Advanced Engineering Materials

ADVANCED ENGINEERING MATERIALS
www.aem-journal.com

RESEARCH ARTICLE OPEN ACCESS

The PRIMA Thesaurus for Materials Science and Engineering

Rossella Aversa[1] | Alexey Boubnov[2] | Dario De Angelis[3] | Catriona Eschke[4] | Maximilian Kabbe[5] | Felix Kraus[1] | Mirco Panighel[3,6] | Richard Thelen[5] | Davide Valentinis[7]

[1]Scientific Computing Center, Karlsruhe Institute of Technology, Karlsruhe, Germany | [2]Institute of Nanotechnology, Karlsruhe Institute of Technology, Karlsruhe, Germany | [3]CNR – Istituto Officina dei Materiali (IOM), Trieste, Italy | [4]Institute of Metallic Biomaterials, Helmholtz-Zentrum Hereon, Geesthacht, Germany | [5]Institute of Microstructure Technology, Karlsruhe Institute of Technology, Karlsruhe, Germany | [6]Department of Physics and Materials Science, University of Luxembourg, Luxembourg City, Luxembourg | [7]Institute for Quantum Materials and Technologies, and Institute for Theory of Condensed Matter, Karlsruhe Institute of Technology, Karlsruhe, Germany

Correspondence: Rossella Aversa (rossella.aversa@kit.edu)

Received: 29 November 2025 | **Revised:** 1 July 2026 | **Accepted:** 2 July 2026

Keywords: findable, accessible, interoperable, and reusable (FAIR) vocabulary | materials science | materials science and engineering (MSE) | thesaurus

ABSTRACT

Materials science and engineering (MSE) is characterized by heterogeneous workflows, often coupled with limited research data management (RDM) practices. In particular, provenance metadata are often confined within individual electronic laboratory notebooks (ELNs), highlighting the need for semantic resources that support findable, accessible, interoperable, and reusable (FAIR) data, while remaining accessible to nonexperts. The presented Provenance information for materials science (PRIMA) Thesaurus addresses this challenge by providing a structured vocabulary for describing experimental and computational workflows, without relying on complex conceptual models or formal axiomatization. Its development, based on an iterative process involving domain experts, includes requirement analysis across multiple techniques, selection and harmonization of concepts, and alignment with existing community standards. The resulting terminology, structured in a few hierarchical layers, is implemented in Simple Knowledge Organization System (SKOS) to ensure flexibility and ease of integration. The applicability of PRIMA as a versatile resource capable of serving diverse scientific disciplines is demonstrated through use cases spanning theoretical and experimental condensed-matter physics, metrology, surface science, and metallic biomaterial research. These examples illustrate the integration of concepts into various ELNs and metadata schemas, showing how this shared semantic layer supports consistent data description, improves discoverability, and enables cross-platform interoperability.

1 | Introduction

Materials science and engineering (MSE) is an extremely vast and highly interdisciplinary field, in which scientists often rely on large-scale research infrastructures, usually combining multiple data, modalities, and techniques, to investigate complex systems and to enable the discovery of new materials [1, 2]. The massive amount of data produced by these facilities, combined with the advent of artificial intelligence (AI), provides the foundation for efficient data-driven research, e.g., to derive aggregated results based on large-scale statistical analyses or to train new AI models [3]. At the same time, however, several factors that hinder the reproducibility of experiments [4] remain prevalent, including inadequate provenance metadata and a multitude of (often incompatible) solutions and tools, due to the huge variety of experimental, computational, and analytical methods in this field [3, 5, 6]. Addressing these challenges requires a systematic approach to research data management (RDM) [3].

It is nowadays widely acknowledged that to fully exploit its potential, data should be findable, accessible, interoperable, and reusable (FAIR) [7] for both humans and machines. While all four principles are essential, the last two are particularly critical in this context, as they directly impact the integration, interpretation, and effective reuse of data across heterogeneous systems and communities [5]. These aspects heavily rely on shared semantics. In particular, FAIR vocabularies are essential to reach interoperability [8] and to ensure that data can be exchanged across multiple domains [5], even when metadata perspectives, purposes, or formats differ.

Recognizing the paramount importance of semantic interoperability in advancing the digital transformation in MSE [3, 5, 9], numerous organizations, from international initiatives to individual research groups, developed mid-level, domain-level, or application-level ontologies [10], vocabularies [11–13], and metadata schemas [14, 15]. However, a recent analysis found that a significant proportion of the available semantic artifacts are not FAIR [10]. Moreover, their development and adoption face significant challenges: on the one hand, a universally adopted, one-size-fits-all standard for metadata representation is difficult to achieve [6]; on the other hand, there is considerable resistance to semantic technologies in MSE due to the limited user expertise and the resulting perceived complexity [9, 16].

In this paper, we aim to address three needs arising from the scenario outlined above: a harmonized representation of mid-level provenance applicable across multiple techniques and communities, a FAIR vocabulary, and a simple solution that lowers the barrier for those unfamiliar with formal ontological frameworks.

To this end, we introduce the Provenance information for materials science (PRIMA) Thesaurus, a FAIR vocabulary to describe the mid-level provenance information of physical and digital entities collected in experimental and computational studies (as first defined by Matthews et al. [17] and adopted by Bunakov et al. [11]) and then archived for further data discovery and sharing. It is designed to support the development of metadata schemas and templates in RDM systems, thereby enhancing data reusability, interoperability, and exchange, and enabling the implementation of filtering and searching functionalities based on shared mid-level terms across data resources, ultimately improving their findability.

The PRIMA Thesaurus, hereafter referred to as PRIMA, is the result of the Metadata Working Group (WG) [18], a collaboration between the Joint Lab “Integrated Model and Data Driven Materials Characterization” (JL MDMC) [19] of the Helmholtz Association, the NFFA-Europe research infrastructure (Nanoscience Foundries and Fine Analysis) [20], and the consortium NFDI-MatWerk (Nationale Forschungsdateninfrastruktur für Materialwissenschaft & Werkstofftechnik) [21] of the German National Research Data Infrastructure.

The paper is organized as follows: an overview of the main existing work in the field is presented in Section 2, while the adopted development methodology is detailed in Section 3. PRIMA is presented in Section 4, with a particular focus on the graphical representation of the concepts and their relations, as well as on its evaluation. To illustrate the flexible use of PRIMA, Section 5 provides an overview of its current applications in different MSE research areas, which are discussed in Section 6 together with future perspectives. Finally, Section 7 summarizes the results and provides the conclusions.

To ease their identification, the concepts of PRIMA are written in bold with capital initial letters throughout the text.

2 | Existing Work in the Field

The broadly adopted high-level schemas, such as Crossref [22], Dublin Core Metadata Initiative (DCMI) metadata terms [23], DataCite [24], and schema.org [25] provide generic, domain-independent descriptions of research outputs. By focusing on common metadata elements, they support interoperability between heterogeneous systems, but are not intended to model in-depth domain knowledge. On the other hand, numerous semantic artifacts have been proposed in the literature to represent the MSE domain, differing in both their level of detail and intended purpose.

At the highest level of abstraction, the Elementary Multiperspective Material Ontology (EMMO) [26] provides a common semantic foundation for representing materials, processes, models, and simulations across physics, chemistry, and materials science. While facilitating interoperability, it often requires domain-specific extensions to adequately represent specialized concepts. The Provenance Ontology PROV-O [27] is widely used as a standard framework for representing provenance information in a domain-independent way. While highly expressive for provenance tracking, it typically requires domain-specific extensions, e.g., the REPRODUCE-ME Ontology [28] or the Platform Material Digital Core Ontology (PMDco) [5], to model scientific workflows in detail.

In addition to ontologies capturing generic concepts related to experimental design and methodology [29], various initiatives developed mid-level vocabularies to describe experimental and computational workflows or entities. For example, the CODATA Research Data Management Terminology (RDMT) Working Group collected and maintained terms and definitions for concepts relevant to RDM [13]; the International Materials Resource Registries (IMRR) Working Group of the Research Data Alliance (RDA) developed the Materials Resource Registry Vocabulary (NMRRVOCAB) [12] for global discovery of MSE data resources; NFFA-Europe designed a Common Vocabulary [11] to describe the lifecycle of data collected in physical and computational nanoscience experiments, partially building upon the Core Scientific Metadata Model (CSMD) [17, 30]; within NFDI, it is relevant citing the MatWerk Ontology (MWO) [31] and the Metadata4Ing ontology [32], respectively developed by the consortia NFDI-MatWerk and NFDI4ING to represent research data and related activitied in materials science and in engineering sciences; finally, the Physics Subject Headings (PhySH) [33] of the American Physical Society is a classification scheme intended to cover all disciplines of physics for browsing and searching functionalities.

At a finer degree of semantic granularity, several domain-level vocabularies have been proposed to model specific aspects and sub-domains within MSE. For instance, the Materials Design

Ontology (MDO) [34] was the first ontology representing solid-state physics concepts including the provenance of calculations; in the field of computational materials science, Novel Materials Discovery (NOMAD) Metainfo [35] is a FAIR metadata schema for describing and managing atomistic calculations; the CODATA-VAMAS Working Group on the Description of Nanomaterials proposed the Uniform Description System for Materials on the Nanoscale (UDS) [36] focused on nanosamples description, while the NanoMine ontology [37] describes processing, structure, and properties of nanomaterials.

As semantic artifacts evolve from the top level to the application level, they gain expressive power for modeling domain-specific concepts at the expense of interoperability. A detailed overview and evaluation of top- to application-level ontologies used in MSE was recently presented by Norouzi et al. [10].

3 | Methodology

The process of developing the PRIMA Thesaurus was initiated in 2021 by the Metadata WG [18], composed of members of JL MDMC [19] and NFFA-Europe Pilot (NEP) [20], and subsequently expanded in 2022 to include NFDI-MatWerk [21]. Such cross-community synergies have proven essential for fostering communication between scientists and developers, facilitating consensus on shared solutions aligned with MSE applications [1]. The methodology adopted for the development of PRIMA [38] can be divided into four parts, detailed in this section.

