Elevated Temperature Tensile Test and Elastic Properties Data of Al-Alloy EN AW-2618A After Different Aging Times and Temperatures*, Zenodo **2023**.
[70] C. Rockenhäuser, B. Skrotzki, *Radii of S-Phase Al2CuMg In Al-Alloy EN AW-2618A After Different Aging Times At 190 °C*, Zenodo **2023**.
[71] T. O'Brien, eLabFTW **2012**, https://www.elabftw.net/ (accessed: January 2024).
[72] M. Schilling, S. Bruns, B. Bayerlein, J. Kryeziu, J. Schaarschmidt, J. Waitelonis, *Adv. Eng. Mater.* **2024**, 2401527.
[73] GitHub - Data Acquisition Pipeline, https://git.material-digital.de/apps/dap (accessed: May 2024).
[74] GitHub - PMD Deployment Guide OntoDocker, https://materialdigital.github.io/pmd-server/pages/services/onto-docker/ (accessed: February 2024).
[75] S. Lohmann, V. Link, E. Marbach, S. Negru, in *Knowledge Engineering and Knowledge Management, EKAW 2014, Lecture Notes in Computer Science*, Vol. *8982* (Eds: P. Lambrix, E. Hyvönen, E. Blomqvist, V. Presutti, G. Qi, U. Sattler, Y. Ding, C. Ghidini), Springer, Cham **2015**, pp. 154–158.
[76] GitHub - facete3, https://github.com/Scaseco/facete3 (accessed: May 2024).
[77] B. Bayerlein, H. Birkholz, S. Fliegener, J. Grundmann, T. Hanke, P. v. Hartrott, M. Schilling, J. Waitelonis, GitHub - PMD Core Ontology (PMDco), https://github.com/materialdigital/core-ontology (accessed: February 2024).
[78] International Organisation for Standardization, Information technology - Top-level ontologies (TLO) - Part 2: Basic Formal Ontology (BFO), Standard, International Organisation for Standardization https://www.beuth.de/de/norm/iso-iec-21838-2/348948268 (accessed: February 2024).
[79] Data Catalog Vocabulary (DCAT) - Version 3, https://www.w3.org/TR/vocab-dcat-3/#dcat-scope (accessed: February 2024).
[80] SKOS Core Guide, https://www.w3.org/TR/swbp-skos-core-guide/ (accessed: February 2024).
[81] Dublin Core Metadata Initiative (DCMI)- Metadata Terms, https://www.dublincore.org/specifications/dublin-core/dcmi-terms/ (accessed: February 2024).
[82] GitHub - MaterialDigital1 Ontology Collection, https://github.com/materialdigital/materialdigital1_ontology_collection (accessed: February 2024).
[83] PROV-O: The PROV Ontology, https://www.w3.org/TR/prov-o/ (accessed: January 2024).
[84] The OPMW-PROV Ontology, https://www.opmw.org/model/OPMW/ (accessed: February 2024).
[85] GitHub - MaterialDigital1 Ontology Collection Evaluation, https://github.com/materialdigital/materialdigital1_ontology_collection/blob/analysis/eval.ipynb (accessed: April 2024).
[86] Results of MaterialDigital1 Ontology Collection Evaluation, https://materialdigital.github.io/materialdigital1_ontology_collection/index.html (accessed: October 2024).
[87] GitHub - EMMO, https://emmo-repo.github.io/ (accessed: February 2024).
[88] International Organization for Standardization, Metallic Materials - Tensile Testing - Part 1: Method of Test at Room Temperature, Standard, International Organisation for Standardization, https://www.beuth.de/en/standard/din-en-iso-6892-1/317931281 (accessed: March 2022).
[89] M. Schilling, B. Bayerlein, P. v. Hartrott, J. Waitelonis, H. Birkholz, P. D. Portella, B. Skrotzki, *Adv. Eng. Mater.* **2024**, 2400138.
[90] M. Schilling, B. Bayerlein, H. Birkholz, P. von Hartrott, J. Waitelonis, GitHub - Tensile Test Ontology (TTO) **2023**, https://github.com/materialdigital/application-ontologies/tree/main/tensile_test_ontology_TTO/ (accessed: January 2024).
