

## Article

# Digitalizing Material Knowledge: A Practical Framework for Ontology-Driven Knowledge Graphs in Process Chains

Elena Garcia Trelles <sup>\*</sup>, Christoph Schweizer <sup>†</sup>, Akhil Thomas <sup>†</sup>, Philipp von Hartrott <sup>†</sup> and Marina Janka-Ramm <sup>†</sup>

Fraunhofer-Institut für Werkstoffmechanik IWM, Wöhlerstr. 11, 79108 Freiburg, Germany; elena.garcia.trelles@iwm.fraunhofer.de (E.G.T.); akhil.thomas@iwm.fraunhofer.de (A.T.); philipp.von.hartrott@iwm.fraunhofer.de (P.v.H.); marina.janka-ramm@iwm.fraunhofer.de (M.J.-R.)

<sup>\*</sup> Correspondence: elena.garcia.trelles@iwm.fraunhofer.de

<sup>†</sup> These authors contributed equally to this work.

**Abstract:** This paper proposes a robust methodology for integrating process-specific data and domain expert knowledge into linked knowledge graphs. These graphs utilize an ontology that provides a standardized vocabulary for material science and facilitates the creation of semantic models for various processes along the digital process chain. A generic template for structuring processes is proposed, simplifying subsequent data retrieval. The templates of specific processes are designed collaboratively by domain and ontology experts, aided by a proposed interview template that bridges the knowledge gap. Following the digitalization of material data through semantic modeling, machine-readable data with contextual metadata is stored in a graph database, which can be efficiently queried using the SPARQL language, enabling seamless integration into data pipelines. To demonstrate this approach, a knowledge graph is developed to represent the process chain of AlSi10Mg objects manufactured via permanent mold casting, capturing their complete history from the initial manufacturing step to final non-destructive testing and mechanical characterization. This methodology enhances data interoperability and accessibility while providing context-rich data for training AI models, potentially accelerating new knowledge discovery in material science.



**Citation:** Garcia Trelles, E.; Schweizer, C.; Thomas, A.; von Hartrott, P.; Janka-Ramm, M. Digitalizing Material Knowledge: A Practical Framework for Ontology-Driven Knowledge Graphs in Process Chains. *Appl. Sci.* **2024**, *14*, 11683. <https://doi.org/10.3390/app142411683>

Academic Editor: Douglas O'Shaughnessy

Received: 12 November 2024  
Revised: 3 December 2024  
Accepted: 6 December 2024  
Published: 14 December 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** cast AlSi10Mg; digital process chain; materials ontology; knowledge graph template; knowledge extraction; digitalization of material knowledge

## 1. Introduction

In materials science and engineering (MSE), an interdisciplinary field that focuses on the properties, performance, and applications of materials [1], establishing a well-conceived, reliable, and consistent methodology for producing FAIR (Findable, Accessible, Interoperable, Reusable) material data [2] is essential for accelerating material discovery. Adhering to FAIR principles enhances reproducibility and reduces data ambiguity, which is particularly important in a domain such as MSE, characterized by data scarcity and diversity. The World Wide Web Consortium (W3C) [3] has been instrumental in developing and promoting Semantic Web standards, which include RDF (Resource Description Framework), OWL (Web Ontology Language), and SPARQL (a query language for RDF data) [4]. These standards offer a framework for meaningfully sharing and connecting data across the web, enabling interoperability among various applications and domains. Adopting these standards to transform materials data into knowledge graphs represents a promising approach to generating FAIR data in MSE, facilitating the direct linkage of data produced at various stages of materials processing chains. Furthermore, with the advent of powerful large language models (LLMs) [5], knowledge graphs may serve as a countermeasure to address the performance limitations identified within the MSE domain [6]. For instance, strategies that ground the outputs of LLMs in the factual content of knowledge graphs could significantly enhance their reliability [7].

The present work is motivated by the need for a clear and systematic digitalization workflow that generates high-quality, semantically structured linked material data. To this end, a user-friendly approach for digitalizing material data is introduced, aiming to ensure that the proposed solution can be implemented not just by IT experts but also materials engineers and scientists from diverse backgrounds.

In prescribing a digitalization strategy in MSE, various data sources must be considered, including structured and unstructured data collections and actively generated data in research and production. The examination of existing datasets, such as the Inorganic Crystal Structure Database (ICSD) [8], the Open Quantum Materials Database (OQMD) [9], the Materials Project [10], the Materials Data Facility (MDF) [11], and the Novel Materials Discovery (NOMAD) [12], reveals significant deficiencies. Many of these databases contain incomplete and ambiguous data points and lack interconnectivity. This issue is exacerbated by data silos, where information is stored in isolated systems, making access and integration challenging. The diverse sources and formats of data, often in natural language, render much of it untraceable and unstructured, limiting the discoverability, retrievability, and reusability of data for both research and industry. These issues are well highlighted in the work of Bayerlein et al. [13]. Additionally, the absence of standardized data formats and protocols presents a critical challenge [14]. The diversity of research areas within the materials science community complicates seamless integration and analysis, resulting in fragmented datasets that are difficult to compare or combine. For instance, integrating experimental and computational data remains complex, as the field generates vast amounts of data from various synthesis and characterization experiments. The Materials Experiment and Analysis Database (MEAD) [15] has been developed to manage millions of materials synthesis and characterization experiments; however, the sheer volume and diversity of the data pose substantial challenges for effective integration and analysis.

The challenges posed by data silos, lack of standardization, and difficulties in integrating disparate datasets significantly slow down progress in materials science. These issues hinder the ability to connect material data with contextual information and complicate the analysis of large amounts of information. Addressing these challenges is essential for advancing the field and unlocking new opportunities for discovery and innovation. Semantic web technologies provide a promising solution, ensuring material data is structured, interconnected, and interoperable. In the stack of W3C-endorsed technologies, ontologies are fundamental in adding meaning to data, ensuring it is unambiguous and interoperable [16–18]. They serve as ‘machine-understandable’ representations of a domain, where knowledge is expressed as axioms that can be used to infer new insights from existing data, thereby improving data querying and analysis.

Numerous ontology development efforts have been undertaken to model different subdomains as well as the general MSE domain. Most of these ontologies either do not adhere to a top-level ontology or utilize BFO [19] or EMMO [20] as the top-level. In the general MSE domain, ontologies like PMD-Core [21], MSEO [22], and EMMO are currently popular. Additionally, ontologies have been developed for various subdomains and applications in MSE, such as mechanical testing [23–25] and manufacturing [26,27]. Norouzi et al. [28] highlight that despite numerous ontology design efforts in MSE, there is still a need to adopt ontology design patterns based on explicit competency question formulations in the MSE domain.

This work presents a mid-level ontology that captures generic concepts in the MSE field. It also includes a domain-specific ontology focused on the permanent mold casting process for AlSi10Mg, along with the associated manufacturing and material characterization processes. The digitalization strategy has been intentionally designed to be ontology-agnostic, facilitating collaboration with other ontology development efforts in MSE.

Despite numerous initiatives to develop ontologies in MSE, a lack of focus on practical methods for data annotation and user-friendly workflows that can assist materials scientists in utilizing Semantic Web technologies to produce FAIR data remains. While some related work has been discussed, including a well-documented example of a web-based

data infrastructure for generating FAIR tribological data [29], it is clear that a formalized knowledge extraction process is needed to improve communication between ontology developers and materials science experts. Bridging the gap between these disciplines is essential for increasing the acceptance of ontologies in materials science.

Additionally, it is paramount to showcase the feasibility and applicability of digitalization methods using real-world examples. This is necessary not only to validate and enhance existing ontologies and knowledge graph structures but also to showcase the knowledge gained from digitizing materials science data using Semantic Web technology. A compelling application of generating semantically structured materials data lies in providing context-rich and historically informed data to AI models, which can significantly accelerate material design and optimization processes.

The main objective of this work is to improve the efficiency and effectiveness of digital mapping of chemistry-process-structure-property relationships along industrially relevant process chains through a straightforward and systematic digitalization workflow that generates high-quality, semantically structured, linked materials data. A key focus is building digital process chains, which are essential for managing the product life cycle of any object and ensuring adequate quality control. A methodology for creating digital process chains is particularly critical in engineering, where existing literature has not adequately addressed this need. Consequently, significant efforts have been made to implement the necessary digitalization steps in a user-friendly manner, ultimately contributing to the advancement of digitalization in MSE data.

This investigation introduces a methodology to structure and uniformly represent material data along real-world process chains. The applicability of the method proposed is demonstrated through the AlSi10Mg case study, which is employed in Sections 2 and 3 to provide concrete examples of its utility. Section 2 introduces the underlying experimental dataset for AlSi10Mg, the BWMD Ontology and a flexible methodology for digitalizing MSE domain knowledge, supported by representative examples. An interview template is also included to facilitate interaction with domain experts. The resulting data structures and datasets are summarized in Section 3, presented through real-life use cases for knowledge extraction. A first step toward AI-assisted prediction of material properties is taken based on the generated knowledge base, demonstrating its machine readability. Finally, the results are discussed in Section 4 and concluded in Section 5.

## 2. Materials and Methods

First, the experimental dataset for the cast material AlSi10Mg, developed within the framework of the MaterialDigital project [30], is introduced, as well as the ontology providing the necessary vocabulary to describe the context of the mentioned material dataset. Second, the methodology for digitalizing process-generated data using Semantic Web technology, known as the 'Graph Designer Workflow', is presented. This methodology will be illustrated with examples from the AlSi10Mg dataset, highlighting the various steps involved in the proposed approach.

### 2.1. Description of the Experimental Database for the Cast AlSi10Mg

The AlSi10Mg Aluminium-based alloy is a versatile, lightweight alloy well known for its high strength, dynamic load-bearing capacity, and corrosion resistance. Furthermore, its ductility and conductivity properties can be enhanced via heat treatment. This alloy is commonly used in aerospace and automotive applications, as well as in lightweight design [31]. Within the framework of the BWMD project [30] and out of the 'Use Case Metals', a comprehensive database was generated for the cast AlSi10Mg. To this end, different manufacturing routes were followed, aiming to improve the mechanical properties of the cast AlSi10Mg. Various process parameters were applied in the permanent mold casting process, as well as variations of the amounts of Si and Mg in the melt. After the permanent mold casting process, the batches of material produced were heat treated in two steps: solution annealing and artificial aging. Small variations were also introduced

during these two steps by modifying the duration and the maximum temperature applied during the solution annealing and aging heat treatment steps. The resulting manufacturing routes led to several material batches with different mechanical properties, such as tensile strength, ductility, and hardness.

To investigate and mechanically characterize the different material batches, the heat-treated cast parts were finished as tensile test specimens and metallographic specimens. Those were used to perform tensile tests and Brinell hardness measurements. Additionally, component-like parts were cast, and their defect or porosity distribution was investigated via computer tomography.

The resulting experimental database requires a data architecture capable of seamlessly linking the data produced by the different processes of the manufacturing process chain to each of the manufactured or investigated objects. Therefore, a Semantic Web solution was sought to provide interoperable data and design a general process structure pattern to enable the construction and retrieval of the digital process chain of an object. The aim is to digitalize the different production routes with all the relevant contextual data from a different viewpoint, creating the possibility of storing process information about an object in terms of its chronological history.

## 2.2. The BWMD Ontology

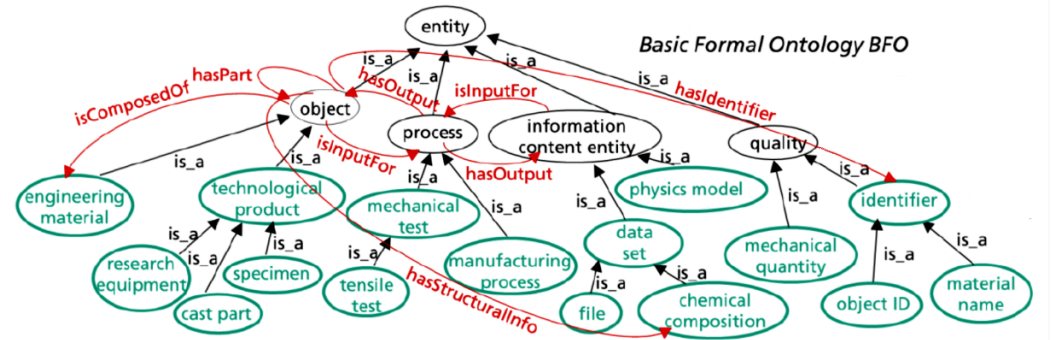
The first step in generating semantically structured data is establishing a common vocabulary to be formalized in the form of an ontology and a data model, both of which are to be determined by the use case.