3.1 | Identification of the Requirements

The Metadata WG was established as an interdisciplinary and voluntary initiative bringing together a core team of 12 participants, complemented by invited guests and consultants who provided feedback on multiple occasions. Their combined expertise spanned 16 experimental and 2 computational techniques: atomic force microscopy (AFM), vertical scanning interferometry (VSI), confocal scanning microscopy (CSM), focused ion beam (FIB), scanning electron microscopy (SEM), transmission electron microscopy (TEM), scanning tunneling microscopy (STM), nano-computed tomography (nanoCT), optical microscopy (OM), optical spectrometry (OS), magnetic resonance imaging (MRI), nuclear magnetic resonance (NMR), low-energy electron diffraction (LEED), X-ray photoelectron spectroscopy (XPS), near-edge X-ray absorption fine structure (NEXAFS), angle-resolved photoemission spectroscopy (ARPES), density functional theory (DFT), and many-body calculations.

Building on this broad range of techniques, we gathered information from the group members on typical scientific workflows and RDM practices in their domains to formulate shared requirements [1, 5, 11]. Their collective experience identified three recurring needs: a unified terminology to ensure consistency and facilitate cross-disciplinary collaborations, extensible metadata templates compatible with both experimental and computational workflows to enhance interoperability and ease data exchange, and a simple semantic structure with minimal hierarchical depth to reduce the learning curve for users. The last requirement, consistent with the limited ontology expertise in MSE mentioned in Section 1, motivated our design choices described in Section 3.3, in alignment with similar works [11, 16].

3.2 | Selection of Concepts and Definitions

The potentially relevant concepts were collected by combining two approaches [38]: input from WG members and their project communities, and reuse of existing work (Section 2). Input from the MSE users was collected in tabular form, with relevant concepts listed for each technique and alternative definitions provided for each concept. The definitions were based as much as possible on available citations from books, manuals, or existing vocabularies, and further elaborated and adapted during the discussions, without any use of AI. When required, concepts and definitions were adapted to reflect the needs and use cases of the involved communities. Iterative loops of discussion were conducted to examine and refine existing work in the field.

When surveying the existing work in the field (Section 2), we first evaluated well-established, domain-independent schemas, such as Crossref [22], DCMI metadata terms [23], DataCite [24], schema.org [25], and PROV-O [27]. As expected, although valuable for general metadata representation, these schemas proved to be too high-level for our purposes. For the selection and description of relevant concepts, we found particularly appropriate the mid-level NFFA Common Vocabulary [11] and the MWO [31], as well as the CODATA RDMT [13], the NMRRVOCAB [12], the EMMO [26], and the Metadata4Ing ontology [32]. For some terms, especially the ones related to the computational workflow, we consulted Wikipedia [39], the MDO [34], and the PhySH [33].

Iterative rounds of discussion within the WG were conducted to examine the collected concepts, reach consensus on their inclusion, and refine their definitions. As a selection criterion, concepts were retained if they were relevant for describing the workflows of at least 80% of all techniques (15 out of 18). In cases where specific variants of a concept (e.g., **Data Acquisition**) pertained exclusively to experimental or computational studies (e.g., **Measurement** or **Simulation**, respectively), they were required to be relevant for at least 80% of the corresponding techniques (13 experimental and 2 computational).

3.3 | Development of the PRIMA Thesaurus

The collaborative sessions described in Section 3.2, focused on consensus building and decision-making, resulted in a plain-text glossary of 45 concepts, known as the MDMC-NEP Glossary of Terms. Feedback from the broader scientific community was encouraged by publishing it on Zenodo [40] and on the project websites [19, 20]. The results presented in this paper build on such additional insights gained from scientists, technicians, and software developers, and include a revision of concepts and an improved consistency in the definitions and in the hierarchical structure.

Taking into account the requirement for a simple semantic structure (Section 3.1), a thesaurus was chosen as the most appropriate representation. A thesaurus is defined as ≪a controlled vocabulary arranged in a known order and structured so that the various relationships among terms are displayed clearly and identified by standardized relationship indicators [...] employed reciprocally≫ [41]. Compared to ontologies, thesauri provide a simpler and more lightweight semantic structure that is easier to design, maintain, and apply, while still enabling effective information organization, retrieval, and interoperability. Their use of standardized hierarchical and associative relationships supports an adequate knowledge representation without requiring the formal complexity of fully specified semantic relations [42].

We modeled PRIMA in Simple Knowledge Organization System (SKOS) [43], a linked data standard recommended by the World Wide Web Consortium (W3C), developed to represent in an interoperable way the structure and content of Knowledge Organization Systems (KOSs) such as thesauri, taxonomies, and classification schemes. SKOS provides an extension vocabulary in Resource Description Framework (RDF) specifically designed for describing concept schemes, enabling their referencing, sharing, and reuse in semantic web applications [43]. We selected SKOS because it is intended for lightweight concept hierarchies and offers structural semantics with a lower entry barrier for users compared to the full logical expressiveness of ontologies, typically represented in the Web Ontology Language (OWL).

SKOS allows concepts to be annotated using generic notes (e.g., `skos:note`) or more specific ones (e.g., `skos:scopeNote` or `skos:historyNote`), and to be mapped to concepts in other schemes [43]. Following the principle of minimal ontological commitment [44], which underlies the design of SKOS, the modeling of PRIMA was deliberately kept simple, avoiding the use of transitive relations like `skos:broaderTransitive` and `skos:narrowerTransitive` [42]. Consistent with this approach, the mappings to other concept schemes were modeled using `skos:closeMatch`, a nontransitive property, rather than `skos:exactMatch`. After each iteration involving the addition or removal of concepts or revisions to their definitions, the Thesaurus was automatically validated using Skosify [45] to ensure overall consistency.

As an illustrative example of our modeling process, we discuss the concept of **Research User**, which was interpreted as the equivalent of an agent in PROV-O [27]. Thus, `skos:related` was used to encompass `prov:wasAttributedTo`, `prov:wasGeneratedBy`, and `prov:wasAssociatedWith`. We considered it as a `skos:closeMatch` of DataCite's `Contributor` [24], and, correspondingly, we mapped **User Role** to a selection of items from the `contributorType` [24] controlled list, completed by the custom **Instrument Scientist**. To date, DataCite has not published an official SKOS or RDF release of its version 4.6 with resolvable uniform resource identifiers (URIs); thus, we resorted to the version 4.4 URIs [46] and indicated the alignment to version 4.6 in the `skos:note`. The impact of this discrepancy is negligible, because the controlled lists in the two versions differ only for the `Translator`, new in 4.6 [24], which is not part of the subset we selected.

Throughout the development process, multiple tools were adopted to lower the barrier to collaborative work and facilitate discussions. Early work relied on shared online tables and textual documents, while in later stages the Editor for Vocabularies to Know Semantics (EVOKS) [47] was employed to collaboratively and easily create concepts in SKOS, and JSONforms [48] was used to support the entry of terms, the rendering of custom JSON schemas, and the visualization of controlled lists.

3.4 | Maintenance and Curation

PRIMA is conceived as a dynamic resource that requires continuous maintenance to remain up to date and aligned with its intended scope. Rather than envisioning major periodic releases, its evolution follows an incremental curation model focused on the regular refinement of individual concepts or the inclusion of new terms reflecting the evolving needs of the community. As mentioned in Section 3.3, this approach led to the evolution of the MDMC-NEP Glossary [40] into the PRIMA Thesaurus, supporting progressive adaptation while preserving its overall structure.

Community engagement plays a central role in the curation process. For this reason, the Metadata WG organizes monthly meetings to provide a structured forum for collecting user feedback and consolidating updates.

4 | Results

In this section, we describe the PRIMA Thesaurus, focusing on different aspects: the references to access it, the relations among its key concepts including a graphical representation, and its evaluation against some recent guiding principles for semantic resources.

4.1 | The PRIMA Thesaurus

The current version of the PRIMA Thesaurus, provided as Supporting Information to this paper, contains a total of 54 `skos:Concepts`, consisting of 42 main concepts and an additional controlled list of 12 **User Roles**. As a result of the development process described in Section 3, they were considered relevant and appropriate by the members of the Metadata WG [18] for describing and framing the mid-level provenance of processes performed and of data generated in a **Study**.

To support its accessibility as well as future extensions and updates, we published PRIMA in two instances of the SKOSMOS vocabulary service [49], hosted by NFDI-MatWerk [50] and by NFFA-Europe [51].

A unique persistent uniform resource locator (PURL), used as a URI, was assigned to the thesaurus [52] and to each concept via the Persistent Identifiers for Semantic Artifacts (PIDA) service [53]. This way, the URIs of the individual concepts can be dereferenced to retrieve both the machine-readable RDF representation and the human-readable description in the NFDI-MatWerk SKOSMOS instance [50], either via Linked Data and Representational State Transfer (REST) application programming interface (API) or through the graphical web interface. This setup facilitates the integration into external workflows and semantic applications, as demonstrated in Section 5.

4.2 | The Relations: Graphical Representation

The concepts of PRIMA are related to each other at different levels and can be grouped focusing on specific aspects. In the graphical representations presented in this section, the different thematic blocks are reported separately, to depict the relations among the relevant concepts with the due level of detail. For the sake of simplicity, only the primary pairs of related terms are connected by an arrow, i.e., the possibly existing transitive relations are not shown; for the same reason, when the arrow represents a one-directional relation (`skos:narrower`), the inverse one (`skos:broader`) is omitted. Complete details on the related terms are included in the definitions, which can be consulted in the Supporting Information or in the SKOSMOS instances [50, 51]. We remark that the formal adoption of a rigorous ontological framework goes beyond the purpose of this work.

4.2.1 | Overview

Figure 1 gives an overview of the PRIMA Thesaurus. A **Study** is performed by one or more **Research Users**, who have their respective **User Role**(s) in the **Project**. The **Study** includes one or more processes, among: **Model Preparation** or **Processing and Treatment** (in case of experimental or computational workflow, respectively), **Data Acquisition** (**Measurement** in case of experimental workflow, **Calculation** or **Simulation** in case of computational workflow), **Data Processing**, **Data Analysis**, and **Data Curation**. Each process, which consists of a set of actions, can be repeated multiple times. The results of a **Study** are reported in **Scientific Publications**, supported and complemented by the collected **Publication Data**.