[91] B. Bayerlein, M. Schilling, GitHub - Precipitate Geometry Ontology (PGO) **2023**, https://github.com/materialdigital/application-ontologies/tree/main/precipitate_geometry_ontology_PGO (accessed: February 2024).
[92] B. Bayerlein, M. Schilling, M. Curran, C. E. Campbell, A. A. Dima, H. Birkholz, J. W. Lau, *Integr. Mater. Manuf. Innovation* **2024**, *13*, 915.
[93] B. Bayerlein, M. Schilling, J. Lau, GitHub - Microscopy Ontology (MO) **2023**, https://github.com/materialdigital/application-ontologies/tree/main/microscopy_ontology_MO (accessed: February 2024).
[94] M. Mutz, M. Perovic, P. Gümbel, V. Steinbauer, A. Taranovskyy, Y. Li, L. Beran, T. Käfer, K. Dröder, V. Knoblauch, A. Kwade, V. Presser, D. Werth, T. Kraus, *Energy Technol.* **2023**, *11*, 2200681.
[95] V. Nebel, M. Mutz, Y. Heim, D. Werth, in *Proc. of Int. Conf. on Information Technology and Applications* (Eds: A. Ullah, S. Anwar,

D. Calandra, R. Di Fuccio), Springer Nature Singapore, Singapore **2024**, pp. 45–54.
[96] GitHub - DIGIT RUBBER, https://github.com/TIBHannover/DIGIT-RUBBER?tab=readme-ov-file (accessed: January 2024).
[97] GitHub - SciGlass, https://github.com/epam/SciGlass (accessed: April 2024).
[98] SciGlass Next, https://sciglass.uni-jena.de/ (accessed: April 2024).
[99] J. Friis, S. Gouttebroze, *Emmo-repo/domain-microstructure: V0.1.0*, GitHub **2023**.
[100] P. Del Nostro, G. Goldbeck, D. Toti, *Appl. Ontol.* **2022**, *17*, 401.
[101] GitHub - Mechanical Testing Ontology, https://gitlab.com/kupferdigital/ontologies/mechanical-testing-ontology (accessed: January 2024).
[102] H. B. Nasrabadi, T. Hanke, M. Weber, M. Eisenbart, F. Bauer, R. Meissner, G. Dziwis, L. Tikana, Y. Chen, B. Skrotzki, *Comput. Ind.* **2023**, *153*, 104016.
[103] GitHub - Copper Key Ontology, https://gitlab.com/kupferdigital/ontologies/copper-key (accessed: January 2024).
[104] Deutsches Kupferinstitut, DKI Copper Key, https://standalone.kupferschluessel.de/standalone.php?lang=english (accessed: January 2024).
[105] B. Meng, F. Dehn, J. F. Unger, D. A. Shepherd, E. Tamsen, S. Pirskawetz, in *21st Ibausil Int. Conf. on Building Materials*, Vol. *6* (Eds: H.-M. Ludwig, M. Lieboldt, T. Baron), Ernst & Sohn GmbH & CO. KG., Weimar, Germany, September **2023**, pp. 1505–1515, https://onlinelibrary.wiley.com/doi/full/10.1002/cepa.2955.
[106] GitHub - CommonCoreOntologies, https://github.com/CommonCoreOntology/CommonCoreOntologies (accessed: March 2024).
[107] T. Bjarsch, H. Beinersdorf, K. Drechsler, J. Reimann, *Ontology-Based Material Data Management In Additive Manufacturing Domain*, FEMS EUROMAT23, Frankfurt am Main **2023**.
[108] G. Xiao, D. Calvanese, R. Kontchakov, D. Lembo, A. Poggi, R. Rosati, M. Zakharyaschev, in *Proc. of the 27th Int. Joint Conf. on Artificial Intelligence, IJCAI-2018. Int. Joint Conf. on Artificial Intelligence Organization*, Stockholm, Sweden, July **2018**, pp. 5511–5519, https://doi.org/10.24963/ijcai.2018/777.
[109] Steel Processing Ontology, https://w3id.org/steel/ProcessOntology/ (accessed: April 2024).