An ontology consists of a collection of classes (hierarchically organized following the class-subclass relation), properties, and constraints. It aims to describe domain knowledge at a reasonable granularity for better reuse and understanding of the digitalized domain data. Hereafter, it is important to clarify that the classes of an ontology encapsulate a range of categories of objects or concepts that share similar attributes and behaviors. Meanwhile, the instances of a class represent a particular individual inside a given class.

Within the framework of the project MaterialDigital [30], a domain ontology managing class hierarchy, object properties, and data properties was designed and implemented. This domain ontology was the product of the united efforts and extensive communication between different MSE domain experts. The developed domain module was directly merged with the BFO 2.0 top-level ontology [32] and received the name of 'BWMD Ontology' [33,34]. The BWMD Ontology contains two layers: a mid-level ontology layer to model material-intensive process chains and a domain layer to model all process details related to the permanent mold casting process for AlSi10Mg, manufacturing, and characterization processes (mechanical and analytical). The process chains modeled by the mid-level ontology concepts (classes and properties) include the processed objects and corresponding process-related data (raw data and analyzed properties), the involved process parameters, the involved machines, the necessary research equipment, infrastructure, and software. The BWMD Ontology also enables a simple description of the material structure, which can be specified for the chemical composition, the material volume, and the material surface. The BWMD Ontology does not include axioms beyond subclassing, and domains and range restrictions for object and data properties.

Figure 1 schematically represents the connections between BWMD ontology classes and BFO classes. It exemplifies how the owl: class *engineering material* from the BWMD Ontology is modeled as a subclass of (*is\_a*) the BFO owl: class *object*. The BFO classes are shown in black and the classes specific to the BWMD Ontology are displayed in green. Details about how instances of these classes are related are expressed by the object properties, depicted in red. This straightforward representation illustrates how an *object* can interact with a process as an 'input' or 'output' instance. A *chemical composition* could be associated with a particular *object* through the object property *hasStructuralInfo*, and each *object* is associated with a specific object identifier through the object property *hasIdentifier*, which points to an instance of the class *identifier*, that could be more precisely represented

by one of its subclasses, e.g., *object ID* (to identify an object) or *material name* (to identify a material). On the other hand, the concept of *dataset* (e.g., a *chemical composition*) is modeled as a subclass of the IAO (Information Artifact Ontology [35]) owl: class *information content entity*, where an instance of *information content entity* can be directly connected to a process, both as input and as output instance.



**Figure 1.** The BWMD process instance modeling philosophy: BFO classes depicted in black, BWMD classes in green, and BWMD object properties in red.

### 2.3. The Graph Designer Workflow

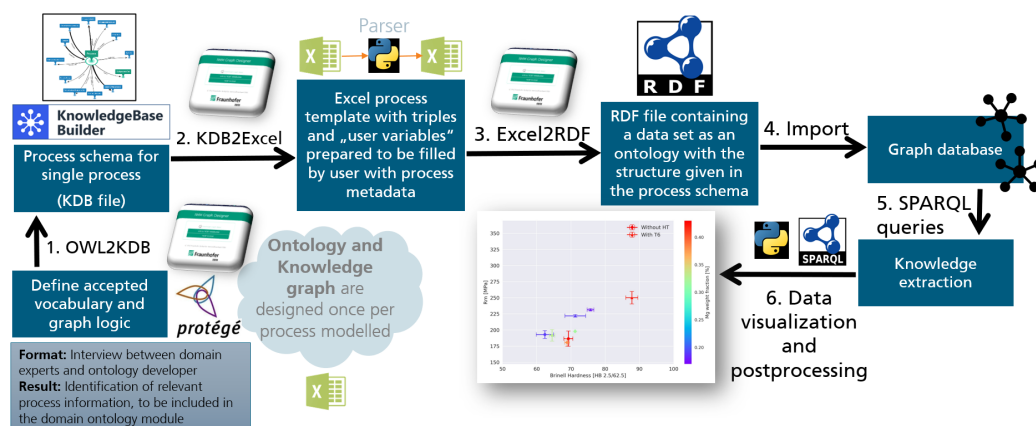
The Graph Designer Workflow is a generic digitalization workflow used to model the metadata of arbitrary processes as an interoperable manageable knowledge graph. It includes an ontology, a process data model or process graph template, a set of tools (the Graph Designer Tool), and a data pipeline. It is generic enough to work for different processes, ontologies, and data models, with the outcome being structured data presented as ontology-based knowledge graph(s).

The Graph Designer Workflow is also coupled with a data modeling philosophy, which relies on the concept of creating a data model (or graph) for each impartible process part of the process chain of a labeled product, attaching to each single process its corresponding relevant metadata. In this context, ‘process chain’ refers to the sequence of processes. An example of such an impartible process, including the input and output objects, is represented in Figure 2.



**Figure 2.** Schematic representation of the permanent mold casting process including the ingot as input and the cast part as output.

The introduced digital workflow for data structuring is illustrated in Figure 3 and comprehends the following steps:



**Figure 3.** The Graph Designer Workflow. Extended documentation regarding its application and tools can be found in [36].

1. Domain knowledge extraction: the first step when semantically modeling a new process is establishing the necessary vocabulary to describe the knowledge and data to be digitalized. This step is vital and will determine the quality and richness of the digitalized process data. The successful completion of the digitalization process is contingent on efficient communication between domain experts and ontology developers. An interview template was created to overcome the existent knowledge gap between these two involved parties. The interview template enables structured interviews and serves as an information exchange protocol. The output of this step results in the process-specific shared vocabulary being integrated into the domain ontology selected for the concrete use case. The formalization of the new extracted domain-specific vocabulary into an existing ontology (or into a new domain ontology module) is then conducted with the help of the software Protégé [37] or a customized Python script;
2. Creation of impartible process graph templates: at this point, a single process step is generically represented as a knowledge graph (from now on denoted as 'Process Graph'). For this purpose, the software InfoRapid KnowledgeBase Builder [38] is used, which offers a graphical user interface for generating graphs. The Process Graph template generation of an impartible process consists of modifying a generic Process Graph template based on the information gathered from the interview template filled in the previous step and consequent established domain-specific vocabulary (domain ontology). This step is accomplished with the help of the functionality OWL2KDB of the 'Graph Designer Tool' [36], which allows the user to import concepts from an existing ontology into the graphing tool. Creating a process-specific Process Graph template is an iterative process. It might take a few hours to days to semantically represent all relevant parts of the modeled process, depending on its level of description detail and complexity. The generated graph for describing the relevant metadata of the digitalized process is then converted into an Excel template. This substep is performed with the help of another functionality of the Graph Designer Tool, referred to as KDB2Excel;
3. Acquisition of process datasets: the previously generated Excel template is duplicated and filled in by the domain experts of the respective process step. In principle, the filling can also be script-controlled. In addition to the captured metadata, the raw data is collected and linked in the filled or annotated Excel template;
4. RDF graph data generation: once the Excel templates have been filled with all the available process metadata, those are converted into RDF data via the Graph Designer Tool functionality Excel2RDF. The generated graphs follow the structure defined in the specific Process Graph template and are compatible with the semantic structure of the generic Process Graph template, used as core pattern;

5. Linking of process datasets: this step comes in addition when aiming to build digital process chains out of the different semantic data generated by reproducing the Graph Designer Workflow for each digitalized process along the supply chain. In the last step of data structuring, the process steps are concatenated according to the sequence of the physical process chain.

At the end of the data structuring procedure, the entire knowledge graph of the process chain is available in a graph database (e.g., [39]) and can be imported and exported using the standardized RDF format for knowledge graphs. Each of the mentioned steps will be explained in detail in the upcoming sections, along with real-life examples extracted from the database introduced in Section 2.1.

### 2.3.1. Systematic Extraction of Domain Expert Knowledge

The interview template supports the knowledge extraction process, which guides the information exchange between the ontology developer and the domain expert. The work of the ontology developer—from outlining the process to developing the ontology—is thus facilitated, and errors can be minimized. A draft version of the interview template focused on the semantic description of an experimental process can be downloaded in [40]. The interview template is based on the different phases of an experimental process: the setup, the execution, and the outcome. This structure is intended to assist the domain expert in completing the template.

After a short introduction text for the domain expert in the first tab of the template, the process metadata, software, and equipment tabs follow as part of the setup phase. Here, among other things, the contact information of the process owner, but also general information about the process, used software, and used equipment are requested. This is followed by process input and process control parameters as part of the execution phase. These tabs give the ontology developer a lot of important information about the test object (e.g., if preparation steps were performed) as well as information about all parameters controlling the process. Figure 4 displays the table of process control parameters. The yellow hint box is intended to support the domain expert in selecting the control parameters. For each parameter, a definition in English (and for this case also in German), the data type (e.g., string, integer), and the unit (e.g., °C, %) are expected additionally. This list provides the ontology developer with all information about the process control parameters for direct integration into the domain ontology.

quantity	definition EN	definition DE	data type of value	unit

They can be seen as a necessary input in order to control the process, e.g.:

- temperature
- strain rate
- mechanical strain amplitude
- frequency
- period
- comment

**Figure 4.** Detail of the interview Excel sheet regulating the information exchange between the domain expert and the ontology developer: one section of the interview is reserved for collecting the list of parameters that determine a specific outcome of the process to be modeled, in the BWMD Ontology referred to through the class *ProcessControlParameters*.

In the last phase, the results are handled. The process output and time stamp tabs are provided in the Excel sheet for this purpose. The process output describes, among other things, any changes undergone by the test object and all data generated during the process. The time stamp determines the start and end time of the process. Two additional tabs:

input and output data description, allow the domain expert to insert screenshots of the respective data and add a short comment.

Formally, the template is structured so that yes/no questions are included in the first level. If there is a need for further questions for the corresponding answer, these only then unfold. This ensures a clear presentation.

### 2.3.2. Creating a Process Graph: The Semantic Data Model

The design of the semantic data model is done based on the generic Process Graph template, illustrated in Figure 5. Its core structure consists of 12 node types and 8 object properties directly connected to the Process node. If irrelevant to the design of the Process, nodes, and associated object properties can be deleted. During the Process template design, the category of the Process node in Figure 5 needs to be refined (e.g., by *PermanentMoldCasting*). It is important to note, that the referred template is built on, based on BWMD Ontology vocabulary. The nodes represent instances of BWMD classes connected to other instances through object properties from the BWMD Ontology. In the following, the meaning of the core nodes and object properties are explained:

- *isInputFor* and *hasOutput* are relations to describe the chronological order along the process chain and are attached to nodes, which are either the input for a process (*Object* and *DataSet*) or the output thereof (*Object* and *ProcessDataSet*). During process design, the categories of the ingoing and outgoing objects can be further specified/refined. The same holds for the ingoing *DataSet*. The node *ProcessDataSet* collects all data, that is either generated during the course of a process (e.g., raw data) or is the result thereof (e.g., analyzed material properties);
- The setpoints and controlled variables of the process are assigned to the node *ProcessParameterSet*;
- Machines, measurement equipment and consumables used within the process are allocated to the node *ProcessSetup*;
- The node *InfrastructureEquipmentSet* merges the necessary infrastructure (e.g., connections for cooling water or compressed air);
- Any kind of software (e.g., for data acquisition, unit control, simulation) and scripts (e.g., for data analysis) are assigned to the node *SoftwareArchitecture*;
- The node *Procedure* is used, if the process is performed following some rules and standards;
- the node *Operator* designates the operator of the process;
- The node *SubprocessSet* integrates all subprocesses, which are helpful to refine the logic or sequence of the process, but do not need to be described in full detail.