4.2.2 | The System

According to the definition (see Supporting Information), a **System** can be a digital or a physical entity. An example of a digital **System** is the **Model**, which is the output of **Model Preparation**. A generic workflow to produce a **Model** is shown in Figure 2: **Research Data**, typically (but not limited to) **Reference Data** from previous **Studies** and/or from literature, after some kind of **Data Processing**, is used as input for the **Model Preparation**. The resulting output of the process, performed by one or more **Research Users**, possibly using **Research Software**, is the **Model**.

A physical **System**, instead, may be a **Material** or a **Sample** (which is defined to be constituted by at least one **Material**). A physical **System** could also be, e.g., a 2D electron gas constituting a part of the **Sample**, namely the electrons confined within a thin layer of the **Material**. As illustrated in Figure 3, a physical **System** usually undergoes some kind of **Processing and Treatment** to be prepared for further processes: one or more **Materials** can be treated to produce a **Sample**, or a **Sample** can be prepared for a **Data Acquisition**.

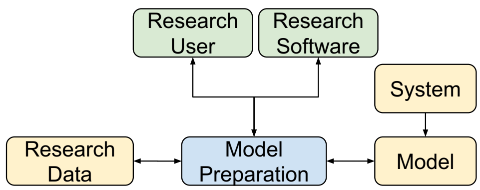

FIGURE 2 | Main concepts related to **Model Preparation**. Different colors are used to group concepts into three intuitive categories: concepts referring to objects (yellow), to agents (green), and to processes or actions (blue). This categorization is intended solely to facilitate visualization, and does not correspond to a formal ontological distinction. Bi-directional arrows represent `skos:related` and one-directional arrows represent `skos:narrower`. The reciprocal `skos:broader` are omitted from the figure for simplicity.

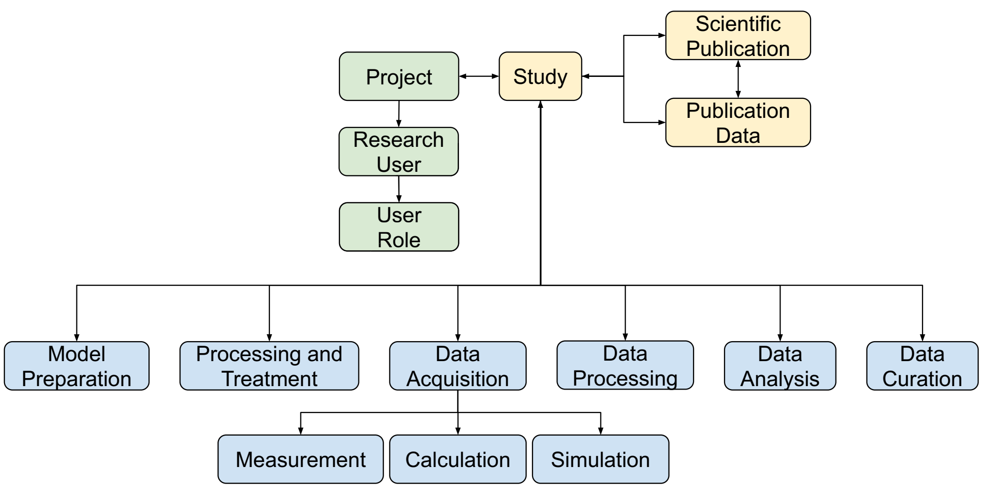

FIGURE 1 | Overview of the PRIMA Thesaurus. Different colors are used to group concepts into three intuitive categories: concepts referring to objects (yellow), to agents (green), and to processes or actions (blue). This categorization is intended solely to facilitate visualization, and does not correspond to a formal ontological distinction. Bi-directional arrows represent `skos:related` and one-directional arrows represent `skos:narrower`. The reciprocal `skos:broader` and most of the relations to **Research User** are omitted to avoid overcrowding the figure.

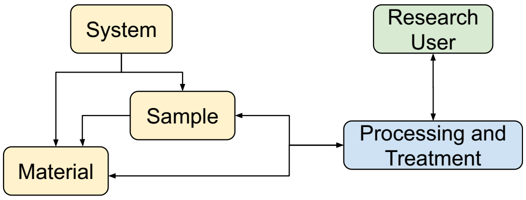

FIGURE 3 | **Processing and Treatment** for the physical **System**. Different colors are used to group concepts into three intuitive categories: concepts referring to objects (yellow), to agents (green), and to processes or actions (blue). This categorization is intended solely to facilitate visualization, and does not correspond to a formal ontological distinction. Bi-directional arrows represent `skos:related` and one-directional arrows represent `skos:narrower`. The reciprocal `skos:broader` are omitted from the figure for simplicity.

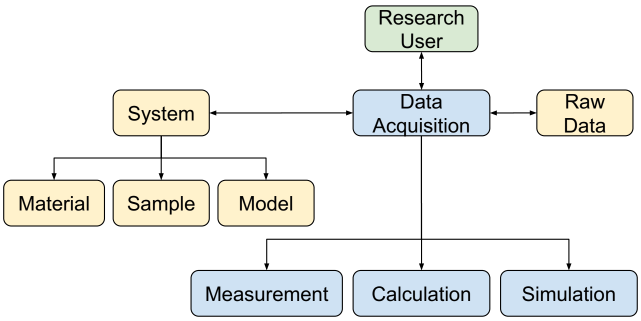

FIGURE 4 | **Data Acquisition** for both physical and digital **Systems**. Different colors are used to group concepts into three intuitive categories: concepts referring to objects (yellow), to agents (green), and to processes or actions (blue). This categorization is intended solely to facilitate visualization, and does not correspond to a formal ontological distinction. Bi-directional arrows represent `skos:related` and one-directional arrows represent `skos:narrower`. The reciprocal `skos:broader` and most of the relations to **Research User** are omitted to avoid overcrowding the figure.

In fact, a **System**, whether physical or digital, is the primary input of any **Data Acquisition**, which is referred to as **Measurement** in the case of an experimental workflow and **Calculation** or **Simulation** in the case of a computational workflow. By definition (see Supporting Information), the purpose of any **Data Acquisition** is to collect **Raw Data** about a **System**. Figure 4 illustrates how these concepts are related.

4.2.3 | *Data*

Figure 5 offers a schematic representation of the different types of data considered in PRIMA and how the concepts are related. **Research Data** is defined as any kind of existing or newly generated data which is analyzed or considered as a basis for reasoning, discussion, or calculation in a research context. According to this definition, **Raw Data**, **Processed Data**, and **Analyzed Data** generated in the course of a **Study**, but also **Reference Data** from previous **Studies** or from literature, are particular types of **Research Data**. Any **Research Data** may be described by **Metadata** and, together with it, may be aggregated into a **Dataset**. A **Dataset** may also consist of other **Datasets**.

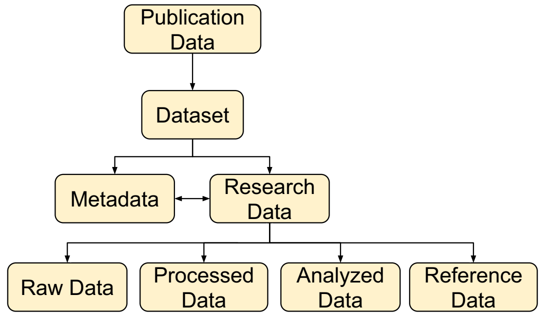

FIGURE 5 | Schematic representation of the different types of data considered in the PRIMA Thesaurus and their relations. Bi-directional arrows represent `skos:related` and one-directional arrows represent `skos:narrower`. The reciprocal `skos:broader` are omitted from the figure for simplicity.

Publication Data consists of one or more **Datasets** that, in addition to **Data Curation**, have undergone quality assessment and are considered scientifically reliable.

4.3 | Evaluation

We evaluated PRIMA against the three guiding principles to develop Computable Biomedical Knowledge (CBK) infrastructure [54]: FAIRness, trustworthiness, and openness (FAIR + TO). Originally conceptualized for the application in the health context, this approach can be easily extended to other fields.

4.3.1 | *FAIRness Assessment*

The PRIMA Thesaurus was created following the available recommendations to make semantic artifacts FAIR [8, 55–57]. Being mutually consistent, we referred to the most recently published FAIR Vocabulary Features (FVFs) [8] to perform the FAIRness assessment. We ascertained full compliance with all the FVFs, as reported in Table 1. Furthermore, we performed an additional assessment with the FAIR-Checker [58]. PRIMA passed all checks except one concerning identifiers. This warning arises because FAIR-Checker, developed mainly for life-science resources, expects PID namespaces specific to that domain. PRIMA instead uses stable, human-readable identifiers derived from truncated term labels. These IDs are unique and persistent, and therefore still meet the FAIR requirements even though they do not follow the namespace pattern assumed by the tool.

4.3.2 | *Trustworthiness Assessment*

Trustworthiness concerns aspects like information about sources, usage, or possible extensions, as well as ownership, publication, and distribution [54]. In the absence of suitable tools, a quantitative evaluation cannot be performed; nevertheless, we can qualitatively argue that the PRIMA Thesaurus complies with all these aspects, as detailed in the following.

We used DCMI metadata terms [23] for PRIMA descriptive and administrative metadata to provide information about, e.g., creator (`dc:creator`), creation date (`dct:created`), license (`dct:license`), and rights holder (`dct:rightsHolder`).

TABLE 1 | FAIRness assessment results of the PRIMA Thesaurus.

FAIR vocabulary feature (FVF)	Compliance
FVF-1: Vocabulary and their terms are assigned globally unique and persistent identifiers	Full
FVF-2: Vocabularies and their terms have rich metadata	Full
FVF-3: Vocabularies and their terms can be accessed using the identifiers, preferably by both humans and machines	Full
FVF-4: Vocabularies and their terms are registered or indexed in a searchable engine	Full
FVF-5: Vocabularies and their terms are retrievable using a standardized (preferably open, free and universally implementable) communication protocol and allow for authentication and authorization, where necessary	Full
FVF-6: Vocabularies and their terms are persistent over time and are appropriately versioned	Full
FVF-7: Vocabularies and their terms use a formal, accessible and broadly applicable, and preferably machine-understandable language for knowledge representation	Full
FVF-8: Vocabularies and terms use qualified references to other vocabularies	Full
FVF-9: Vocabularies and terms are described with a plurality of accurate and relevant attributes	Full
FVF-10: Vocabularies are released with a standard (preferably machine-readable) data usage license	Full
FVF-11: Vocabularies meet domain-relevant community standards	Full

Following the recommendations by Cox et al. [55], we did not embed any version information in the vocabulary path or in the concept URIs. Rather, we included a per-concept lightweight provenance and versioning information using properties from DCMI metadata terms [23] (e.g., `dct:modified`, `dct:isReplacedBy`, `dct:replaces`), OWL [59] (`owl:deprecated`, `owl:priorVersion`), and SKOS [60] (`skos:changeNote`, `skos:historyNote`). Moreover, for each definition adapted or adopted from an existing source, we reported the attribution of the original reference.