[110] Miro Concept Boards, https://miro.com/de/ (accessed: October 2024).
[111] Conceptboard, https://conceptboard.com/ (accessed: October 2024).
[112] Y. Chen, M. Schilling, P. v. Hartrott, H. B. Nasrabadi, B. Skrotzki, J. Olbricht, *Integr. Mater. Manuf. Innovation* **2022**, *11*, 545.
[113] H. Knublauch, R. W. Fergerson, N. F. Noy, M. A. Musen, in *The Semantic Web - ISWC 2004: Third Int. Semantic Web Conf., Hiroshima, Japan, Proc., Volume 3298 of Lecture Notes in Computer Science* (Eds: S. A. McIlraith, D. Plexousakis, F. van Harmelen), Springer, Berlin, Heidelberg, Germany **2004**, pp. 229–243, https://doi.org/10.1007/978-3-540-30475-3_17.
[114] Protégé, https://protege.stanford.edu/ (accessed: October 2024)
[115] M. Blum, B. Ell, P. Cimiano, in *Proc. of the 14th Workshop on Ontology Design and Patterns (WOP 2023)*, Athen, Greece, November **2023**.
[116] M. G. Skjæveland, D. P. Lupp, L. H. Karlsen, H. Forssell, in *The Semantic Web-ISWC 2018: 17th Int. Semantic Web Conf., Proc., Part I 17*, Springer, Monterey, CA, October **2018**, pp. 477–494.
[117] G. Dziwis, L. Wenige, L. Meyer, M. Martin, in *Proc. of the 1st Int. Workshop on Semantic Industrial Information Modelling (SemIIM 2022) Co-located with the 19th Extended Semantic Web Conf. ESWC 2022, Volume 3355 of CEUR Workshop Proc.* (Eds: A. Waaler, E. Kharlamov, B. Zhou, D. Zhou), CEUR-WS.org, Greece, Crete, May **2022**, pp. 1–15, https://ceur-ws.org/Vol-3355/ontoflow.pdf.
[118] GitHub - Concrete Ontology, https://github.com/BAMresearch/LebeDigital/tree/main/lebedigital/ConcreteOntology (accessed: February 2024).
[119] GitHub - Stahldigital Ontology, https://github.com/materialdigital/stahldigital-ontology-public/tree/main/docs (accessed: April 2024).
[120] A. R. Akbashev, S. V. Kalinin, *Nat. Mater.* **2023**, *22*, 270.
[121] GitHub - MaterioMiner, https://gitlab.cc-asp.fraunhofer.de/iwm-micro-mechanics-public/datasets/materio-miner (accessed: April 2024).
[122] GitHub - Materials Mechanics Ontology, https://gitlab.cc-asp.fraunhofer.de/iwm-micro-mechanics-public/ontologies/materials-mechanics-ontology (accessed: April 2024).
[123] H. Babaei Giglou, J. D'Souza, S. Auer, in *The Semantic Web – ISWC 2023: 22nd Int. Semantic Web Conf., Athens, Greece, November 6–10, 2023, Proc., Part I*, Springer-Verlag, Berlin, Heidelberg, November **2023**, pp. 408–427, ISBN 978-3-031-47239-8, https://doi.org/10.1007/978-3-031-47240-4_22.
[124] Y. Liu, B. L. Elsworth, T. R. Gaunt, *Bioinformatics* **2023**, *39*, btad169.
[125] C. Saetia, J. Phruetthiset, T. Chalothorn, M. Lertsutthiwong, S. Taerungruang, P. Buabthong, in *Proc. of TextGraphs-17: Graph-based Methods for Natural Language Processing* (Eds: D. Ustalov, Y. Gao, A. Panchenko, E. Tutubalina, I. Nikishina, A. Ramesh, A. Sakhovskiy, R. Usbeck, G. Penn, M. Valentino), Association for Computational Linguistics, Bangkok, Thailand, August **2024**, pp. 53–60, https://aclanthology.org/2024.textgraphs-1.4
[126] R. N. Murthy, S. T. Potu, A. Thomas, L. Mishra, N. Prange, A. R. Durmaz (preprint), ChemRxiv 2024.