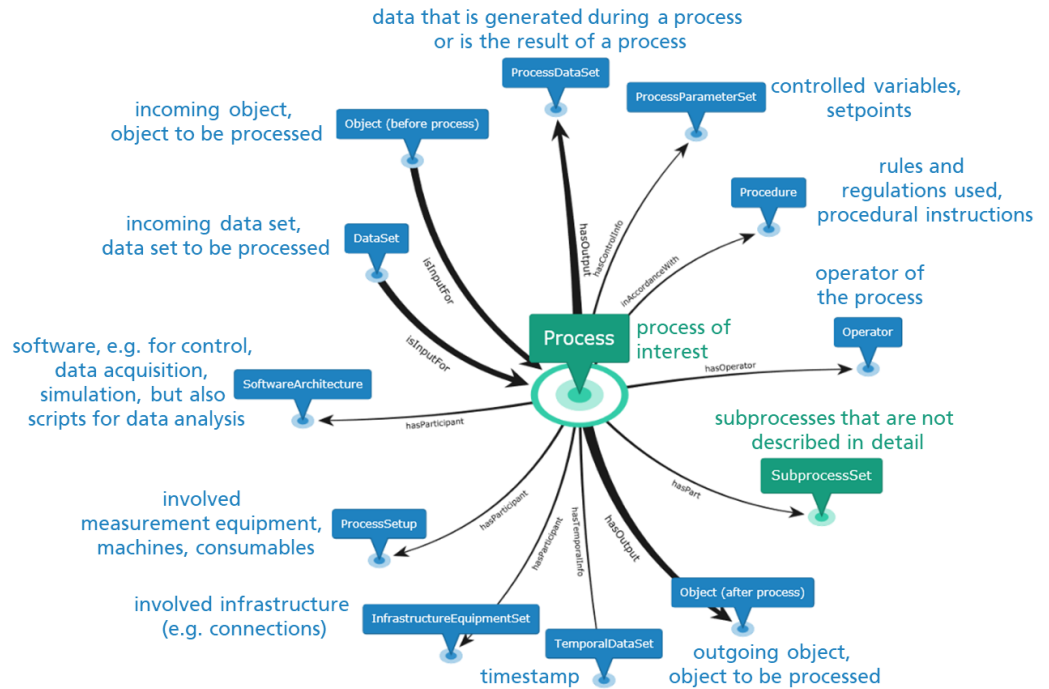
The generic graph template is usually extended and modified but without changing the core structure. This is part of the semantic modeling strategy and is of great importance in generating interoperable linked data for the different digitalized processes and retrieving information based on these common patterns. After the generic Process Graph template is adjusted to semantically describe the metadata generated by the specific modeled process, it is converted into an equivalent Excel template with the functionality **KDB2Excel** from the Graph Designer Tool.

### 2.3.3. Template for Metadata Acquisition

The Excel template for metadata acquisition is exemplary displayed for the *PermanentMoldCasting* process in Figure 6 and contains three sheets. The Spreadsheet *user\_variables* contains the user view for metadata acquisition. Here, the value assignment or data annotation takes place. The names in the column **Variable name** refer to the literal node names from the process graphs (more details later). At this point, it is expected that the domain experts can annotate their data by filling the column **Variable value** of the Excel template with literal values.

In the *gdtriples* spreadsheet, the knowledge graph template structure is stored line by line using node names and categories (subject/object name/type) and edge names

and categories (predicate name/type). This spreadsheet is not intended to be changed by the user.



**Figure 5.** The generic Process Graph template: a common pattern to describe a process.

In the last spreadsheet *namespaces*, the mapping of RDF namespace prefixes to namespaces is stored. It is also not intended to be changed by the user.

### 2.3.4. Handling Cardinalities in Process Graph Templates

In the generic modeling of processes as Process Graphs, one is regularly confronted with the problem that the number of, e.g., outgoing objects from a separation process or the number of chemical elements with corresponding weight fractions can vary within a chemical composition. This problem is solved by adding a suffix to the edge name (e.g., *hasOutput* or *hasPart*) in the process graph. The suffix {1..n} means that the following nodes occur at least once or n-times. Where n is a positive integer, thus allowing to model optional properties or relations. Example: Separating *hasOutput* {1..n} Object.

**1. User view**

Variable name	Variable value
Object ID (ingot)	
Name of producer (ingot)	
Material name (ingot, cast part)	
Unit symbol (temperature, general material info)	
Value (casting temperature, general material info)	
Value (mold temperature as supposed, general material info)	
Unit symbol (density as supposed, general material info)	
Value (density as supposed, general material info)	
Date time start of process (yyyy-mm-dd hh:mm:ss)	
Date time end of process (yyyy-mm-dd hh:mm:ss)	
Object ID (cast part)	
Name of inert gas atmosphere (standard: Argon 4.8)	
Material name (permanent mold)	

**3. Ontology IRI and namespace**

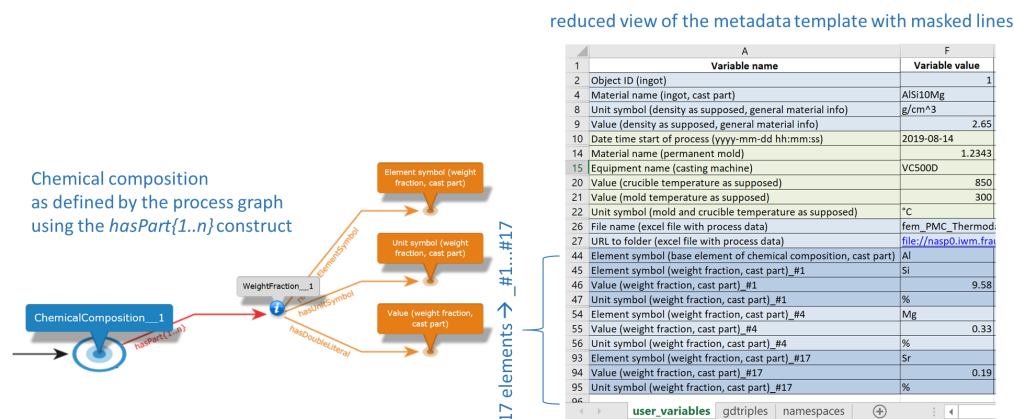
	A	B
1	prefix	qname
2	_default	<a href="https://www.iwm.fraunhofer.de/ontologies/bwmd-ontology">https://www.iwm.fraunhofer.de/ontologies/bwmd-ontology</a>

**2. gdtriples**

	A	B	C	D	E	F
	Subject name	Subject type	Predicate name	Predicate type	Object name	Object type
1	Ingot_1	Ingot	isInputFor	isInputFor	PermanentMoldCasting_1	PermanentMoldCasting
3	PermanentMoldCasting_1	PermanentMoldCasting	hasOutput	hasOutput	CastPart_1	CastPart
4	GeneralMaterialInfo_1	GeneralMaterialInfo	isInputFor	isInputFor	PermanentMoldCasting_1	PermanentMoldCasting
5	PermanentMoldCasting_1	PermanentMoldCasting	hasPart	hasPart	SubprocessSet_1	SubprocessSet
6	SubprocessSet_1	SubprocessSet	hasPart	hasPart	Melting_1	Melting
7	ProcessSetup_1	ProcessSetup	hasPart	hasPart	InertGasAtmosphere_1	InertGasAtmosphere

**Figure 6.** Details of the Excel template for metadata acquisition: (1.) *user\_variables*: user view including the node names of the literals from the process graphs; (2.) *gdtriples*: logic of the process graph; (3.) *namespaces*: ontology IRI and namespace.

The curly brace is removed when the Process Graph template is converted to an Excel file. Literals that follow such a {1..n} construct are marked with the combination of '#' and a sequence number within the Excel template in the user view. A variable-value pair marked in this way can be duplicated by the user in the Excel template section *user\_variables* as needed, as illustrated by Figure 7.



**Figure 7.** Exemplary use of the {1..n} construct with the example of user-specific duplication of chemical elements within a chemical composition.

### 2.3.5. Conversion of Metadata Templates to RDF

The completed Excel templates are converted into RDF using the functionality **Excel2RDF** of The Graph Designer Tool. By assigning values in the Excel template, the Process Graph becomes a graph instance populated with process-specific literal values. All nodes, interpreted as instances of an ontology class, are assigned a unique resource identifier (URI) containing the original node name and a unique identifier. The original node names of the InfoRapid KnowledgeBase Builder software are also retained in the RDF file as labels of the type `rdfs:label`. Similarly, the original node category is kept as `rdf:type`, which implies that the resource is an instance of an `owl:class`.

### 2.4. Linking of Process Datasets

Once all processes have been semantically modeled and all the generated RDF files have been uploaded into a graph database, the next step towards building digital process chains for all the digitalized objects is to link all the individual process graphs generated. The linking is done via the incoming and outgoing objects or datasets of the involved processes and is realized by SPARQL statements creating `owl:same_as` relations. This is achieved by locating all objects modeled as the output of one specific process using the pattern *Process\_x hasOutput Object\_x*, and stating that this object is indeed the same object modeled as *inputFor* the following chronological process, e.g., *Process\_y*, within the physical process chain. However, it is represented as a different instance, e.g., *Object\_y*. Consequently, the statements added to the graph database containing the disconnected process graphs will follow the pattern *Object\_x isSameAs Object\_y*. With the assistance of a reasoner, it can then be inferred that *Object\_x* was indeed *inputFor Process\_y*.

## 3. Results

The following section presents the main characteristics of the generated knowledge database for the 'Use Case Metals' from the MaterialDigital project [30], which is publicly available in RDF format in the Fordatis repository [33]. The added value of the linked data structures to build process chains will be illustrated using exemplary SPARQL queries, representing common knowledge extraction needs in the MSE domain. The integration of machine-readable and context-rich data into a straightforward AI algorithm (a decision regression tree) to predict material properties will be demonstrated by showcasing the

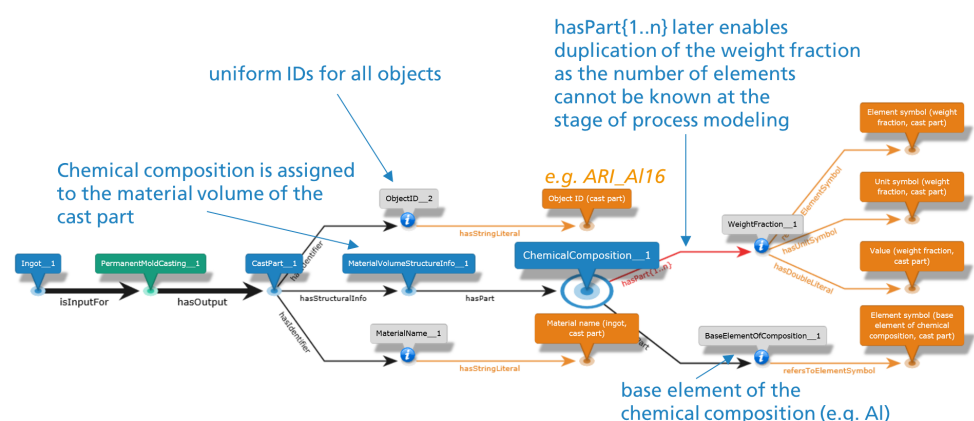
task of identifying the optimal production route to maximize the tensile strength of the investigated AlSi10Mg cast alloy.

### 3.1. Knowledge Base: Manufacturing, Non-Destructive and Destructive Testing of the AlSi10Mg

Using the example of the permanent mold casting process, the application of the BWMD Ontology for the digital representation of MSE processes in the form of process graphs is demonstrated.

Figure 8 shows a portion of the Process Graph of the permanent mold casting process, in which the object identifier (*Object\_ID\_2*) and chemical composition (*ChemicalComposition\_1*) are assigned to the cast part (*CastPart\_1*) after the permanent mold casting process (*PermanentMoldCasting\_1*). The ID assignment construct is uniformly inserted in all processes. When assigning the chemical composition (*ChemicalComposition\_1*), it is made clear that this is a chemical composition that should represent the entire material volume (note the node *MaterialVolumeStructureInfo\_1*). To be as generic as possible when creating the process graph, the particular modeling construct presented in Figure 8 is used for assigning the weight fractions (*WeightFraction\_1*) of the chemical composition. The *hasPart* relation connected to the chemical composition nodes is supplemented by the suffix {1..n}. This is to express that the following nodes can occur at least once. This construct is automatically translated during the conversion into an Excel template for metadata acquisition, as already introduced in Section 2.3.4 so that a user can specify any number of chemical elements and their weight fractions in the Excel template. From the node *WeightFraction\_1*, the element symbol (e.g., Mg, Si), the unit symbol (here %), and the value of the weight fraction are referenced. Additionally, the symbol of the base element (here Al) can be specified.