To ease the adoption of PRIMA and support its further development and extension to other domains, descriptions and documentation are provided online [18, 38, 40]. Finally, in addition to its publication in SKOSMOS [50, 51] for web searching and browsing, we registered PRIMA [61] in the Basic Register of Thesauri, Ontologies & Classifications (BARTOC) [62] for improved discoverability.

4.3.3 | Openness Assessment

The final principle of the FAIR + TO approach [54] is openness. While this aspect extends beyond the notion of accessibility (see FVF-5 in Section 4.3.1), the two concepts address distinct but complementary requirements: accessibility ensures that data and knowledge can be retrieved and used under well-defined conditions, whereas openness promotes their availability with minimal restrictions, supporting broader dissemination and reuse. In this context, the authors emphasize that knowledge should be ≪as open as possible and only closed as necessary≫ [54], in full alignment with FAIR principles [7].

The PRIMA Thesaurus fulfills both accessibility and openness: it can be accessed using the Hypertext Transfer Protocol Secure (HTTPS) protocol, without user identification or access restriction. Moreover, it is released under the Creative Commons CC BY [63] license, which also applies to the individual concepts to promote their broad reuse.

5 | Applications

All interviewed users considered semantic interoperability essential for enabling the interpretation and exchange of data across heterogeneous systems and platforms, regardless of format (Section 3.1). However, the use of shared semantics and RDM practices was found to vary considerably across the surveyed community, with some users employing electronic laboratory notebooks (ELNs), while others had not yet adopted consistent solutions. For the latter group, the introduction of well-defined metadata schemas in the workflow represents a crucial step forward compared to locally storing data with minimal, if any, documentation.

This section presents two applications of the PRIMA Thesaurus in line with the described ecosystem: first, its integration into ELNs, where the implementations vary depending on the underlying technologies and platform design, and then its use to support the development and harmonization of metadata schemas, in the absence of a dedicated RDM platform.

The use cases described for each application were proposed by interested members of the Metadata WG to improve semantic interoperability in their scientific practices. The impact of the implemented solutions was qualitatively assessed by comparing the situations before and after the introduction of PRIMA. A rigorous analysis of the impact of PRIMA goes beyond the purpose of this paper and will be approached in future work.

5.1 | Integration into ELNs

The adoption of FAIR vocabularies in RDM systems such as ELNs and laboratory information management systems (LIMSs) is essential to achieve semantic interoperability. A standardized terminology ensures consistency in describing workflows, reduces ambiguity, and ultimately supports reproducibility. It also facilitates the search and filtering of relevant information related to the performed activities and to the collected research data [64].

In this context, ELNs play a key role in integrating FAIR vocabularies to generate semantically annotated forms which help to structure information and to plan and design new experiments, eventually reducing the number of attempts, thereby ensuring a more efficient research activity and minimizing the risk of misinterpretation or loss of meaning [3, 9]. With FAIR vocabularies integrated, ELNs would be not only tools for data and metadata capture, but also for enforcing input consistency in accordance with community or domain standards [65]. This is particularly relevant when recording experimental parameters or associating research users with their institution and their role in conducting a study.

Beyond maintaining the accuracy of metadata to track the provenance within individual laboratories, globally unique identifiers assigned to terms enable data interpretation and exchange between teams and disciplines, even across different platforms and institutions [5]. In fact, by relying on URIs that point to the same underlying concepts, different ELNs can ensure consistent identification of entities, thereby facilitating data comparison.

Moreover, FAIR vocabularies support seamless data exchange with other systems often connected to ELNs and LIMSs, such as acquisition software or databases, enhancing machine actionability and enabling automation. This, in turn, reduces errors associated with manual data entry and accelerates structured access to research data, taking advantage of well-structured metadata to boost advanced data analysis approaches and machine learning algorithms [64].

The following use cases outline the application of PRIMA across three ELNs currently in use within the community (Kadi4Mat [66], eLabFTW [67], and Herbie [68]), illustrating a progression from early-stage adoption to more advanced integration.

5.1.1 | *Use Case: Kadi4Mat*

Kadi4Mat [66] is a generic and open-source virtual research environment combining the features of a data collaboration platform with those of an ELN. It is routinely used to record the settings of multiple metrology measurements carried out for instrumental quality assessment and calibration at the Institute of Microstructure Technology of the Karlsruhe Institute of Technology. Multiple surface metrology instruments are located in the laboratory, ranging from AFM, CSM, VSI to OM and SEM.

The high-quality data required in metrology are usually obtained by combining multiple techniques; for instance, interferometry and microscopy are combined in interferometric microscopy to improve resolution and vertical range [69]. While the specific parameters of each single piece of equipment are known from the manufacturer, broader concepts pertaining to the workflow are usually neglected. In fact, terms like **System**, **Experiment**, and **Material** are deliberately not defined in the International Vocabulary of Metrology [70]. The lack of shared higher-level semantics creates issues when workflows involving multiple techniques have to be designed or performed, or when the output data have to be interpreted. The terms of the PRIMA Thesaurus provide such a missing link.

As an illustrative example, three **Measurements** were performed on the same **Sample**, namely a calibration gauge from Bruker, using a Bruker AFM and two optical metrology **Instruments**: a Mahr CSM and a Bruker VSI. The resulting **Datasets** were managed within Kadi4Mat, where they were annotated and linked to the corresponding **Instruments** via JavaScript Object Notation for Linked Data (JSON-LD), and subsequently published on Zenodo [71–73]. Their structured representation, complemented by the semantic interoperability provided by PRIMA, facilitated the direct comparison of the **Datasets** and allowed the performance characteristics of the different **Techniques** to be determined. The results of this **Study** demonstrated the capability of all three **Instruments** to resolve µm-sized patterns with a height difference of 180 nm.

Currently, Kadi4Mat lacks support to link **Measurements** to the corresponding **Samples**. Such connection, particularly when information about sample preparation is included, would significantly enhance the value of the data, e.g., for the kind of iterative process development supported by AI. As an interim solution, the description of the Bruker calibration gauge [74] was registered in the NFDI-MatWerk Metadata Repository [75] and additionally included as a JSON file within each **Dataset** [71–73]. For this purpose, an extended metadata schema for experimental **Samples** [76] was adopted, providing a detailed description of both **Sample** characteristics and preparation steps.

In future developments of Kadi4Mat, the PRIMA concept URIs will be integrated using JSON-LD, and the **Sample** metadata schema will be adopted as a template that users can complete and link to the corresponding **Measurements**. This will further streamline the user experience by consolidating all the relevant features within a single service.

5.1.2 | *Use Case: eLabFTW*

eLabFTW [67] is a free and open-source web-based ELN which includes a laboratory inventory management system. While it does not formally define any semantic structure, it provides two high-level classes of entities which cannot be renamed, complemented by configurable templates which can be tailored to different workflows and research domains. This design ensures a basic level of structure and high flexibility; however, it may encourage the proliferation of ad hoc descriptions hindering semantic harmonization of datasets, particularly when they are exchanged across other platforms.

To address this limitation, a mapping between the eLabFTW terminology (formatted in italic) and the corresponding PRIMA concepts (in bold and capitalized) was established, and a set of standardized templates was developed accordingly.

The two high-level classes, for which templates can be defined, are *experiments* and *resources*. The *experiments*, representing the log of the laboratory activity, are mapped to PRIMA concepts referring to processes or actions, while the *resources*, associated with the digital representation of physical objects or abstract entities, are mapped to the concepts referring to objects (Section 4.2).

These classes belong to one or more *teams* of *users*, corresponding to **Projects** of **Research Users**.

In the described use case, *resource templates* were defined for **Sample**, **Instrument**, **Equipment**, **Consumable**, and **Study**, while *experiment templates* were defined for specific typologies of **Processing and Treatment** (namely sputtering, annealing, and evaporation) and **Measurement**, according to the specific **Technique** used (STM or XPS). One or more instances of each concept were then initialized from the corresponding templates.

With this setup, it was possible to uniformly represent surface science **Studies** performed at the "Instituto Officina dei Materiali" research institute of the Italian National Research Council (CNR-IOM) and at the University of Luxembourg. Figure 6 illustrates an example of a typical workflow: a **Sample** instance, containing relevant information about constituent **Material**(s), **Sample Holder**(s), and/or **Sample carrier**(s), is linked to the **Study** and to the instances of the individual **Processing and Treatments** it has undergone, where used **Equipment** or **Consumables** are tracked. The **Sample** is also linked to the instances of the **Measurements** carried out on it, including the **Technique**, the **Settings** of the **Instrument** and other **Equipment**, as well as the URL to the **Data Collaboration Platform** where acquired **Raw Data** is stored. Each **Measurement** is connected to the corresponding **Study** and **Instrument** instances. The latter contain the technical description of the specific **Instruments** and their **Laboratory** location.

It is worth noting that relations in eLabFTW are bidirectional, resulting in a hierarchical network of interlinked *experiments* and *resources*. The integration of PRIMA in this platform contributes to the unambiguous identification of the actions performed on a given sample using different techniques. This ultimately enables data comparison, which is particularly required, for example, in correlative characterization.

5.1.3 | *Use Case: Herbie*

Herbie [68] is an open-source semantic ELN originally tailored for metallic biomaterials research and subsequently applied to other fields of materials science, medicine, and coastal research. It is designed to enable structured, interoperable, and machine-readable documentation of experimental studies. To ensure rich and meaningful metadata as well as full provenance, Herbie provides interconnected web forms facilitating manual data entry for each experimental process or research entity in the workflow. Herbie generates the forms automatically from RDF-based semantic artifacts, such as ontologies and thesauri, with additional constraints specified in Shapes Constraint Language (SHACL).