[127] T. Xie, Y. Wan, W. Huang, Z. Yin, Y. Liu, S. Wang, Q. Linghu, C. Kit, C. Grazian, W. Zhang, I. Razzak, B. Hoex, DARWIN Series: Domain Specific Large Language Models for Natural Science, https://arxiv.org/abs/2308.13565 (accessed: April 2023).
[128] A. M. Bran, S. Cox, O. Schilter, C. Baldassari, A. D. White, P. Schwaller, *Nat. Mach. Intell.* **2024**, *6*, 525.
[129] GitHub - PMD Knowledge Graph, https://materialdigital.github.io/pmdkg/ (accessed: March 2024).
[130] Virtuoso Open-Source Edition, https://vos.openlinksw.com/owiki/wiki/VOS (accessed: March 2024).
[131] LodLive, http://en.lodlive.it/ (accessed: March 2024).
[132] Triply, https://yasgui.triply.cc/ (accessed: March 2024).
[133] GitHub - SHMARQL, https://github.com/epoz/shmarql (accessed: March 2024).
[134] MSE Knowledge Graph v1.0, https://demo.fiz-karlsruhe.de/matwerk/ (accessed: March 2024).
[135] A. Gangemi, V. Presutti, *Handbook on Ontologies, Chapter Ontology Design Patterns*, 2nd ed., Springer, Berlin, Heidelberg, Germany **2009**, pp. 221–243, ISBN 978-3-540-92673-3.
[136] R. A. Falbo, G. Guizzardi, A. Gangemi, V. Presutti, in *Proc. of the 4th Int. Conf. on Ontology and Semantic Web Patterns - Volume 1188, WOP'13. CEUR-WS.org, Aachen, DEU*, Sydney, Australia, October **2013**, pp. 14–26.
[137] OBO Foundry, https://obofoundry.org/ (accessed: March 2024).
[138] L. Vogt, P. Strömert, N. Matentzoglu, N. Karam, M. Konrad, M. Prinz, R. Baum, FAIR 2.0: Extending the FAIR Guiding Principles to Address Semantic Interoperability https://arxiv.org/abs/2405.03345 (accessed: September 2024).
[139] D. Garijo, M. Poveda-Villalón, *Applications and Practices in Ontology Design, Extraction, and Reasoning, Volume 49, Chapter Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web*, IOS Press, Amsterdam, Netherlands **2020**, pp. 39–54.
[140] M. Poveda-Villalón, D. Garijo, A. N. Gonzalez-Beltran, C. Jonquet, Y. Le Franc, in *FAIR Principles for Ontologies and Metadata in Knowledge Management, FOAM2024*, Enschede, Netherlands, July **2024**, pp. 1–14.

[141] Fast Ontodocker, https://git.material-digital.de/ya-fanchen/fast-ontodocker (accessed: January 2024)
[142] S. Kempf, M. Krug, F. Puppe, *Appl. Intell.* **2022**, *53*, 9513.
[143] B. Mieller, S. Ben Hassine, J. Töpfer, C. Priese, A. Bochmann, B. Capraro, S. Stark, U. Partsch, C. Fresemann, *Adv. Eng. Mater.* **2024**, 2401042.
[144] J.-B. Lamy, *Artif. Intell. Med.* **2017**, *80*, 11.
[145] H. R. Schmidtke, M. Leemhuis, J. Mertens, R. Courant, J. Maas, Ö. L. Özçep, in *2023 19th Int. Conf. on Intelligent Environments (IE)*, Uniciti, Mauritius, June **2023**, pp. 1–8, https://ieeexplore.ieee.org/document/10179095 (accessed: February 2024).
[146] J. Mertens, M. Leemhuis, Ö. Özçep, H. Schmidtke, J. Maas, in *Electroactive Polymer Actuators and Devices (EAPAD) XXV* (Eds: H. R. Shea, I. A. Anderson, J. D. W. Madden), Vol. *12482*, International Society for Optics and Photonics, SPIE **2023**, p. 124820Z, https://doi.org/10.1117/12.2661222.