Another important benchmark of the digitalization strategy with the Graph Designer Workflow is how the process control parameters are semantically connected to the controlled process; an example is illustrated in Figure 9 through the collection node *ProcessParameterSet\_1* from the core structure introduced in Figure 5. All control variables of the permanent mold casting process are connected to the collection node via the *hasPart* relation. For example, the nominal values of the mold temperature, the crucible temperature, and the casting pressure are listed (not all process parameters are shown). All three process parameters are physical quantities assigned a value and a unit symbol. At this point, whether the literal (value) to be specified later by the user is, for example, a value of the type double, integer or string can be expressed.



**Figure 8.** Description of the assignment of an ID and chemical composition to a cast part.

The completion of all the steps of the Graph Designer Workflow as described in Section 2.3 led to a comprehensive knowledge base containing linked data for all the processes listed in Table 1. Figure 10 displays all the linked instances part of the BWMD dataset, a knowledge graph including 278 digitalized process steps linked based on their chronological order within the corresponding process chain they belong to, and 180 instances of the BFO class object (digitalized intermediate products).

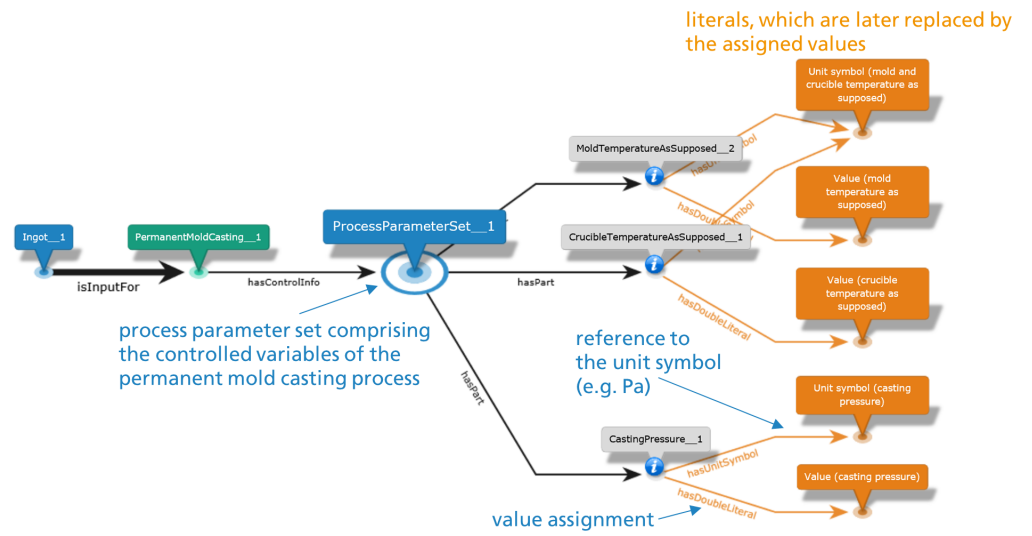


Figure 9. Description of process parameters and handling of physical quantities and unit symbols.

Table 1. List of digitalized processes and corresponding BWMD Ontology class in the BWMD dataset.

Process	Category
permanent mold casting	bwmd:PermanentMoldCasting
3D X-ray computer tomography	bwmd:ThreeDimensionalXRayComputerTomography
wire eroding	bwmd:WireEroding
turning	bwmd:Turning
separating	bwmd:Separating
solution annealing	bwmd:SolutionAnnealing
artificial aging	bwmd:ArtificialAging
tensile test	bwmd:QuasiStaticTensileTest
Brinell indentation	bwmd:BrinellIndentation
casting simulation	bwmd:CastingSimulation

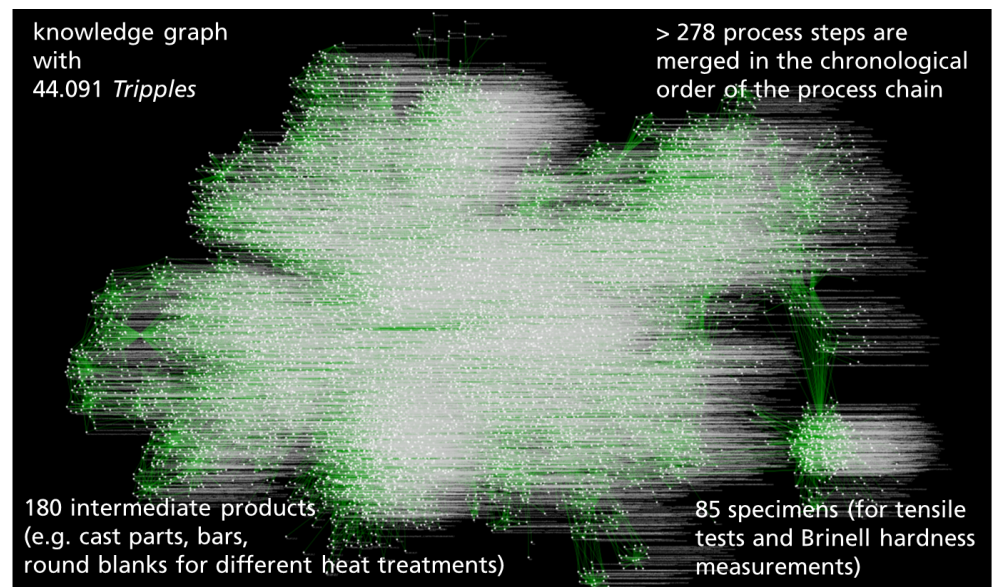


Figure 10. Visualization of the complete knowledge graph.

### 3.2. Knowledge Extraction: SPARQL Queries

The complete knowledge graph depicted in Figure 10 is stored in RDF format. Hence, all the stored information in the graph can be accessed via SPARQL queries. The SPARQL queries are formulated following the core pattern of the generic Process Graph template. The extraction of knowledge from the developed database will be demonstrated through three SPARQL queries designed to retrieve specific information from the large knowledge graph:

1. All the Brinell hardness values;
2. The process control parameters based on the given identifier of one input object;
3. The digital process chain of a particular object based on its label or identifier.

#### 3.2.1. Example 1: Retrieve the Brinell Hardness of All Tested Specimens

Appendix A.1 outlines the SPARQL query algorithm designed to extract the average Brinell hardness values and their corresponding units for all metallographic specimens modeled as input for the Brinell indentation process. It achieves this by matching specific patterns in the WHERE clause and binding the relevant values to the variables defined in the SELECT statement. For a more intuitive interpretation of the query structure, a graphical representation has been created using the ‘sparqlgraphviz’ package for Python [41], as shown in Figure 11. This visualization illustrates the pattern followed by the SPARQL query to identify all instances of the class MetallographicSpecimen, which are modeled as input for an instance of the class BrinellIndentation (process\_node). Each output instance of the BrinellIndentation process is represented by an instance of the ProcessDataSet class, linked to a specific average Brinell hardness value and its corresponding unit. Additionally, every instance of the MetallographicSpecimen class is connected to an instance of the ObjectID class, which is associated with a specific specimen name through the data property hasStringLiteral.

The information retrieved by the SPARQL query is presented in Table 2, displaying only the first five matches for clarity.

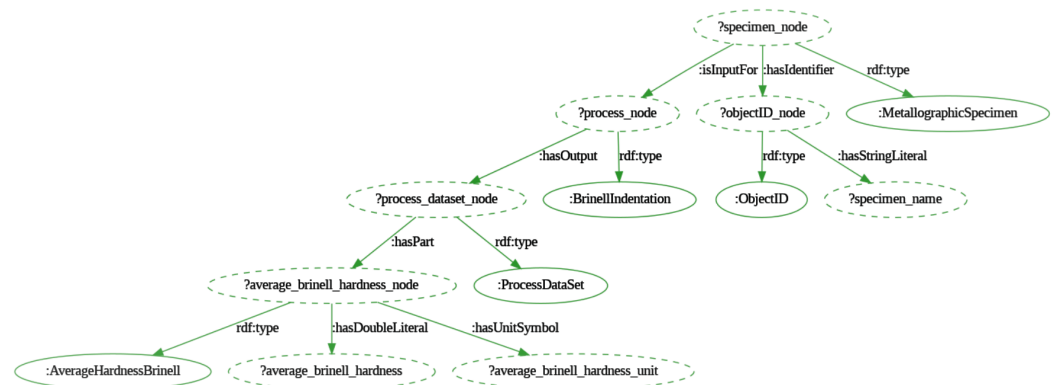


Figure 11. Visualization of Brinell hardness SPARQL query algorithm.

Table 2. First five matches retrieved by the Brinell hardness SPARQL query.

Specimen_Name	Average_Brinell_Hardness	Average_Brinell_Hardness_Unit
ARI_AI19_Stab4_R2	87.8	HB 2.5/62.5
ARI_AI16_S4_R1	65.6	HB 2.5/62.5
ARI_AI26_Stab16_R1	63.7	HB 2.5/62.5
ARI_AI22_Stab8_R9	68.2	HB 2.5/62.5
ARI_AI22_Stab8_R4	68.2	HB 2.5/62.5

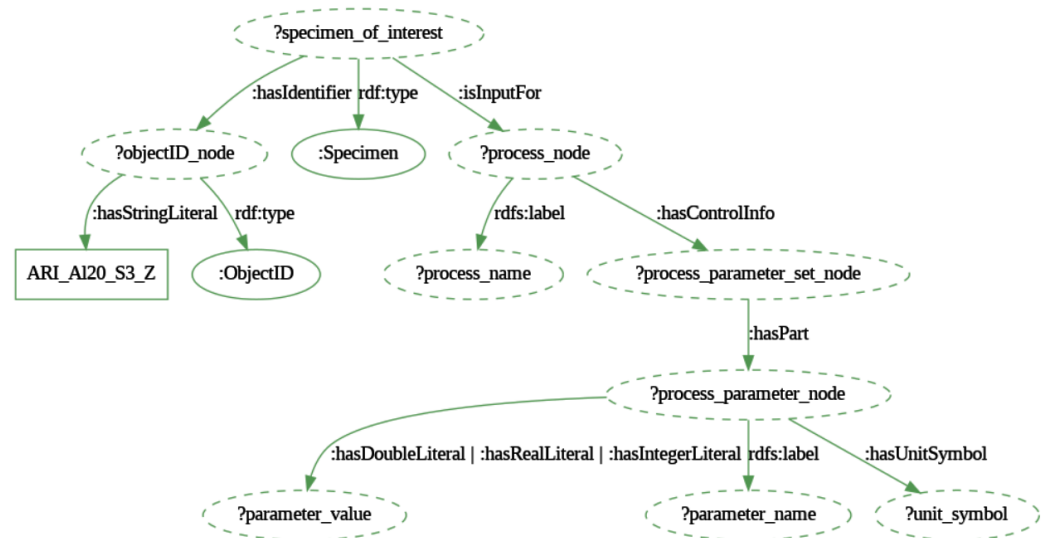
### 3.2.2. Example 2: Retrieve the Process Control Parameters Corresponding to One Specific Object Identifier

This example demonstrates how data structures based on the generic Process Graph template enable the search for process control parameters given a specimen identifier or label. In the present case, the user seeks information regarding a specific specimen, referred to as *specimen\_of\_interest* and labeled with the object identifier ARI\_A120\_S3\_Z. The user wants to know:

- Which process(es) utilized this specimen, i.e., the associated *process\_name*;
- Which set of parameter quantities did control the process(es) for which the *specimen\_of\_interest* was used for;
- The values and corresponding units of the set of parameters controlling the output of the process(es).

This knowledge can be extracted from the RDF graph by running the SELECT query presented in Appendix A.2. A graphical representation of this query is depicted in Figure 12. It shows how first the specimen with given object identifier ARI\_A120\_S3\_Z is located, then follows the pattern given by the object property *isInputFor* and finds the process for which this specimen was used. Finally, through the object-property relationship *hasControlInfo* linking the *process\_node* and the *process\_parameter\_set\_node* instances, finds the connection to the *process\_parameter\_node* instance(s) and their attached data property value(s).

The query result returning the process control parameters, values and units, as well as the process they belong to, for the specimen with given identifier ARI\_A120\_S3\_Z are summarized in Table 3. This result clarifies that the investigated specimen was employed in a quasi-static tensile test, controlled by the specified values for crosshead separation rate and original gauge length, along with their respective units.



**Figure 12.** Visualization of the SPARQL query algorithm retrieving the process control parameters corresponding to one specific object identifier.