By default a discipline-agnostic framework, Herbie does not provide its own semantic structure. Therefore, the RDF-based implementation of the PRIMA Thesaurus was applied as a domain-level laboratory setup for metallic biomaterials research at the Helmholtz-Zentrum Hereon. It was used to represent various consecutive **Techniques** applied for the **Processing and Treatment** of **Materials** and **Samples**, as well as **Measurements** and resulting **Raw Data**, including utilized **Consumables**, **Settings** of **Equipment**, and affiliated **Research Users** members of **Projects**. Figure 7 shows an example of a typical workflow modeled in Herbie, where magnesium

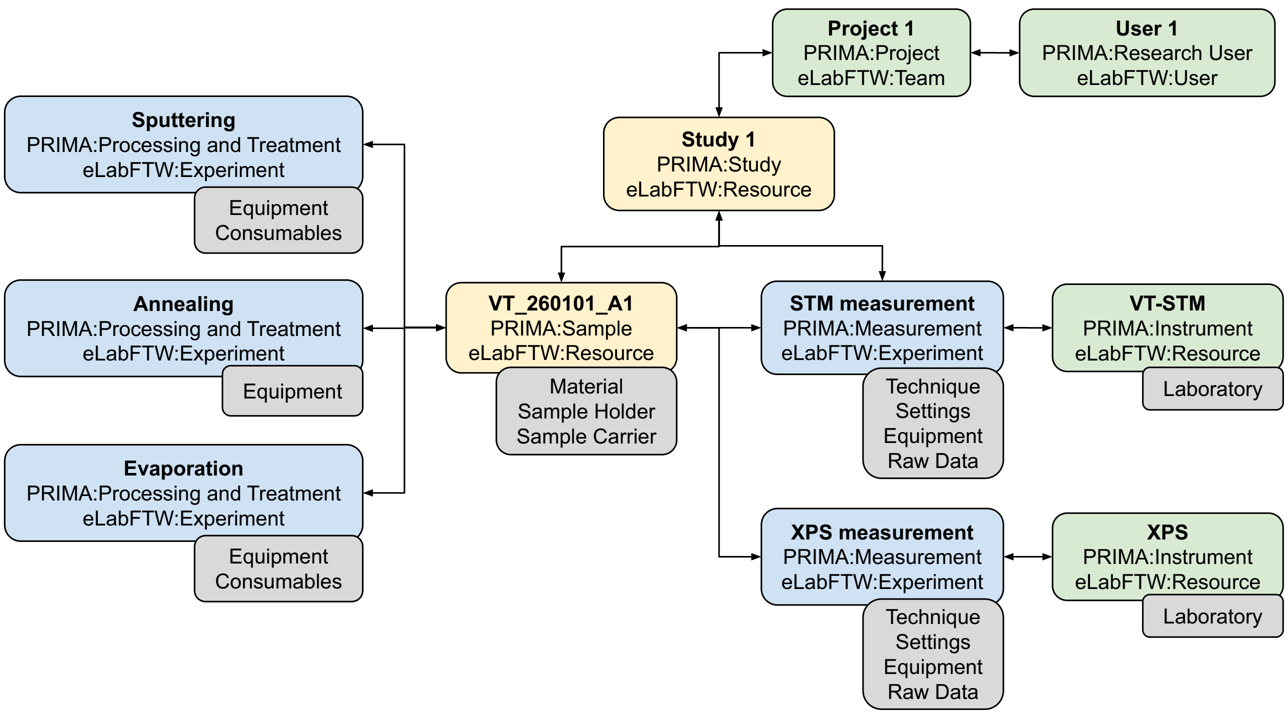

FIGURE 6 | Workflow example showing the mapping of PRIMA into eLabFTW for a typical study consisting of a sample prepared and measured using STM and XPS. Each block represents an instance, where the assigned name, the PRIMA mapped concept, and the corresponding eLabFTW terminology are indicated. The color code is the same as used in Section 4.2. Gray insets list the PRIMA concepts referenced within each instance. The arrows show the links between instances. The relations to **Research User** are omitted to avoid overcrowding the figure.

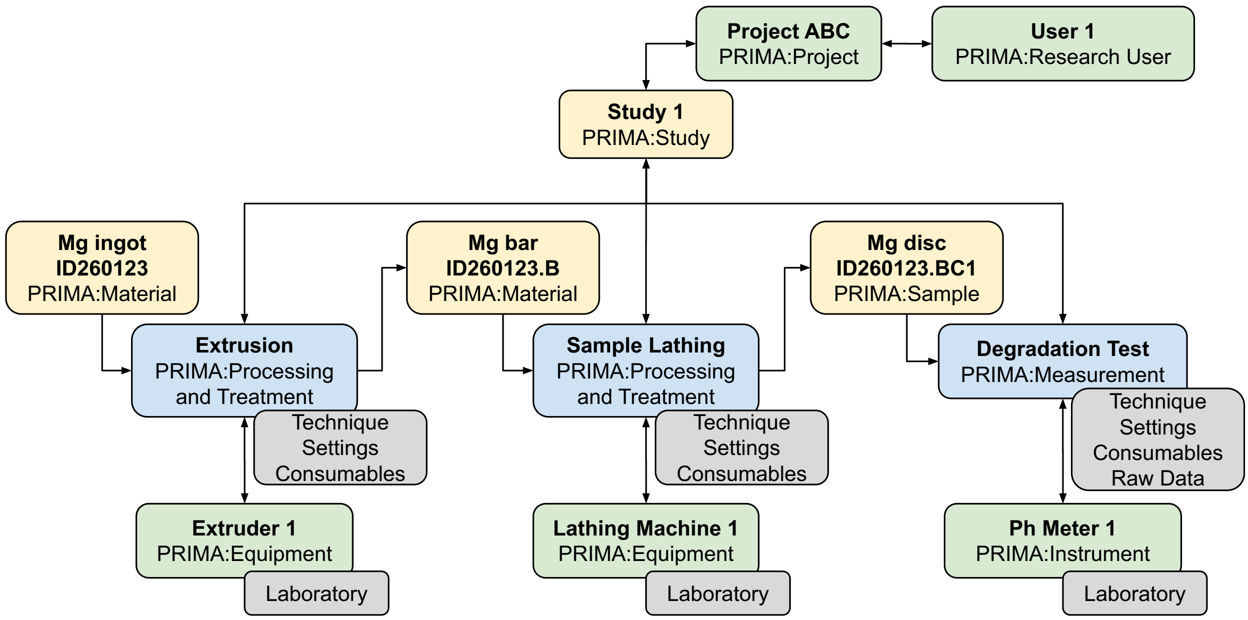

FIGURE 7 | Example showing the mapping of PRIMA into Herbie for a typical workflow in metallic biomaterials research. Each block represents an instance, where the assigned name and the PRIMA mapped concept are indicated. The color code is the same as used in Section 4.2. Gray insets list the PRIMA concepts referenced within each instance. The arrows show the links between instances. The relations to **Research User** are omitted to avoid overcrowding the figure.

ingots are extruded, lathed into discs, and then tested for their degradation behavior. The integration of PRIMA provides a consistent semantic structure for annotating the generated datasets, making them easier to search, filter, exchange, and reuse.

Building upon this semantic foundation, we further extended PRIMA to define an application-specific metadata schema for nanoCT measurements [77]. This schema supports the standardized recording and annotation of nanoCT experiments within Herbie, thereby promoting reproducibility, semantic interoperability, and enhanced data discoverability in materials characterization workflows.

5.2 | Metadata Schemas

As seen in Section 5.1, metadata handling and semantic structuring are typically implemented in ELNs by expert developers, often relying on complex ontologies that remain largely transparent to end users. While such rigorous semantic technologies ensure robustness, they also represent a significant barrier for scientists not adopting RDM tools or practices. In these cases, the use of metadata schemas embedding the PRIMA Thesaurus is an ideal solution.

A metadata schema is a ≪logical plan showing the relationships between metadata elements, normally through establishing rules for the use and management of metadata specifically as regards the semantics, the syntax and the optionality (obligation level) of values≫ [78, 79]. The file containing the actual descriptive information about a given resource, i.e., an instance of the metadata schema validated against it, is called a metadata document. Metadata schemas and corresponding metadata documents can be written in different formats; we adopted JSON [80], but equivalent implementations can be realized using, e.g., Extensible Markup Language (XML).

As an application, we developed a set of modular metadata schemas describing some of the concepts illustrated in Section 4.2, namely **Material** [81, 82], **Sample** [83, 84], **Model** [85, 86], **Raw Data** [87, 88], **Analyzed Data** [89, 90], and **Dataset** [91, 92]. They were selected as they constitute the main input or output of processes in typical experimental or computational workflows. It is worth noting that such workflows are intended as possible examples among several alternatives, and that not all the presented concepts necessarily apply to every case. The developed metadata schemas capture the minimal descriptive information shared across the techniques listed in Section 3.1 and provide a common basis that can be extended and customized to represent application-specific parameters. The use cases illustrated in this section demonstrate how the employment of PRIMA concepts allows the same metadata schemas to represent different scenarios, whereas examples of extensions are discussed in Section 6.

The metadata schemas were registered under the CC BY [63] license in the NFDI-MatWerk Metadata Repository [75] and the NFFA-Europe MetaRepo [93], two instances of the MetaStore [94], a metadata repository and schema registry developed and hosted at the Karlsruhe Institute of Technology.

5.2.1 | *Technical Realization*

Independently of their structure or format, metadata schemas should use vocabularies that follow the FAIR principles [7]. In our context, this means that the contained metadata elements should include the persistent identifiers of the concepts for disambiguation, interoperability, and reuse. This way, different metadata documents can be compared at the semantic level despite variations in design, leading to an easier interpretation of the described resources.

To this end, a custom keyword (e.g., `URI`) can be used to embed the concept URIs as attributes in the metadata schemas, as shown in Figure 8. As an alternative, another custom keyword (e.g., `termURI`) can be used to include the Skosmos REST API endpoints [49] to dynamically retrieve human-readable information associated with the concepts. The endpoint is built according to the path:

$$S + \mathrm{API_S} + r + \mathrm{URI_c} + f$$

where S is the server, $\mathrm{API_S}$ is the Skosmos REST API, r is the request, $\mathrm{URI_c}$ is the concept URI, and f is the format in which the response should be returned (JSON in our case). An example is shown in the middle panel of Figure 9.