[147] J. Maas, M. Leemhuis, J. Mertens, H. Schmidtke, R. Courant, M. Dahlmann, S. Stark, A. Böhm, K. Pagel, M. Hinze, D. Pinkal, M. Wegener, M. F.-X. Wagner, T. Sattel, H. Neubert, Ö. Özçep, *Adv. Eng. Mater.* **2024**, *2302208*, https://doi.org/10.1002/adem.202302208.
[148] GitHub - SmaDi-OBDMA-System, https://github.com/SmaDi-OBDMA/SmaDi-OBDMA-system/tree/main (accessed: January 2024).
[149] NFDI - Nationale Forschungsdaten Infrastruktur, https://www.nfdi.de/ (accessed: March 2024).
[150] FAIRmat, https://www.fairmat-nfdi.eu/fairmat (accessed: March 2024)
[151] NOMAD Materials Science Data Managed and Shared, https://nomad-lab.eu/nomad-lab/ (accessed: March 2024).
[152] FAIR-DI - FAIR Data Infrastructure for Physics, Chemistry, Materials Science, and Astronomy e.V., https://www.fair-di.eu/fair-di/ (accessed: March 2024).
[153] NFDI-MatWerk - Nationale Forschungsdaten Infrastruktur für Materialwissenschaft & Werkstofftechnik, https://www.nfdi-matwerk.de (accessed: March 2024)
[154] C. Eberl, M. Niebel, E. Bitzek, T. Dahmen, F. Fritzen, P. Gumbsch, T. Hickel, S. Klein, F. Mücklich, M. S. Müller, J. Neugebauer, C. Pauly, M. Politze, H. Sack, S. Sandfeld, R. Schwaiger, P. Slusallek, R. Stotzka, A. Streit, M. Zimmermann, *Consortium Proposal NFDI-MatWerk*, Zenodo **2021**, https://doi.org/10.5281/zenodo.5082837.
[155] EOSC - European Open Science Cloud, https://digital-strategy.ec.europa.eu/de/policies/open-science-cloud (accessed: March 2024)
[156] Gaia-X, https://www.bmwk.de/Redaktion/EN/Dossier/gaia-x.html (accessed: March 2024).
[157] Catena-X, https://catena-x.net/en/ (accessed: April 2024).
[158] Manufacturing-X, https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/Manufacturing-X.html (accessed: April 2024).
[159] Plattform Industrie 4.0, https://www.plattform-i40.de/IP/Navigation/EN/Home/home.html (accessed: April 2024).
[160] FAIR. Data Spaces, https://www.nfdi.de/fair-data-spaces/?lang=en (accessed: March 2024).
[161] IDSA - International Data Spaces Association, https://internationaldataspaces.org/ (accessed: March 2024).
[162] DIADEM, https://www.cnrs.fr/en/press/diadem-exploratory-priority-research-programme-and-infrastructure-linking-materials-and-ai (accessed: March 2024).
[163] MatNavi - NIMS Material Database, https://mits.nims.go.jp/ (accessed: March 2024).
[164] MGI - Materials Genome Initiative, https://www.mgi.gov/ (accessed: March 2024).
[165] A. White, *MRS Bull.* **2012**, *37*, 715.
[166] RDA - Research Data Alliance, https://www.rd-alliance.org/ (accessed: March 2024).
[167] RDA/CODATA - Materials Data, Infrastructure & Interoperability IG, https://www.rd-alliance.org/node/939 (accessed: March 2024).
[168] Harmonised Terminologies and Schemas for FAIR Data in Materials Science and Related Domains WG, https://www.rd-alliance.org/groups/harmonised-terminologies-and-schemas-fair-data-materials-science-and-related-domains-wg (accessed: March 2024).
[169] MaRDA - Materials Research Data Alliance, https://www.marda-alliance.org/ (accessed: March 2024).
[170] IOF - Industrial Ontologies Foundry, https://oagi.org/pages/industrial-ontologies (accessed: March 2024).
[171] EMMC - European Materials Modelling Council, https://emmc.eu/ (accessed: April 2024).
[172] OntoCommons, https://ontocommons.eu/ (accessed: April 2024).
[173] TIB Terminology Service, https://terminology.tib.eu/ts/ (accessed: March 2024).
[174] MatPortal.org, https://matportal.org/ontologies (accessed: March 2024).