**Table 3.** Result of the control process parameters query for the specimen with object identifier ARI\_A120\_S3\_Z.

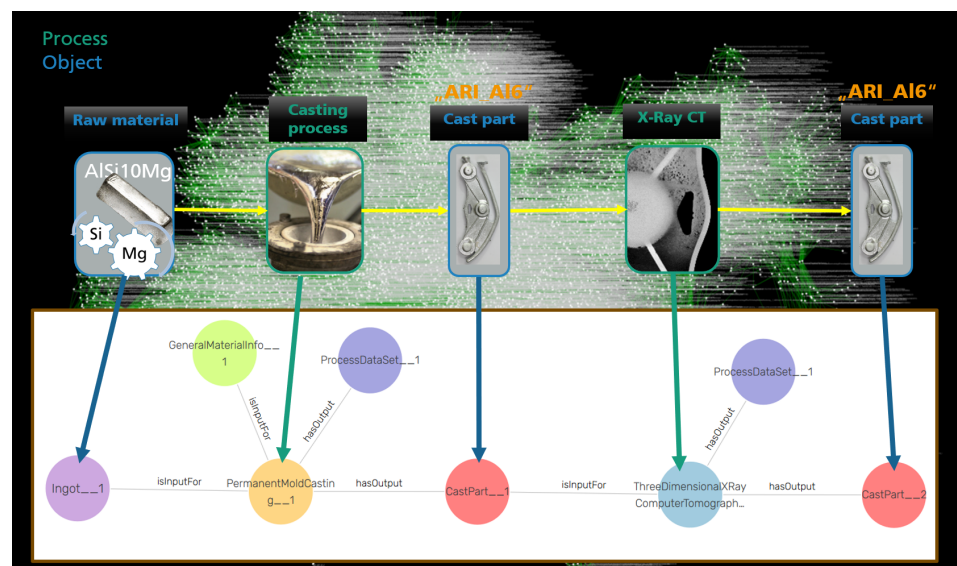
Process_Name	Parameter_Name	Parameter_Value	Unit_Symbol
QuasiStaticTensileTest__1	CrossheadSeparationRate__1	$3.0 \times 10^{-2}$	mm/s
QuasiStaticTensileTest__1	OriginalGaugeLength__1	$9.79 \times 10^7$	$\mu\text{m}$

### 3.2.3. Example 3: Retrieve the Digital Process Chain of One Specific Object

The main scope of this article was to demonstrate the added value of the Graph Designer Workflow to create MSE semantically structured linked process data to build digital process chains, herewith, making all digitalized product history traceable along their corresponding product process chain, often referred to as product life cycle. To accomplish this, a generic SPARQL query was designed based again on the generic digitalization methodology established and following the pattern determined by the generic Process Graph template. This query must enable the user to extract the entire product history of a specific object only based on its given object identifier. The developed generic SPARQL query algorithm for extracting the process chain of a particular object and exemplary demonstrated for the cast part with object identifier ARI\_A16, is presented in Appendix A.3.

It is important to note that to run this query, graph inference needs to be turned on. In addition, the ruleset 'owl2-rl' was selected [42]. The CONSTRUCT keyword included at the beginning of the SPARQL query determines the format of the query result, in this case, in the form of a graph consisting of multiple triple patterns. The overall purpose of this query is to construct a graph pattern based on the specified pattern included within the WHERE clause, which first finds the specimen\_of\_interest based on the given object identifier (in this example 'ARI\_A16' and secondly describes the relationship between entities, connected inwards or backward through the object property chronologicalConnection, whereas the object properties isInputFor and hasOutput are modeled as sub-properties of chronologicalConnection in the BWMD Ontology. The filter included in the WHERE clause ensures that only triples with predicates of type hasOutput, isInputFor, or hasIdentifier are included in the query result.

Figure 13 illustrates graphically the result of the SPARQL process chain query for the object labeled as ARI\_A16. This image displays schematically the mapping between the domain expert (in this example, the material science engineer) knowledge regarding the physical process chain of the object investigated to its digital version, preserving the domain expert knowledge and making it available for other users with no direct possession of such information, as it could be the case of the product supply manager.



**Figure 13.** Visualization of the process chain corresponding to the object ARI\_A16 with the help of the Visual Graph tool of GraphDB database engine [39].

### 3.3. From Knowledge Extraction to New Knowledge Generation

Generating rich semantic structured MSE data with connected process graphs along process chains offers significant benefits. This data can be utilized to feed AI models, enabling new insights, optimization of production steps or mechanical properties, and per-

forming reliable predictions. In this section, a Decision Tree Regression model is trained and validated to illustrate how the RDF graph created for the studied AlSi10Mg in the form of chronologically linked process data can be exploited to find the optimal production parameters aiming to improve the tensile strength ( $R_m$ ) of the investigated material. For this didactic purpose, the relatively simple Decision Tree Regression model (the *DecisionTreeRegressor* model from the 'sklearn.tree' module included in the 'scikit-learn' Python package [43]) was selected. This model works by continuously splitting a training dataset based on selected feature variable values, resulting in a tree-like structure that performs predictions on the target variable(s). As a result, a trained model capable of assessing the impact of the included feature variables on the target is generated.

The first step in the modeling procedure was to query from the knowledge base all the conducted tensile tests including the previous history information of each tested specimen (the chemical composition from the casting process and the undergone heat treatment steps with their corresponding process control parameters), as well as the resulting material tensile strength from the performed tensile tests. Secondly, the target and feature variables were selected, where:

- $R_m$  was chosen as the **target variable** or independent variable, to make predictions on;
- The magnesium weight fraction ('weightfraction\_Mg[%]'), the annealing duration ('duration\_annealing[min]'), the annealing temperature ('temperature\_annealing[°C]'), the aging duration ('duration\_aging[min]') and the aging temperature ('temperature\_aging[°C]'), represent the **feature variables** or dependent variables expected to influence the target variable.

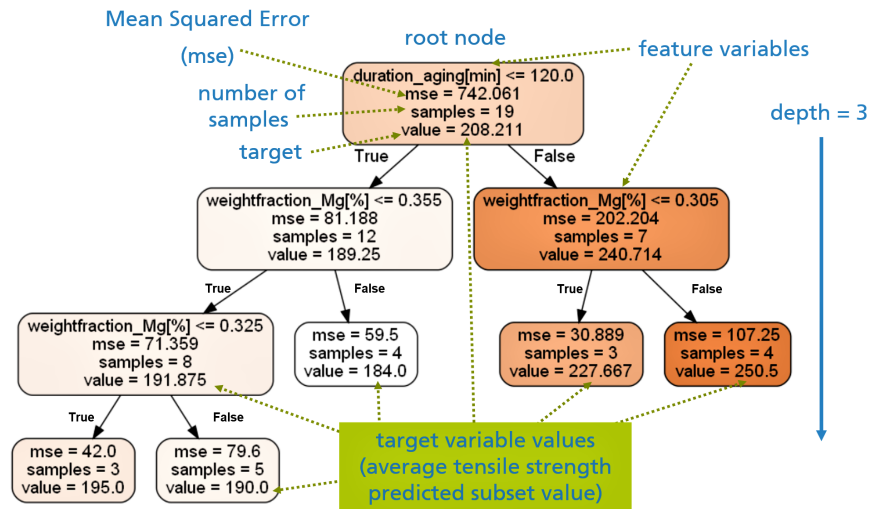
In the next step, the retrieved dataset is randomly split into **training data** (70 % of the dataset) and **validation data** (30 % of the dataset). Then, the training data is used to train the Decision Tree Regression model, which is configured through the following hyperparameters:

- `max_depth = 3`, controlling the maximum depth of the tree;
- `min_samples_split = 6`, which determines the minimum amount of samples required to split an internal node;
- `min_samples_leaf = 3`, representing the minimum amount of samples required per leaf.

In this study case, the model splits the **training data** into subsets applying the default criterion (the Mean Squared Error) by selecting a feature and a threshold value of the selected feature and searching for the split that minimizes the error of each subset. The splitting process stops once the maximum depth of the tree or the minimum samples per split is reached. The tree grows per split with newly generated branches, composed of nodes representing a decision (internal node) or a prediction (leaf nodes). Figure 14 displays the tree-like form of the fitted Decision Tree Regression model. The model predicts that the optimal production route to maximize the tensile strength of the cast AlSi10Mg is achieved for an amount of Mg in the cast melt greater than 0.305 % and an aging duration greater than 120 min, where the prediction of the target variable results from averaging the tensile strength value of a data subset composed by four samples. It is observed that other included feature variables in the fitted model are not displayed in the tree; this is explained by the low-scored level of importance of these missing variables. The scored level of importance values for performing predictions with the trained model of each considered feature variable are collected in Table 4. These values indicate that the main factor influencing the value of the tensile strength (based on the used training data) of the investigated material is the aging duration applied during the heat treatment process. Meanwhile, the amount of Mg in the melt added during the casting process seems to have a small influence on the targeted tensile strength value. In contrast, the other considered production parameters do not have any relevant impact on the tensile strength value.

The model prediction quality has been assessed based on the R-squared score or coefficient of determination, whose value indicates how well the fitted model can capture the variability of the target variable. The showcased fitted model performs with an R-squared equal to 0.89, relatively close to 1, and therefore delivers reasonably good predictions.

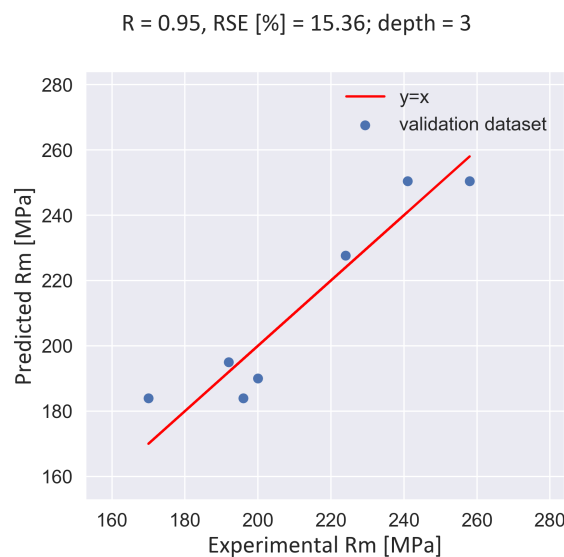
The last step to follow at this stage is to test the goodness of the model to perform predictions of the target value based on the left side **validation data**. The validation of the trained Decision Tree Regression model is depicted in Figure 15. The data points represent the predicted tensile strength values by the trained model versus the experimental ones from the validation dataset. The red line stands for the perfect prediction line.



**Figure 14.** Decision Tree Regression model to predict the tensile strength of the cast AlSi10Mg based on process chain graph data.

**Table 4.** Level of importance for making predictions.

Feature Variable	Level of Importance [%]
duration_aging [min]	91.4
weightfraction_Mg [%]	8.6
temperature_aging [°C]	0.0
temperature_annealing [°C]	0.0
duration_annealing [min]	0.0



**Figure 15.** Validation of the trained Decision Tree Regression model with a correlation coefficient ( $R$ ) of 95%, a coefficient of determination ( $R^2$ ) of 89%, and a relative standard error (RSE) of the slope equal to 15.36%.

#### 4. Discussion

While many papers have focused on developing material ontologies [13,44–47], fewer have addressed the design of knowledge graphs for structuring material data through knowledge graph patterns (e.g., [48,49]). Many examples of structuring MSE process data focus on ontology-based semantic frameworks designed to digitize a specific process. For instance, PMD-Core provides an ontology-based semantic data model for digitizing tensile tests with the Tensile Test Ontology [47]. Similarly, MSEO offers a data schema for semantically modeling hardness measurements [50]. Moreover, PMD-Core [51] supplies a framework for connecting processes with relevant (meta)data. This framework consists of a coarse-grained structure intended to chronologically connect processes and associate them with their input and output objects. So far, no investigations have concentrated on developing a fined-grain and generic method for digitizing material data using a semantic data model as a core pattern. This is the objective of the generic Process Graph template introduced in Section 2.3.2.

This work aims to design a digitalization strategy focused on developing semantic data models for impartible processes, as outlined in Section 2. Despite the availability of existing material ontologies, an ontology specifically tailored for the digitalization strategy used to create digital process chains within the MSE domain was presented in Section 2.2. The BWMD Ontology employs BFO 2.0 as its top-level ontology, ensuring compatibility with the ISO/IEC 19763 standard [52]. This establishes a solid foundation for creating interoperable domain and application ontologies.