The two approaches address different requirements. The use of persistent URIs aligns more closely with FAIR principles, as it ensures long-term preservation of information. In contrast, the API-based approach provides greater flexibility, as semantic content is not hard-coded within the metadata schema but retrieved dynamically from the source. This is particularly advantageous for vocabularies that are frequently updated or shared across multiple platforms, as changes are automatically propagated without requiring any modifications. For the use cases presented in this paper, the use of persistent URIs was preferred to ensure stability.

The known drawback of both approaches is that explicit schema rules must be written to validate them, as only standard JSON Schema keywords are interpreted by validators [95]. An alternative to address this issue is to use native linked data formats like JSON-LD [96]; we plan to explore this option in Kadi4Mat, as mentioned in Section 5.1.1.

```
"institution":
{
    "description": "Hierarchical entity which hosts one or more Laboratories.",
    "URI": "https://purls.helmholtz-metadaten.de/skosmos/prima/institution",
    "type": "string"
},
```

FIGURE 8 | Example of how a PRIMA concept URI is embedded as an attribute in JSON metadata schema.

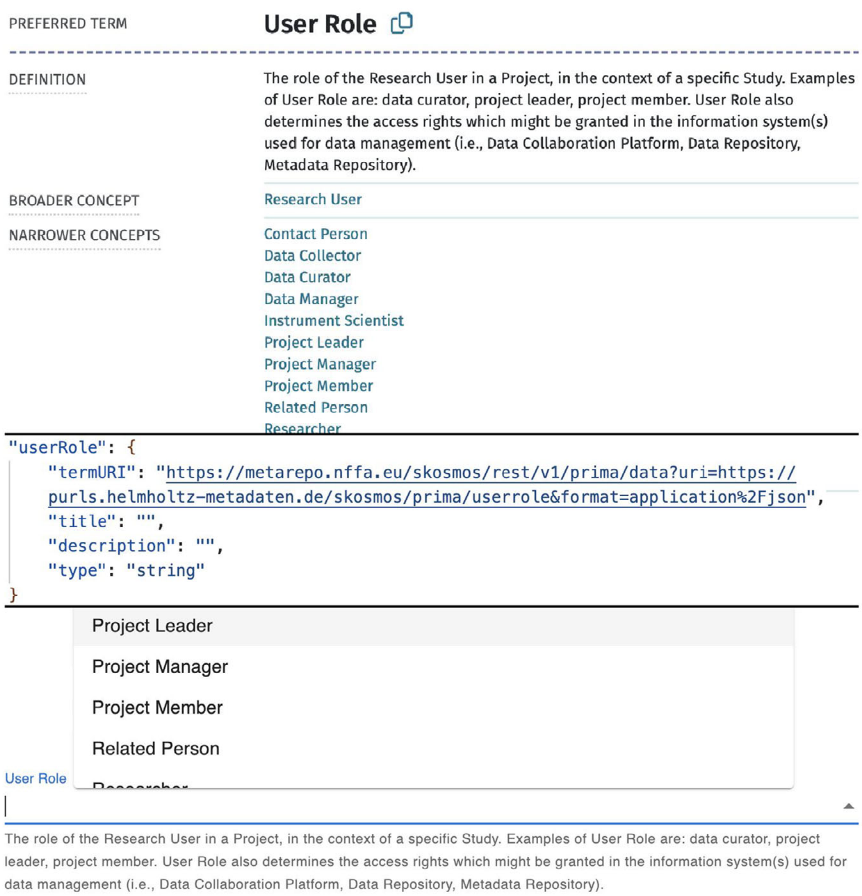

FIGURE 9 | Top panel: controlled list of **User Roles**, modeled as `skos:narrow`, as it appears in Skosmos. Middle panel: use of the custom keyword "`termURI`" in a JSON metadata schema to retrieve the concept title, definition, and narrower concepts from the Skosmos REST API. Bottom panel: how the Metadata Editor renders the machine-readable information by resolving the `termURI` endpoint.

To ease the use of metadata schemas by scientists with no prior technical expertise with JSON, a dedicated tool was developed within NFFA-Europe: the Metadata Editor [97]. This open-source, cross-platform desktop application provides an intuitive graphical interface that abstracts the underlying JSON Schema structure, thereby simplifying metadata document creation, editing, and management.

Among its features, the Metadata Editor is configured to dynamically retrieve semantic information from PRIMA by resolving the associated Skosmos REST API endpoints [49]. This information is rendered directly within the user interface, providing human-readable definitions and controlled lists, modeled as `skos:narrower`, during metadata compilation, as shown in Figure 9. While persistent URIs remain the preferred solution for ensuring stability and interoperability, this complementary approach keeps the displayed information aligned with updates in the underlying vocabularies, supporting more consistent annotation.

The metadata schemas were developed following a modular design. In contrast to a monolithic structure, this approach offers practical advantages by supporting flexible description at different stages of the workflow, as well as individual referencing. For instance, a **Sample** is described once and then referenced, rather than duplicated, in the metadata document of any **Measurement** performed on it. This lightweight provenance is implemented by including in each metadata schema an object called `parents`, consisting of an array of references. When such references are URLs pointing to metadata documents deposited in MetaStore instances, the Metadata Editor is configured to iteratively resolve and aggregate them into a manifest-like representation of the provenance workflow associated with a given resource.

5.2.2 | *Use Case: Theoretical Condensed-Matter Physics*

In theoretical condensed-matter physics publications, the focus is often on the derivation of fundamental equations and on the comparison with experimental or numerical data. While model settings or simulation parameters are described in considerable detail, a standardized description of the complete workflow is usually overlooked. The absence of structured metadata based on shared semantics may lead to ambiguity in the definition of concepts, making it difficult to align results across studies. The resulting poor traceability of models, methods, and derived datasets limits transparency and hinders systematic reuse or extension of published results. In the presented use case, we demonstrate how the introduction of metadata schemas integrating PRIMA concepts provides a possible high-level overview of the computational workflows.

The **Dataset** metadata schema [91] was used to describe the **Publication Data** [98, 99] corresponding to numerical and analytical results from a toy model of superconducting non-Fermi liquids. The latter term denotes exotic but ubiquitous phases of strongly correlated electrons, observed in the normal state of several classes of unconventional superconductors, such as cuprates, pnictides, and heavy-fermion compounds. This phenomenology cannot be reconciled with conventional pictures of standard metals based on the Landau Fermi liquid. The **Model**, described in two **Scientific Publications** [100, 101], is a variant of the Yukawa-Sachdev-Ye-Kitaev (YSYK) approach: it consists of fermions on a lattice with fully disordered hopping parameters, i.e., where hoppings between lattice sites are drawn from a Gaussian statistical distribution with zero mean and given nonzero variance. In addition, the fermions are strongly coupled to an onsite scalar boson through Gaussian-distributed fermion–boson interactions, again with zero mean and fixed variance. Such interaction determines at once the softening and overdamping of the bosons, the destruction of Fermi-liquid quasiparticles for fermions in the normal state, and the pairing interaction that enables superconductivity below a critical temperature T_c. We complemented the reported extensive information with a structured metadata document for the **Model** [102].

The **Datasets** [98, 99] associated with the **Scientific Publications** [100, 101] comprise **Computations** of thermodynamic (entropy, critical temperature, condensation energy, phase stiffness) and spectroscopic (spectral functions, optical conductivity) properties for the lattice-YSYK **Model**. Thanks to the modular design (Section 5.2.1), both corresponding **Dataset** metadata documents [103, 104] reference the same **Model** [102] in the `parents` field. This way, the relations between the various research entities become explicitly encoded, making the workflow transparent, machine-readable, and easier to interpret and reuse.

5.2.3 | *Use Case: Experimental Condensed-Matter Physics*

In complex experimental setups, such as synchrotron radiation techniques or cryo-electron microscopy, several issues may arise when metadata is not properly structured and shared semantics are not adopted. Key problems include poor interoperability between datasets generated by different instruments or laboratories, polysemy, and difficulties in aligning equivalent concepts across studies, thereby limiting the potential for automated data comparison and large-scale knowledge extraction. The PRIMA Thesaurus provides a common semantic layer to describe and relate datasets, thereby supporting their interpretation and reuse.

The described use case shows the adoption of PRIMA concepts at different stages of a complex experimental research workflow. A **Study** was performed combining ab initio **Calculations** with surface science **Measurements** using experimental **Techniques** such as STM, LEED, XPS, NEXAFS, and ARPES to characterize the structural and electronic properties of a two-dimensional crystalline diboron trioxide (B_2O_3) polymorph down to the atomic level. The results and conclusions were reported in a **Scientific Publication** [105], and the associated **Publication Data** [106] was deposited in the Materials Cloud [107] data repository.

The metadata documents were compiled and interconnected using the `parents` field (Section 5.2.1) to represent the provenance chain linking the **Material** [108] to the derived **Sample** [109], and in turn the **Sample** to the corresponding **Dataset** [110]. As a result, the research outcomes become more reproducible, discoverable, and reusable: experiments can be accurately reconstructed, datasets can be queried and compared across repositories, and advanced computational approaches such as data mining and machine learning are enabled.

6 | Discussion

The PRIMA Thesaurus is a mid-level vocabulary positioned between the high-level interoperable schemas (like Dublin Core [23] or DataCite [24]) and the domain- and application-level ontologies. More precisely, its semantic scope is intended to bridge the gap between the representation of research data and workflows in MSE (e.g., MWO [31]) or in related fields (e.g., Metadata4Ing [32]) and the modeling of specific aspects or sub-domains, like processing, structure, and properties of materials (e.g., PMDco [5]), solid-state physics (e.g., MDO [34]), nanomaterials (e.g., NanoMine [37]), or atomistic calculations (e.g., NOMAD Metainfo [35]).

PRIMA builds upon the NFFA-Europe Common Vocabulary [11] and the CSMD [17], expanded to represent a broader community, and enriched by formal SKOS hierarchical relations and PIDs. Despite this increased level of formalization, feedback from our community indicates that, compared with the existing ontologies [10], the PRIMA Thesaurus appears to be easier for scientists to understand, consult, and integrate into everyday workflows.