[175] FEMS - Federation of European Materials Societies, https://www.fems.org/event/fems-euromat-2023-1 (accessed: April 2024).
[176] OMERO, https://www.openmicroscopy.org/omero/ (accessed: May 2024).
[177] D. Beckett, T. Berners-Lee, E. Prud'hommeaux, G. Carothers, RDF 1.2 Turtle - Terse RDF Triple Language, https://www.w3.org/TR/2024/WD-rdf12-turtle-20240418/ (accessed: March 2024).
[178] M. Sporny, D. Longley, G. Kellogg, M. Lanthaler, P.-A. Champin, N. Lindström, JSON-LD 1.1 - A JSON-BASED SERIALIZATION for Linked Data, https://www.w3.org/TR/2020/REC-json-ld11-20200716/ (accessed: 2020).
[179] DigiBatMat, https://www.tu-braunschweig.de/iwf/fup/abp/forschungsprojekte-1/digibatmat (accessed: January 2024).
[180] DigiBatMat Project, https://www.materialdigital.de/project/14 (accessed: January 2024).
[181] DIGIT RUBBER, https://www.dikautschuk.de/forschung/aktuelle-projekte/projektil-5104-1-2-1/ (accessed: January 2024).
[182] DIGIT RUBBER Project, https://www.materialdigital.de/project/12 (accessed: January 2024).
[183] DiProMag, https://www.dipromag.de/ (accessed: January 2024).
[184] DiProMag Project, https://www.materialdigital.de/project/2 (accessed: January 2024).
[185] DiStAl Project, https://www.materialdigital.de/project/7 (accessed: January 2024).
[186] GlasDigital, https://www.bam.de/Content/DE/Projekte/laufend/GlasDigital/glasdigital.html (accessed: January 2024).
[187] GlasDigital Project, https://www.materialdigital.de/project/4 (accessed: January 2024).
[188] iBain, https://www.iwm.fraunhofer.de/de/warum-fraunhofer-iwm/loesungen-fuer-produktlebenzyklus/digitalisierung-in-der-werkstoff technik/ibain.html (accessed: January 2024).
[189] iBain Project, https://www.materialdigital.de/project/8 (accessed: January 2024).
[190] KNOW-NOW, https://www.bam.de/Content/DE/Projekte/laufend/KNOW-NOW/know-now.html (accessed: January 2024).
[191] KNOW-NOW Project, https://www.materialdigital.de/project/9 (accessed: January 2024).
[192] KupferDigital, https://fem-online.de/kupferdigital-datenoekosystem-fuer-die-digitale-materialforschung-auf-basis-ontologie-basierter-digitaler-repraesentationen-von-kupfer-und-kupferlegierungen-2/ (accessed: January 2024).
[193] KupferDigital Project, https://www.materialdigital.de/project/1 (accessed: January 2024).

[194] LeBeDigital, https://www.bam.de/Content/DE/Projekte/laufend/LeBeDigital/lebedigital.html (accessed: January 2024).
[195] LeBeDigital Project, https://www.materialdigital.de/project/3 (accessed: January 2024).
[196] ODE_AM, https://www.tuhh.de/idac/research/ode-am (accessed: January 2024).
[197] ODE_AM Project, https://www.materialdigital.de/project/13 (accessed: January 2024).
[198] SensoTwin, https://www.asg.ed.tum.de/lcc/oeffentliche-projekte/simulation/sensotwin-sensorintegrierter-digital-twin-fuer-hochleistungs-faserverbundanwendungen/ (accessed: January 2024).
[199] SensoTwin Project, https://www.materialdigital.de/project/5 (accessed: January 2024).
[200] SmaDi, https://www.tu.berlin/emk/forschung/projekte/smadi (accessed: January 2024).
[201] SmaDi Project, https://www.materialdigital.de/project/10 (accessed: January 2024).
[202] StahlDigital, https://www.iwm.fraunhofer.de/de/warum-fraunhofer-iwm/loesungen-fuer-produktlebenzyklus/digitalisierung-in-der-werkstofftechnik/stahldigital.html (accessed: January 2024).
[203] StahlDigital Project, https://www.materialdigital.de/project/6 (accessed: January 2024).