The BWMD Ontology is the product of a two-year collaborative effort among MSE experts from various German research institutes initiated in 2018. Although this paper references its non-modularized version for simplicity, a modularized version exists and is published in [53]. This modularized version includes a mid-level module for general MSE concepts and a domain module for specific concepts related to mechanical testing procedures, microstructural features, and specimen preparation. Additional modules have been created to address various use cases beyond the AlSi10Mg dataset. Extensions of the BWMD Ontology that describe other case studies or projects can be explored in [54]. For the AluTrace use case, an application ontology module was created to digitalize the Laser Bed Powder Fusion process to additively manufacture AlSi10Mg and to discover optimal production routes (including different heat treatments) with the help of computer tomography scans, mechanical testing (tensile and fatigue), defect analysis and metallographic investigations. The AluTrace dataset is well-documented in [55]; individual process-graph patterns, RDF data, and SPARQL queries are available for download, as well as the BWMD-based ontology modules created to describe the use case. The AluTrace dataset exposes the versatility of the proposed framework.

Nevertheless, this article does not advocate for a specific ontology to digitalize material data. Parallel developments of mid-level and domain-specific ontologies, such as the PMD-Core Ontology [21] from the Material Digital Platform initiative [56], are equally suitable for modeling MSE data. The PMD-Core Ontology was initially developed with a bottom-up approach to align with the EMMO [20] and has since been adapted for interoperability with BFO [51]. A unique advantage of the PMD-Core Ontology is that it is built on the W3C Provenance Ontology (PROV-O) [57]. This foundation provides a standardized framework for managing various data sources, facilitating data integration from different origins, and creating digital workflows [58]. While the BWMD Ontology manages data integration from diverse sources, it does not count with a specialized vocabulary for developing workflows. Conversely, the BWMD Ontology includes an extensive vocabulary for semantically describing the MSE domain. Developing a BWMD ontology module based on the PMD-Core mid-level ontology will be advantageous in leveraging the synergies between both ontologies. Furthermore, another ontology, the MSEO [22], developed as part of the Materials Open Laboratory initiative, also aligns with the BFO principles. The MSEO uses the Common Core Ontologies (CCO) [59] as its framework, and the community involved in its development has worked to integrate most of the BWMD concepts into its

domain vocabulary. This effort is demonstrated by 977 mappings between the MSEO and BWMD [60]. Besides, a mechanical testing domain module [24] was created within the Urwerk project [61], partially covering the MSE domain. Creating mappings between these MSE ontologies can enhance interoperability and mutual benefits. It is worth mentioning that showcasing the application of the Graph Designer Workflow with the BWMD Ontology was a merely practical decision, as the fitting ontology solution was already available for semantically describing the AlSi10Mg dataset. However, transferring all the necessary domain concepts to a different mid-level ontology, such as the PMD-Core Ontology, is not an issue.

In addition to the Graph Designer Tool, other digitalization tools are available for material data. Tools like OpenRefine [62], TARQL [63], and Ontopanel [50] offer different approaches for digitalizing and structuring MSE data. However, the advantages of the Graph Designer Workflow, as presented in this work, are significant. The Graph Designer Tool [64] provides a user-friendly solution for uploading ontology vocabularies into a generic Process Graph template, managing multiple namespaces, and converting the Graph Template into a standard Excel Table for data annotation. This minimizes errors from misinterpreting domain vocabulary and allows domain experts to populate the Process Graph template with real data.

The proposed MSE data digitalization strategy successfully achieves the goal of bridging the existing knowledge gap between domain and ontology experts. While not technological, this challenge poses significant obstacles in digitalizing material science knowledge. Effective information exchange between domain and ontology experts is crucial for designing appropriate semantic models. Understanding the meaning of classes when incorporating new vocabulary into the domain ontology is essential for selecting suitable upper classes from mid-level or top-level ontologies. The quality of this exchange will determine the value of the generated semantically structured data, with a richer vocabulary providing deeper context and knowledge. The interview template introduced in Section 2.3.1 facilitates efficient information exchange, guiding the classification of new concepts according to the semantic schema of the generic Process Graph template in Section 2.3.2. This significantly aids ontology experts in adjusting templates and creating compliant data models. The challenges in knowledge extraction are often overlooked in literature despite being well-known in the ontology development community. Additionally, using NLP (Natural Language Processing) to extract information from domain-specific literature is a powerful tool for semi-automated domain ontology creation.

The generic aspect of the Graph Designer Workflow implies its possible application with other ontologies or semantic data models. The presented work demonstrates its applicability based on the BWMD Ontology and the generic Process Graph template as a semantic data model to efficiently build digital process chains for the cast AlSi10Mg use case. The BWMD Dataset RDF graph, depicted in Figure 10, has been introduced in Section 3. This RDF graph is publicly available in the Fordatis database [33]. The primary objective of this work is to present a generic digitalization method for semantically modeling various processes using different ontologies as base vocabularies and instantiating schemas or generic Process Graph templates. The proposed methodology is demonstrated across ten different processes (see Table 1), showcasing the generality of the solution and the semantic depth of the BWMD Ontology. As the disparity of modeled process data increases, more concepts are required to encapsulate the knowledge, leading to an enriched ontology vocabulary. In addition to the RDF graph, two Jupyter Notebooks are also accessible with the created knowledge base for the cast AlSi10Mg [33], including further SPARQL queries and data visualization features. This aims to add a dimension of user experience to the presented study, as this article intends to be informative and didactic.

The advantages of the generic Process Graph template become apparent through the introduced SPARQL query examples in Section 3.2. By following the core pattern of the generic process data model, generic query algorithms can be developed to efficiently and easily extract information stored in the RDF graph connecting all the different digitalized

processes. The structure enables retrieval of specific material properties generated via one specific test (see Section 3.2.1) and allows the extraction of other types of information, such as the controlled parameters applied to conduct one specific test (see Section 3.2.2). Unlike traditional SQL databases, graph databases allow for extracting both tabular and graph data based on patterns, thanks to the CONSTRUCT SPARQL query structure. This is showcased in Section 3.2.3, where the digital process chain of one specific object is accessed via SPARQL query. The retrieved object is a piece of the complete RDF Graph and can be visualized as displayed in Figure 13. This example shows the utility of the proposed methodology in creating a data infrastructure that enables a product traceability feature along its digitalized manufacturing process chain. Both the generic Process Graph template and the ‘process chain’ query algorithm could be extended to include further process-related information details such as the CO<sub>2</sub> footprint or the cost associated with the conducted manufacturing step, which could later be used in conjunction with other object history information to find production routes for decreased environmental impact or for reducing production costs of a particular product.

Access to product history information enables the discovery of new production patterns, leading to optimized material properties, as illustrated in Section 3.3. Feeding AI models with context-rich data enhances the material design process, helping to find optimal production paths and allowing for quick solutions to describe phenomena not considered in physical models. If an AI-trained model is fed with a high-quality database (as would be the case of a context-rich graph database), reliable predictions can be made. In the MSE field, this could imply significant cost reduction by creating reduced or optimized material testing programs. Another advantage would be developing robust hybrid models, which would better reproduce the dispersion of apparently similarly generated material experimental data, i.e., generated from similar testing conditions. In such cases, the material scatter can be explained by small differences detected in the control parameters along the digitalized process chain of each tested specimen, e.g., in one or several heat treatment steps, as showcased for the material AlSi10Mg regarding the impact of aging duration on the resulting tensile strength of the material. While this information is common knowledge for the material production expert, it is not necessarily shared by the material engineer who performs simulations with this material, and the information might remain hidden in the form of experimental scatter. Another hidden effect discussed in Section 3.3 is the impact of the variation of the Mg amount in the melt on the tensile strength of the investigated material.

The presented Graph Designer Workflow and the BWMD Ontology provide a robust framework for digitalizing material data and creating digital process chains. Furthermore, the digitalized context-rich semantic MSE data offer the perfect basis for training AI models and generating new reliable knowledge. The generic nature of the proposed digitalization framework and its potential for further development make it a valuable tool for advancing material science and engineering digitalization efforts. While the Graph Designer Workflow offers numerous advantages, limitations are inherent to working with graph databases. These limitations include the complexities associated with semantic data models and the intricacies of SPARQL queries. Nevertheless, recent advancements in text foundation models offer potential improvements in designing systems that interact with graph databases using natural language. Furthermore, processing high-dimensional or highly interconnected data presents several challenges, including increased computational load and potential performance bottlenecks in graph databases. To handle this problem, the development of scalable algorithms (such as the graph query engine Qlever [65]), and the use of sophisticated methods to handle non-linearities and dynamic dependencies are required [66]. Future work will focus on tackling these challenges, improving user interfaces to reduce the dependency on specialized tools (e.g., Protégé, InfoRapid KnowledgeBase Builder), and exploring the application of the proposed digitalization methodology in other domains.

## 5. Conclusions

This work aims to accelerate material optimization and discovery by digitalizing data from material process chains in a connected manner, exposing the composition-process-structure-property chain in MSE. Knowledge graphs link data from various processes within material process chains. Using a generic template as the core structure for creating semantic models simplifies information retrieval, allowing for tracking the complete history of specific labeled objects. Additionally, using ontologies ensures unambiguous and interoperable data enriched with domain knowledge. A digitalization framework, the so-called Graph Designer Workflow, is presented to address these needs. It is an ontology-agnostic robust methodology for storing process-specific data and domain knowledge into linked knowledge graphs. It includes an ontology-based generic Process Graph template and step-by-step instructions for applying the methodology to create digital process chains. The resulting machine-understandable data adheres to the FAIR principles. The entire pipeline is presented in a didactic manner, making it easy for MSE-domain experts to adopt the workflows for their own digitalization requirements.

Furthermore, the proposed digitalization methodology is demonstrated on a use case involving the permanent mold casting of AlSi10Mg, encompassing variations in chemical composition, specimen extraction, heat treatment, non-destructive characterization and mechanical testing. To this end, the BWMD Ontology is used. The resulting connected knowledge graph, or knowledge base, is employed to retrieve and analyze data to investigate the influence of heat treatment and magnesium content on the strength properties characterized by tensile tests. MSE-related examples of knowledge extraction are also presented, including the retrieval of the digital process chain of a labeled object, showcasing the product traceability feature within the Graph Designer Workflow.

From a scientific and technical point of view, the following increments were made concerning the digital mapping of material-intensive process chains:

- The BWMD Ontology was introduced, which is based on the Basic Formal Ontology and provides the necessary mid-level and domain-specific vocabulary for the use case of AlSi10Mg. The mid-level and domain-level parts of the ontology are also available as separate modules. The mid-level ontology module is designed to be general to the MSE domain. Comparisons and connections were made to parallel mid-level ontology design efforts in the MSE domain, highlighting the need to create mappings between these ontologies.
- A generic Process Graph template for mapping material data along process chains was introduced, respecting the taxonomy and semantic rules of the mid-level BWMD Ontology. The Process Graph template enables a digital representation of impartible processes that are part of material process chains. While digitalization efforts like PMD [21] and UrWerk [61] have described modeling of specific processes and process chains in MSE, the present work provides a generic template that can be adapted for modeling any impartible processes. Further, the ‘instantiation’ of this template for ten different processes was presented, demonstrating its versatility.
- The Graph Designer Workflow was presented to generate semantic RDF data. The tools part of the workflow are publicly available in [64]. An interview template is introduced to reduce the knowledge gap between domain and ontology experts, aiming to regulate and formalize information exchange. The digitalization workflow is designed to make the user interaction steps easy and intuitive. Users can design the data schema for describing processes in the interview template and later in a GUI without mastering any schema languages. Tools are created to convert the GUI template to Excel files, which are used for data annotation. Here, users can instantiate data in a flat hierarchy, unlike the requirement to preserve data hierarchy in workflows based on schema languages such as LinkML [67] and JSON-Schema. These advantages have led to the adoption of the Graph Designer Workflow in digitalization efforts involving diverse materials and processing steps. Examples include AluTrace (referenced as the Fraunhofer IWM toolchain [55,68]), iBain (dilatometry, isothermal bainitic trans-

formation, axial and bending resonance fatigue tests of bainitic steel [56]), UrWerk (process chain of 100Cr6 steels including heat treatment, surface finishing, and fatigue tests [61]), H2Digital (hydrogen fatigue test [69]), etc.