It is worth noting that the PRIMA Thesaurus was also re-engineered into a formal ontology [111], which has already been cited and evaluated in a recent work [10]. Details on its design, along with its validation and demonstrative use cases, are described in a paper currently in preparation [112]. Nevertheless, although ontologies ultimately provide the level of expressiveness required for a comprehensive knowledge representation, the focus of this work is on lowering the accessibility barrier for a broader research community with limited expertise in RDM and semantics.

The concepts of PRIMA are consistently defined and mutually referenced in their definitions. Although this limits their use as self-standing and independently reusable resources, it reflects the deliberate choice to develop them as part of a self-consistent collection. They were selected to describe the scientific workflows and the data management infrastructures of the participating projects [19–21], which differ from each other in available techniques, offered services, and adopted solutions. This design flexibility renders PRIMA sufficiently general to be applicable to other MSE contexts and offers the potential to accommodate future extensions, be they individual concepts or modular expansions.

As an example, PRIMA was adopted by NFFA-Europe [20], and custom terms were added to more accurately represent project-specific aspects. Furthermore, reflecting the Metadata WG vision underlying the development of PRIMA, each concept can be described at different levels of detail using a metadata schema. For instance, the concept of **Measurement** was expanded and customized for different **Techniques**, such as SEM [14], MRI [15], and nanoCT [77] (see Section 5.1.3). Similarly, as mentioned in Section 5.1.1, an application-level metadata schema to describe experimental **Samples** was developed, including characterizing properties and preparation steps [76]. Ontologies can be adopted as well, including the ontology for Materials Science Laboratory Equipment (MSLE) [113] for qualifying devices, the Chemical Methods Ontology CHMO [114] for extending the description of processing and treatments, or the Crystallographic Defect Ontology Suite (CDOS) [115] for representing crystallographic defects.

As for every vocabulary, the acceptance of PRIMA largely depends on its interaction with other domains. Thus, it is key to establish intense exchange with related fields and subfields to ensure that terminology, definitions, and relationships are continuously shared, aligned, and updated across disciplinary boundaries. In addition to the adoptions and extensions mentioned above, we planned three main mechanisms to put this exchange in practice:

1. Joint schools with other projects and initiatives. Over the past 2 years, we have participated in the iEntrance school [116], the NFFA-Europe school [117], the NFDI-MatWerk schools [118, 119], and the NFDI-MatWerk/FAIRmat school [120]. We intend to maintain a strong presence at these and similar training events in the future.
2. Regular active dialog with various institutions and initiatives to establish new synergies and promote harmonization. We already published a paper in collaboration with the German Federal Institute for Materials Research and Testing (BAM) focused on creep testing [121] and contributed to the release of the Electron Microscopy Glossary [95] in the frame of the Helmholtz Metadata Collaboration (HMC) [122]. We are successfully cooperating with the Bruno Kessler Foundation (FBK) and FAIRmat [123] to focus on nanofabrication processes and to achieve a domain-agnostic description of samples. Moreover, we identified common ground with the fields of engineering sciences and computational chemistry and have maintained a regular exchange with NFDI4Ing [124] and NFDI4Chem [125] since 2021 with the aim of continuously harmonizing terminology across disciplines; on a global level, we are actively involved in the RDA “Sample Type Classification” WG [126].
3. Tutorials and videos to facilitate adoption and further promote dissemination. We recently initiated these activities [127], and we plan to continue investing in them in the future.

A key field of application of the PRIMA Thesaurus is the integration into RDM systems. In particular, ELNs are becoming central to the daily workflow of researchers, serving as the primary interface for recording experimental procedures, parameters, and results. By adopting PRIMA concepts within structured forms, ELNs enable semantic annotation directly at the point of data entry, ensuring consistency, interoperability, and machine-actionability of the generated data, while preserving flexibility and accessibility for end users. This integration effectively bridges the gap between research and RDM practices. We observe that in RDM systems where interaction occurs exclusively through user-friendly forms capable of concealing semantic complexity (e.g., Herbie [68]), the employment of the more expressive PRIMA Ontology [111] derived from the Thesaurus could be preferred.

Finally, to further ensure consistency and facilitate data integration across heterogeneous systems, the semantic descriptions should be harmonized across RDM systems. A notable contribution in this direction was given by the ELNdataBridge [6], offering a seamless data exchange between two ELNs (Herbie [68] and Chemotion [128]) and assessing the integrability of 14 other open-source ELNs for future extension. We are also collaborating with FAIRmat to extend the NOMAD infrastructure [129, 130], currently limited to atomistic computational data, also to experimental data. Recognizing that the actual implementation of PRIMA or any other vocabulary ultimately depends on the specific priorities of each tool, the use cases presented in Section 5.1 reflect ongoing collaborations and synergistic dialog with ELN developers that we are confident to further deepen and expand in the future.

7 | Conclusions

The PRIMA Thesaurus outlined in this paper is a SKOS-based vocabulary compliant with the FAIR principles, designed to describe the mid-level provenance information of computational and experimental studies in the MSE domain. Its main purpose is to improve data findability, reusability, and interoperability while remaining simple enough to support its adoption by scientists not familiar with semantic technologies.

We demonstrated the benefits of adopting PRIMA across two complementary applications: its integration into ELNs and its use within metadata schemas. In all use cases, its adoption enhances the consistency and interpretability of research data, facilitates workflow transparency and dataset comparison, and supports advanced data-driven approaches across heterogeneous platforms and diverse scientific contexts.

The concepts of PRIMA, shared across different techniques, are envisioned to be regularly maintained to always reflect advances in the field. Because they are individually associated with persistent identifiers, these updates do not compromise the stability and validity of their references. They can be adopted by other initiatives in the MSE domain, even from the early stages of the research. In fact, by integrating them into Data Management Plans (DMPs) and RDM systems like ELNs or LIMSs, they can be consistently used from the planning of the activities, throughout all the phases of the study, up to the scientific publication and its associated publication data. Moreover, the persistent URIs assigned to the concepts enable disambiguation, long-term interoperability, and seamless data integration into repositories and RDM systems.

As a future perspective, this semantic layer can be integrated into automated scientific workflows, acting as the logical bridge between raw experimental or computational output and high-level data discovery. This will allow algorithms to perform precise queries and cross-domain correlations beyond what is achievable with unstructured data.

Author Contributions

Rossella Aversa: conceptualization (lead), data curation (lead), formal analysis (lead), investigation (lead), methodology (lead), supervision (lead), writing – original draft (lead), writing – review & editing: (lead). **Alexey Boubnov:** formal analysis (equal), methodology (equal). **Dario De Angelis:** conceptualization (equal), data curation (equal), formal analysis (equal), methodology (equal), validation (equal), writing – review & editing: (equal). **Catriona Eschke:** conceptualization (equal), methodology (equal), software (lead), writing – review & editing (equal). **Maximilian Kabbe:** conceptualization (equal), methodology (equal), writing – review & editing (equal). **Felix Kraus:** data curation (equal), methodology (equal), validation (equal), writing – review & editing (equal). **Richard Thelen:** (conceptualization (equal), data curation (equal), methodology (equal), validation (equal), writing – review & editing (equal). **Mirco Panighel:** (conceptualization (equal), data curation (equal), formal analysis (equal), methodology (equal), validation (equal), writing – review and editing (equal). **Davide Filippo Valentinis:** (data curation (equal), methodology (equal), validation (equal), writing – review and editing (equal).

Acknowledgments

This work was carried out with the support of: the Joint Laboratory Model and Data-driven Materials Characterization (JL MDMC), a cross-centre platform of the Helmholtz Association; the EU's H2020 framework program for research and innovation under grant agreement n. 101007417, NFFA-Europe PILOT; the consortium NFDI-MatWerk, funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under the National Research Data Infrastructure - NFDI 38/1 - project number 460247524; the research program "Engineering Digital Futures" of the Helmholtz Association of German Research Centers; the Helmholtz Metadata Collaboration Platform. We thank Prof. L. Ghiringhelli for the useful discussions and the anonymous reviewers for their detailed and constructive feedback, which led to significative improvements in the manuscript.

Open Access funding enabled and organized by Projekt DEAL.

Funding

This work was supported by European Commission (101007417), Deutsche Forschungsgemeinschaft (460247524), Helmholtz-Gemeinschaft (JLMDMC, HMC).

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. B. Bayerlein, T. Hanke, T. Muth, et al., "A Perspective on Digital Knowledge Representation in Materials Science and Engineering," *Advanced Engineering Materials* 24, no. 6 (2022): 2101176.

2. T. L. Burnett and P. J. Withers, "Completing the Picture through Correlative Characterization," *Nature Materials* 18 (2019): 1041–1049.

3. J. Kimmig, S. Zechel, and U. S. Schubert, "Digital Transformation in Materials Science: A Paradigm Change in Material's Development," *Advanced Materials* 33, no. 8 (2021): 2004940.

4. M. Baker, "1,500 Scientists Lift the Lid on Reproducibility," *Nature* 533 (2016): 452.

5. B. Bayerlein, M. Schilling, H. Birkholz, et al., "PMD Core Ontology: Achieving Semantic Interoperability in Materials Science," *Materials & Design* 237 (2024): 112603.

6. M. Starman, F. Kirchner, M. Held, et al., "ELNdataBridge: Facilitating Data Exchange and Collaboration by Linking Electronic Lab Notebooks via API," *Journal of Cheminformatics* 17 (2025): 86.

7. M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, et al., "The FAIR Guiding Principles for Scientific Data Management and Stewardship," *Scientific Data* 3 (2016): 1–9.

8. F. Xu, N. Juty, C. Goble, S. Jupp, H. Parkinson, and M. Courtot, "Features of a FAIR Vocabulary," *Journal of Biomedical Semantics* 14 (2023): 6.

9. M. Dembska, O. Helle, I. Yadav, and D. Peters, “Proceedings of the First International Workshop on Semantic Materials Science: Harnessing the Power of Semantic Web Technologies in Materials Science co-located with the 20th International Conference on Semantic Systems (SEMANTiCS 2024),” in CEUR Workshop Proceedings, ed. A. Valdestilhas, H. Li, P. Lambrix, and H. Sack (2024).