- A simple and interpretable machine learning model illustrates the benefit of creating semantic-rich MSE data. A Decision Tree Regression model is used to predict the maximum tensile strength as a function of heat treatment and chemical composition.

This work successfully accelerates materials discovery by digitalizing data from material process chains using a generic methodology. The resulting knowledge graphs adhere to the FAIR principles and provide a valuable resource for materials-domain experts. Future work will focus on addressing the challenges of semantic data modeling and further improving user experience.

**Author Contributions:** Conceptualization: C.S., P.v.H., A.T. and E.G.T.; methodology: C.S., P.v.H., A.T., M.J.-R. and E.G.T.; software: P.v.H. and A.T.; validation: C.S., P.v.H., A.T. and E.G.T.; formal analysis: C.S., P.v.H., A.T. and E.G.T.; investigation: C.S., P.v.H., A.T. and E.G.T.; resources: C.S., P.v.H., A.T. and E.G.T.; data curation: C.S., P.v.H., A.T. and E.G.T.; writing—original draft preparation: E.G.T., C.S., A.T. and M.J.-R.; writing—review and editing: E.G.T., C.S., P.v.H. and A.T.; visualization: C.S., E.G.T., A.T. and M.J.-R.; supervision: E.G.T., C.S. and A.T.; project administration, C.S.; funding acquisition, C.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** The study has received funding from the ‘Ministerium für Wirtschaft, Arbeit und Wohnungsbau Baden-Württemberg’ under the project MaterialDigital, funding number 3-4332.62-IWM/21.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The complete BWMD dataset, including two Jupyter Notebooks with integrated SPARQL queries to directly interact with the graph data, is published and available for download under [MaterialDigital Dataset](#) accessed on 5 December 2024.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

FAIR	Findable, Accessible, Interoperable, and Reusable
AI	Artificial Intelligence
IT	Information Technology
ICSD	Inorganic Crystal Structure Database
OQMD	Open Quantum Materials Database
MDF	Materials Data Facility
NOMAD	Novel Materials Discovery
MSE	Materials Science and Engineering
W3C	World Wide Web Consortium
BFO	Basic Formal Ontology
OWL	Web Ontology Language
LLMs	Large Language Models
IAO	Information Artifact Ontology
RDF	Resource Description Framework
EU	European Union
SLACKS	Semantic Linked Data Stack
MMKG	Materials Modeling Knowledge Graph
BWMD	Baden Württemberg Material Digital
SPARQL	SPARQL Protocol and RDF Query Language
PMD	Platform Material Digital
TARQL	SPARQL for Tables
NLP	Natural Language Processing
GUI	Graphical User Interface
JSON	JavaScript Object Notation

## Appendix A SPARQL Queries

### Appendix A.1 SPARQL Query to Retrieve the Brinell Hardness of All Tested Specimens

```

PREFIX : <https://www.iwm.fraunhofer.de/ontologies/bwmd-ontology#>
SELECT
?specimen_name
?average_brinell_hardness
?average_brinell_hardness_unit
WHERE {
    # pattern for specimen name
    ?specimen_node a :MetallographicSpecimen;
        :hasIdentifier ?objectID_node;
        :isInputFor ?process_node.
    ?objectID_node a :ObjectID;
        :hasStringLiteral ?specimen_name.

    # pattern for brinell test results
    ?process_node a :BrinellIndentation;
        :hasOutput ?process_dataset_node.
    ?process_dataset_node a :ProcessDataSet;
        :hasPart ?average_brinell_hardness_node.
    ?average_brinell_hardness_node a :AverageHardnessBrinell;
        :hasDoubleLiteral ?average_brinell_hardness;
        :hasUnitSymbol ?average_brinell_hardness_unit.
}

```

### Appendix A.2 SPARQL Query to Retrieve the Process Control Parameters Corresponding to One Specific Object Identifier

```

PREFIX : <https://www.iwm.fraunhofer.de/ontologies/bwmd-ontology#>
SELECT
?process_name
?parameter_name
?parameter_value
?unit_symbol
WHERE{
?specimen_of_interest a :Specimen;
    :hasIdentifier ?objectID_node.
?objectID_node a :ObjectID;
    :hasStringLiteral "ARI_A120_S3_Z".
?specimen_of_interest :isInputFor ?process_node.
?process_node rdfs:label ?process_name;
    :hasControlInfo ?process_parameter_set_node;
    rdfs:label ?process_name.
?process_parameter_set_node :hasPart ?process_parameter_node.
?process_parameter_node rdfs:label ?parameter_name;
    (:hasDoubleLiteral|:hasRealLiteral|:hasIntegerLiteral)
    ?parameter_value;
    :hasUnitSymbol ?unit_symbol.
FILTER (?parameter_value != xsd:double("NaN"))
}

```

### Appendix A.3 SPARQL Query to Retrieve the Digital Process Chain of One Specific Object

```

PREFIX : <https://www.iwm.fraunhofer.de/ontologies/bwmd-ontology#>
CONSTRUCT{

```

```

?some_node ?p_outer ?some_other_node.
}
WHERE{
# Find the specimen of interest
?specimen_of_interest a :Object;
    :hasIdentifier ?objectID_node.
?objectID_node a :ObjectID;
    :hasStringLiteral "ARI_A16".
# Apply pattern for process chain
?specimen_of_interest (~:chronologicalConnection)*
| (:chronologicalConnection)* ?some_node.
{
?some_node ?p_outer ?some_other_node.
filter(?p_outer=:hasOutput||?p_outer=:isInputFor||?p_outer=:hasIdentifier).
}
}

```

## References

- Callister, W.; Rethwisch, D. *Materials Science and Engineering: An Introduction*; Wiley: Hoboken, NJ, USA, 2020.
- Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [[CrossRef](#)] [[PubMed](#)]
- Web Standards. 2024. Available online: <https://www.w3.org/standards/> (accessed on 27 November 2024).
- Rachapalli, J.; Khadilkar, V.; Kantarcioglu, M.; Thuraisingham, B. Towards fine grained RDF access control. In Proceedings of the 19th ACM Symposium on Access Control Models and Technologies, London, ON, Canada, 25–27 June 2014; pp. 165–176. [[CrossRef](#)]
- OpenAI. ChatGPT: Applications and Impacts. 2022. Available online: <https://openai.com/research/chatgpt> (accessed on 27 November 2024).
- Murthy, R.N.; Potu, S.T.; Thomas, A.; Mishra, L.; Prange, N.; Durmaz, A.R. Ontology-conformal recognition of materials entities using language models. *ChemRxiv* **2024**. [[CrossRef](#)]
- Pan, S.; Luo, L.; Wang, Y.; Chen, C.; Wang, J.; Wu, X. Unifying large language models and knowledge graphs: A roadmap. *IEEE Trans. Knowl. Data Eng.* **2024**, *36*, 3580–3599. [[CrossRef](#)]
- Bergerhoff, G. Data base for inorganic crystal structures. *Comput. Phys. Commun.* **1984**, *33*, 79–84. [[CrossRef](#)]
- Saal, J.E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM* **2013**, *65*, 1501–1509. [[CrossRef](#)]
- Jain, A.; Ong, P.; Hautier, G.; Chen, W.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K.A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation Commentary: The Materials Project: A materials genome. *APL Mater.* **2016**, *1*, 011002. [[CrossRef](#)]
- Blaiszik, B.; Chard, K.; Pruyne, J.; Ananthakrishnan, R.; Tuecke, S.; Foster, I. The Materials Data Facility: Data Services to Advance Materials Science Research. *JOM J. Miner. Met. Mater. Soc.* **2016**, *68*, 2045–2052. [[CrossRef](#)]
- Draxl, C.; Scheffler, M. NOMAD: The FAIR concept for big data-driven materials science. *MRS Bull.* **2019**, *44*, 124–132. [[CrossRef](#)]
- Bayerlein, B.; Hanke, T.; Muth, T.; Riedel, J.; Schilling, M.; Schweizer, C.; Skrotzki, B.; Todor, A.; Moreno Torres, B.; Unger, J.F.; et al. A perspective on digital knowledge representation in materials science and engineering. *Adv. Eng. Mater.* **2022**, *24*, 2101176. [[CrossRef](#)]
- Hill, J.; Mulholland, G.; Persson, K.; Seshadri, R.; Wolverton, C.; Meredig, B. Materials science with large-scale data and informatics: Unlocking new opportunities. *MRS Bull.* **2016**, *41*, 399–409. [[CrossRef](#)]
- Soedarmadji, E.; Stein, H.; Suram, S.; Guevarra, D.; Gregoire, J. Tracking materials science data lineage to manage millions of materials experiments and analyses. *Npj Comput. Mater.* **2019**, *5*, 79. [[CrossRef](#)]
- Decker, S.; Erdmann, M.; Fensel, D.; Studer, R. *Ontobroker: Ontology Based Access to Distributed and Semi-Structured Information*; Springer US: New York, NY, USA, 1999; pp. 351–369. [[CrossRef](#)]
- Antoniou, G.; Harmelen, F.V. *Web Ontology Language: OWL*; Springer: Berlin/Heidelberg, Germany, 2004; pp. 67–92. [[CrossRef](#)]
- Ariza-Colpas, P.; Piñeres-Melo, M.; Nieto-Bernal, W.; Morales-Ortega, R. WSIA: Web Ontological Search Engine Based on Smart Agents Applied to Scientific Articles. In *Advances in Swarm Intelligence: 10th International Conference, ICSI 2019, Chiang Mai, Thailand, 26–30 July 2019*; Springer International Publishing: New York, NY, USA, 2019; pp. 338–347. [[CrossRef](#)]
- Smith, B.; Almeida, M.; Bona, J.; Brochhausen, M.; Ceusters, W.; Courtot, M.; Dipert, R.; Goldfain, A.; Grenon, P.; Hastings, J.; et al. *Basic Formal Ontology 2.0: Specification and User's Guide*; 2015; Version 2.0. <https://raw.githubusercontent.com/BFO-ontology/BFO/master/docs/bfo2-reference/BFO2-Reference.pdf>.

20. Horsch, M.; Chiacchiera, S.; Schembera, B.; Seaton, M.; Todorov, I. Semantic interoperability Based on the European Materials and Modelling Ontology and its Ontological Paradigm: Mereosemantics. In Proceedings of the 14th WCCM-ECCOMAS Congress, virtual, 11–15 January 2021. [CrossRef]
21. PMD Core Ontology (PMDco) v2.0.8. 2024. Available online: <https://github.com/materialdigital/core-ontology> (accessed on 2 October 2024).
22. Hanke, T. MSEO Ontology. 2023. Available online: <https://matportal.org/ontologies/MSEO> (accessed on 15 June 2023).
23. Austin, T.; Bullough, C.; Gagliardi, D.; Loveday, M.; Leal, D. *A Guide to the Development and Use of Standard Compliant Data Format for Engineering Materials Test Data*; European Committee for Standardisation (CEN-CENELEC): Brussels, Belgium, 2010.
24. Domain: Mechanical Testing. 2024. Available online: <https://github.com/emmo-repo/domain-mechanical-testing?tab=readme-ov-file> (accessed on 7 October 2024).
25. Schilling, M. Towards Digitalization of Materials in PMD: An Application Ontology of the Tensile Test. 2021. Available online: <https://opus4.kobv.de/opus4-bam/frontdoor/index/index/docId/534811> (accessed on 5 December 2024).
26. Usman, Z.; Young, R.; Chungoora, N.; Palmer, C.; Case, K.; Harding, J. Towards a formal manufacturing reference ontology. *Int. J. Prod. Res.* **2013**, *51*, 6553–6572. [CrossRef]
27. Ali, M.; Yang, R.; Zhang, B.; Furini, F.; Rai, R.; Otte, J.; Smith, B. Enriching the functionally graded materials (FGM) ontology for digital manufacturing. *Int. J. Prod. Res.* **2020**, *59*, 5540–5557. [CrossRef]
28. Norouzi, E.; Waitelonis, J.; Sack, H. The landscape of ontologies in materials science and engineering: A survey and evaluation. *arXiv* **2024**, arXiv:2408.06034. [CrossRef]
29. Brandt, N.; Garabedian, N.T.; Schoof, E.; Schreiber, P.J.; Zschumme, P.; Greiner, C.; Selzer, M. Managing FAIR tribological data using Kadi4Mat. *Data* **2022**, *7*, 15. [CrossRef]
30. Schweizer, C.; Reichenbach, R.; Butz, A.; Lienhard, J.; Herrmann, T.; Preußner, J.; Hartrott, P.v.; Friedmann, V.; Wessel, A.; Thomas, A.; et al. *Abschlussbericht zu MaterialDigital*; 2020. [CrossRef]
31. Metal solutions EOS Aluminium AlSi10Mg Material Data Sheet. 2024. Available online: [https://www.eos.info/de/var/assets/03\\_system-related-assets/material-related-contents/metal-materials-and-examples/metal-material-datasheet/aluminium/material\\_dasheet\\_eos\\_aluminium-alsi10mg\\_en\\_web.pdf](https://www.eos.info/de/var/assets/03_system-related-assets/material-related-contents/metal-materials-and-examples/metal-material-datasheet/aluminium/material_dasheet_eos_aluminium-alsi10mg_en_web.pdf) (accessed on 5 December 2024).
32. Basic Formal Ontology (BFO) 2.0 OWL. 2020. Available online: <https://raw.githubusercontent.com/BFO-ontology/BFO/master/releases/2.0/bfo.owl> (accessed on 5 December 2024).
33. Schweizer, C.; Reichenbach, R.; Butz, A.; Lienhard, J.; Herrmann, T.; Preußner, J.; Hartrott, P.v.; Friedmann, V.; Wessel, A.; Thomas, A.; et al. MaterialDigital Dataset. 2023. Available online : <https://fordatis.fraunhofer.de/handle/fordatis/323> (accessed on 9 December 2024). [CrossRef]
34. Schweizer, C.; Reichenbach, R.; Butz, A.; Lienhard, J.; Herrmann, T.; Preußner, J.; Hartrott, P.v.; Friedmann, V.; Wessel, A.; Thomas, A.; et al.; BWMD Ontology. 2022. Available online: [https://gitlab.cc-asp.fraunhofer.de/gf7\\_public/iwm-gdtool/-/blob/c08de6d4c7bdcd4c4cf0c29fa3d7a99302ce8a2b/BWMD\\_legacy/BWMD\\_Ontologie\\_2020-08-12.owl](https://gitlab.cc-asp.fraunhofer.de/gf7_public/iwm-gdtool/-/blob/c08de6d4c7bdcd4c4cf0c29fa3d7a99302ce8a2b/BWMD_legacy/BWMD_Ontologie_2020-08-12.owl) (accessed on 9 December 2024).
35. Information Artifact Ontology (IAO). 2024. Available online: <https://github.com/information-artifact-ontology/IAO/> (accessed on 23 October 2024).
36. Garcia Trelles, E.; Hartrott, P.v. DIGITAL-WORKFLOW-STEPS-USING-THE-GRAPH-DESIGNER-TOOL. 2022. Available online: [https://gitlab.cc-asp.fraunhofer.de/gf7\\_public/iwm-gdtool/-/wikis/DIGITAL-WORKFLOW-STEPS-USING-THE-GRAPH-DESIGNER-TOOL](https://gitlab.cc-asp.fraunhofer.de/gf7_public/iwm-gdtool/-/wikis/DIGITAL-WORKFLOW-STEPS-USING-THE-GRAPH-DESIGNER-TOOL) (accessed on 9 December 2024).
37. Stanford Center for Biomedical Informatics Research. Protégé. 2021. Available online: <https://protege.stanford.edu/> (accessed on 5 December 2024).
38. Inforapid KnowledgeBase Builder. 2021. Available online: <https://www.inforapid.com/> (accessed on 5 December 2024).
39. Ontotext. GraphDB. Available online: <https://graphdb.ontotext.com/documentation/10.2/> (accessed on 5 December 2024).
40. Janka-Ramm, M.; Garcia Trelles, E. Interview Template for Domain Knowledge Extraction. 2023. Available online: [https://gitlab.cc-asp.fraunhofer.de/gf7\\_public/iwm-gdtool/-/blob/5e9b251b8cc0e14a3048af7d8bf5da78b79c2ffc/BWMD\\_legacy/Interview\\_format\\_v3.xlsx](https://gitlab.cc-asp.fraunhofer.de/gf7_public/iwm-gdtool/-/blob/5e9b251b8cc0e14a3048af7d8bf5da78b79c2ffc/BWMD_legacy/Interview_format_v3.xlsx) (accessed on 2 June 2023).
41. Aryan, P.R. SparqlGraphViz: Visualize SPARQL Query Results with Graphviz. GitHub Repository. 2021. Available online: <https://github.com/pebbie/sparqlgpviz> (accessed on 9 December 2024).
42. Ontotext. Reasoning. 2023. Available online: <https://graphdb.ontotext.com/documentation/10.7/reasoning.html> (accessed on 3 October 2023).
43. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. 2011. Available online: <https://jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf> (accessed on 9 December 2024).
44. Cheung, K.; Drennan, J.; Hunter, J. Towards an Ontology for Data-driven Discovery of New Materials. In Proceedings of the AAAI Spring Symposium: Semantic Scientific Knowledge Integration, Stanford, CA, USA, 26–28 March 2008 ; pp. 9–14.
45. Valledor, P.; Berrueta, D.; Rubiera, E.; Diaz, D.; Polo, L. Ontology for the Steel Domain. 2010. Available online: <https://ontorule-project.eu/resources/steel.owl> (accessed on 9 December 2024).
46. Alam, M.; Dittmann, F.; Niebel, M.; Lehmann, J.; Dessi, D.; Morgado, J.F.; von Hartrott, P.; Eberl, C.; Gumbsch, P.; Sack, H. Towards digitizing physical entities in materials science. In *FIZ Karlsruhe, White Paper*; 2020. Available online:

- <https://www.fiz-karlsruhe.de/sites/default/files/FIZ/Dokumente/Forschung/ISE/White-Paper/2020-Alam-Dessi-Sack-Towards-Digitizing-Physical-Entities-in-Materials-WhitePaper.pdf> (accessed on 9 December 2024).
47. Schilling, M.; Bayerlein, B.; von Hartrott, P.; Waitelonis, J.; Birkholz, H.; Portella, P.D.; Skrotzki, B. FAIR and Structured Data: A Domain Ontology Aligned with Standard-Compliant Tensile Testing. *Adv. Eng. Mater.* **2024**, 2400138. [CrossRef]
  48. Garabedian, N.T.; Schreiber, P.J.; Brandt, N.; Zschumme, P.; Blatter, I.L.; Dollmann, A.; Haug, C.; Kümmel, D.; Li, Y.; Meyer, F.; et al. Generating FAIR research data in experimental tribology. *Sci. Data* **2022**, 9, 315. [CrossRef]
  49. Bayerlein, B. Towards Interoperability: Digital Representation of a Material Specific Characterization Method 2022. Available online: <https://dgm.inventum.de/app/medialibrary/show/eventItem/d0ab5001-7aed-451a-a0aa-0e68e2f58630>.
  50. Chen, Y.; Schilling, M.; von Hartrott, P.; Nasrabadi, H.B.; Skrotzki, B.; Olbricht, J. Ontopanel: A Tool for Domain Experts Facilitating Visual Ontology Development and Mapping for FAIR Data Sharing in Materials Testing. *Integr. Mater. Manuf. Innov.* **2022**, 11, 545–556. [CrossRef]
  51. Bayerlein, B.; Schilling, M.; Birkholz, H.; Jung, M.; Waitelonis, J.; Mädler, L.; Sack, H. PMD Core Ontology: Achieving semantic interoperability in materials science. *Mater. Des.* **2024**, 237, 112603. [CrossRef]
  52. ISO/IEC 19763-10:2023; Information Technology—Metamodel Framework for Interoperability (MFI)—Part 10: Core Model and Basic Mapping. Technical Committee: ISO/IEC JTC 1/ISO/IEC JTC 1/SC 32 Data Management and Interchange. International Organization for Standardization, International Electrotechnical Commission: Geneva, Switzerland, 2023; p. 24.
  53. The BWMD Ontology. 2021. Available online: [https://gitlab.cc-asp.fraunhofer.de/EMI\\_datamanagement/bwmd\\_ontology](https://gitlab.cc-asp.fraunhofer.de/EMI_datamanagement/bwmd_ontology) (accessed on 2 June 2023).
  54. Materials Open Laboratory MatPortal. Available online: <https://matportal.org/ontologies/BWMD-DOMAIN> (accessed on 30 November 2024).
  55. Huschka, M.; Dlugosch, M.; Friedmann, V.; Trelles, E.G.; Hoschke, K.; Klotz, U.; Patil, S.; Preußner, J.; Schweizer, C.; Tiberto, D. Mat-O-Lab/AluTrace-Data-and-Documentation: V1.0.3. 2023. Available online: <https://zenodo.org/records/7990266> (accessed on 9 December 2024). [CrossRef]
  56. Platform Material Digital. Available online: <https://www.materialdigital.de/> (accessed on 9 December 2024).
  57. PROV-O: The PROV Ontology. 2024. Available online: <https://www.w3.org/TR/prov-o/> (accessed on 3 December 2024).
  58. MatPortal : The Ontology Repository for Materials Science. 2024. Available online: <https://matportal.org/ontologies/PMDCO?p=summary> (accessed on 3 December 2024).
  59. The Common Core Ontologies (CCO). 2024. Available online: <https://github.com/CommonCoreOntology/CommonCoreOntologies> (accessed on 3 December 2024).
  60. MatPortal : The Ontology Repository for Materials Science. 2024. Available online: <https://matportal.org/ontologies/MSEO/?p=summary> (accessed on 3 December 2024).
  61. Fliegenger, S.; Rosenberger, J.; Luke, M.; Domínguez, J.M.; Francisco Morgado, J.; Kobialka, H.U.; Kraft, T.; Tlatlik, J. Digital Methods for the Fatigue Assessment of Engineering Steels. *Adv. Eng. Mater.* **2024**, 2400992. [CrossRef]
  62. OpenRefine. 2022. Available online: <https://openrefine.org/> (accessed on 15 June 2023).
  63. TARQL. 2019. Available online: <https://tarql.github.io/> (accessed on 15 June 2023).
  64. IWM-GDTool. 2022. Available online: [https://gitlab.cc-asp.fraunhofer.de/gf7\\_public/iwm-gdtool](https://gitlab.cc-asp.fraunhofer.de/gf7_public/iwm-gdtool) (accessed on 15 June 2023).
  65. Bast, H.; Buchhold, B. Qlever: A query engine for efficient sparql+ text search. In Proceedings of the 2017 ACM Conference on Information and Knowledge Management, Singapore, 6–10 November 2017; pp. 647–656.
  66. Zhu, M.; Chen, W.; Hu, Y.; Hou, Y.; Liu, L.; Zhang, K. DRGraph: An Efficient Graph Layout Algorithm for Large-scale Graphs by Dimensionality Reduction. *IEEE Trans. Vis. Comput. Graph.* **2020**, 27, 1666–1676. [CrossRef] [PubMed]
  67. Moxon, S.A.; Solbrig, H.; Unni, D.R.; Jiao, D.; Bruskiwich, R.M.; Ballhoff, J.P.; Vaidya, G.; Duncan, W.D.; Hegde, H.; Miller, M.; et al. The Linked Data Modeling Language (LinkML): A General-Purpose Data Modeling Framework Grounded in Machine-Readable Semantics. *Proc. Int. Conf. Biomed. Ontol. (ICBO)* **2021**, 3073, 148–151.
  68. Huschka, M.; Dlugosch, M.; Friedmann, V.; Trelles, E.G.; Hoschke, K.; Klotz, U.E.; Patil, S.; Preußner, J.; Schweizer, C.; Tiberto, D. Position Paper The “AluTrace” Use Case: Harnessing Lightweight Design Potentials via the Materials Data Space®. In Proceedings of the TRUSTS—Trusted Secure Data Sharing Space Workshop: Data Spaces & Semantic Interoperability, Vienna, Austria, 3 June 2022.
  69. H2DIGITAL: Digitale Abbildung Einer Grünen Wasserstoffwirtschaft. 2022. Available online: <https://www.iwm.fraunhofer.de/de/warum-fraunhofer-iwm/forschungsprojekte-des-fraunhofer-iwm/h2digital.html> (accessed on 25 September 2024).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.