10. E. Norouzi, J. Waitelonis, and H. Sack, “Proceedings of the First International Workshop on Semantic Materials Science: Harnessing the Power of Semantic Web Technologies in Materials Science co-located with the 20th International Conference on Semantic Systems (SEMANTiCS 2024),” in CEUR Workshop Proceedings, ed. A. Valdestilhas, H. Li, P. Lambrix, and H. Sack (2024).

11. V. Bunakov, T. Griffin, B. Matthews, and S. Cozzini, *Data Analytics and Management in Data Intensive Domains*, ed. L. Kalinichenko, S. O. Kuznetsov, and Y. Manolopoulos (Springer International Publishing, 2017): 248–262.

12. A. Medina-Smith, C. A. Becker, R. L. Plante, et al., “A Controlled Vocabulary and Metadata Schema for Materials Science Data Discovery,” *Data Science Journal* 20 (2021): 1.

13. CODATA RDM Terminology Working Group, “CODATA Research Data Management Terminology,” 2025, accessed October 2025, https://vocabs.ardc.edu.au/viewById/685.

14. R. Joseph, A. Chauhan, C. Eschke, et al., “Metadata Schema to Support FAIR Data in Scanning Electron Microscopy,” *Supplementary Proceedings of the XXIII International Conference on Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL 2021)*, in CEUR Workshop Proceedings, ed. A. Pozanenko (CEUR-WS.org, 2021), 265–277.

15. R. Aversa, N. MacKinnon, and N. Blumenröhr, “FAIR Data Management Workflow for MRI Data,” in *3rd Workshop on Metadata and Research (objects) Management for Linked Open Science (DaMaLOS 2023) - Co-located with Extended Semantic Web Conference (ESWC)*, (2023).

16. J. Greenberg, K. Polson, S. McClellan, X. Zhao, A. Kalinowski, and Y. An, “Proceedings of the First International Workshop on Semantic Materials Science: Harnessing the Power of Semantic Web Technologies in Materials Science Co-Located with the 20th International Conference on Semantic Systems (SEMANTiCS 2024),” in CEUR Workshop Proceedings, ed. A. Valdestilhas, H. Li, P. Lambrix, and H. Sack (2024): 67–77.

17. B. Matthews, S. Sufi, D. Flannery, et al., “Using a Core Scientific Metadata Model in Large-Scale Facilities,” *International Journal of Digital Curation* 5, no. 1 (2010): 106–118.

18. JL-MDMC Metadata Working Group, accessed October 2025, https://jl-mdmc-helmholtz.de/mdmc-activities/metadata-wg/.

19. JL-MDMC, accessed October 2025, https://jl-mdmc-helmholtz.de.

20. NFFA-Europe, accessed October 2025, https://www.nffa.eu.

21. NFDI-MatWerk, accessed October 2025, https://nfdi-matwerk.de.

22. Crossref, “Schema Version 50,” 2020, accessed October 2025, https://gitlab.com/crossref/schema/-/blob/5.0/schemas/common5.0.xsd.

23. Dublin Core Metadata Initiative (DCMI), “DCMI Metadata Terms,” accessed October 2025, https://www.dublincore.org/specifications/dublin-core/dcmi-terms.

24. DataCite, “Metadata Schema Version 4.6,” 2024, accessed October 2025, https://schema.datacite.org.

25. Schema.org, 2024, accessed October 2025, https://schema.org/.

26. European Materials Modelling Council, The Elementary Multiperspective Material Ontology (EMMO), accessed October 2025, https://emmo-repo.github.io.

27. T. Lebo, S. Sahoo, and D. McGuinness, “PROV-O: The PROV Ontology, W3C Recommendation, World Wide Web Consortium (W3C),” 2013, accessed November 2025, https://www.w3.org/TR/2013/REC-prov-o-20130430/.

28. S. Samuel and B. König-Ries, *The Semantic Web: ESWC 2017 Satellite Events*, ed. E. Blomqvist, K. Hose, H. Paulheim, A. Lawrynowicz, F. Ciravegna, and O. HartigSpringer International Publishing, 2017: 17–20.

29. L. N. Soldatova and R. D. King, “An Ontology of Scientific Experiments,” *Journal of the Royal Society Interface* 3 (2006): 795–803.

30. S. Sufi and B. Matthews, *Knowledge and Data Management in GRIDs*, ed. D. Talia, A. Bilas, and M. Dikaiakos (Springer, 2007): 135–149.

31. A. A. Guzmán, A. Z. Ihsan, S. Fathalla, and A. Gedsun, “The MatWerk Ontology (MWO),” 2024, accessed June 2026, http://purls.helmholtz-metadaten.de/mwo/.

32. D. Iglezakis, G. Lanza, S. Arndt, et al., “Metadata4Ing: An Ontology for Describing the Generation of Research Data Within a Scientific Activity,” (2025), https://doi.org/10.5281/zenodo.14982558.

33. American Physical Society, “Physics Subject Headings,” accessed October 2025, https://physh.org/browse.

34. H. Li, R. Armiento, E. J. Z. Lambrix, et al., *The Semantic Web–ISWC 2020* (Springer International Publishing, 2020), 212–227.

35. L. M. Ghiringhelli, C. Baldauf, T. Bereau, et al., “Shared Metadata for Data-Centric Materials Science,” *Scientific Data* 10 (2023): 626.

36. CODATA-VAMAS Working Group on the Description of Nanomaterials and J. Rumble, “Uniform Description System for Materials on the Nanoscale, Version 2.0,” 2016, https://doi.org/10.5281/zenodo.56720.

37. J. P. McCusker, N. Keshan, S. Rashid, M. Deagen, C. Brinson, and D. L. McGuinness, *The Semantic Web–ISWC 2020*, ed. J. Z. Pan, V. Tamma, C. d'Amato, K. Janowicz, B. Fu, A. Polleres, O. Seneviratne, and L. Kagal (Springer International Publishing, 2020): 144–159.

38. R. Aversa, “Building the MDMC-NEP Glossary, NFDI4ING SIG Metadata and Ontologies Meeting,” 2024, accessed June 2026, https://publikationen.bibliothek.kit.edu/1000168783.

39. Wikipedia, accessed October 2025, https://www.wikipedia.org.

40. R. Aversa, A. Boubnov, D. De Angelis, et al., “The MDMC-NEP Glossary of Terms,” 2024, https://doi.org/10.5281/zenodo.10663833.

41. National Information Standards Organization, ANSI/NISO Z39.19-2005, “(R2010) Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies,” 2010, accessed June 2026, https://www.niso.org/publications/ansiniso-z3919-2005-r2010.

42. T. Baker, S. Bechhofer, A. Isaac, A. Miles, G. Schreiber, and E. Summers, “Key Choices in the Design of Simple Knowledge Organization System (SKOS),” *Journal of Web Semantics* 20 (2013): 35.

43. A. Miles and S. Bechhofer, “SKOS Simple Knowledge Organization System Reference, W3C Recommendation, World Wide Web Consortium (W3C),” 2009, accessed November 2025, https://www.w3.org/TR/skos-reference/.

44. T. R. Gruber, “Toward Principles for the Design of Ontologies Used for Knowledge Sharing?,” *International Journal of Human-Computer Studies* 43, no. 5–6 (1995): 907–928.

45. O. Suominen and E. Hyvönen, *Knowledge Engineering and Knowledge Management*, ed. A. ten Teije, J. Völker, S. Handschuh, H. Stuckenschmidt, M. d'Acquin, A. Nikolov, N. Aussenac-Gilles, and N. Hernandez (Springer, 2012): 383–397.

46. DataCite, “Contributor Type Version 4.4.,” 2022, accessed November 2025, http://purl.org/datacite/v4.4/ContributorType.

47. EVOKS Repository, accessed April 2026, https://github.com/kit-data-manager/EVOKS.

48. JSONForms, accessed April 2026, https://jsonforms.io.

49. Skosmos, accessed April 2026, https://skosmos.org.

50. NFDI-MatWerk SKOSMOS Instance, "PRIMA - PRovenance Information for MAterials Science," accessed November 2025, https://matwerk.datamanager.kit.edu/skosmos/prima.

51. NFFA-Europe SKOSMOS Instance, "PRIMA - PRovenance Information for MAterials Science," accessed November 2025, https://metarepo.nffa.eu/skosmos/prima.

52. R. Aversa, D. Valentinis, C. Eschke, et al., "PRIMA Thesaurus," accessed October 2025, https://purls.helmholtz-metadaten.de/skosmos/prima/.

53. Helmholtz Metadata Collaboration (HMC), "PIDA: Persistent Identifiers for Semantic Artifacts," accessed October 2025, https://purls.helmholtz-metadaten.de/.

54. J. McCusker, L. D. McIntosh, C. Shaffer, et al., "Guiding Principles for Technical Infrastructure to Support Computable Biomedical Knowledge," *Learning Health Systems* 7, no. 3 (2023): e10352.

55. S. J. D. Cox, A. N. Gonzalez-Beltran, B. Magagna, M.-C. Marinescu, and S. Markel, "Ten Simple Rules for Making a Vocabulary FAIR," *PLOS Computational Biology* 17, no. 6 (2021): e1009041.

56. W. Hugo, Y. Le Franc, G. Coen, J. Parland-vonEssen, and L. Bonino, "D2.5 FAIR Semantics Recommendations Second Iteration," 2020, https://doi.org/10.5281/zenodo.4314321.

57. D. Garijo and P.-V. María, *Applications and Practices in Ontology Design, Extraction, and Reasoning, Volume 49 of Studies on the Semantic Web*, ed. G. Cota, M. Daquino, and G. Pozzato (IOS Press, 2020): 39–54.

58. A. Gaignard, T. Rosnet, F. De Lamotte, V. Lefort, and M.-D. Devignes, "FAIR-Checker: Supporting Digital Resource Findability and Reuse with Knowledge Graphs and Semantic Web Standards," *Journal of Biomedical Semantics* 14, no. 1 (2023): 7.

59. W3C, "Web Ontology Language (OWL)," 2012, accessed October 2025, https://www.w3.org/OWL/.

60. W3C, "SKOS Home Page," accessed November 2025, https://www.w3.org/2004/02/skos.

61. PRIMA - PRovenance Information for MAterials Science, BARTOC, accessed November 2025, http://bartoc.org/en/node/20830.

62. A. Ledl, "104. Deutscher Bibliothekartag (Nürnberg, 2015), 2015, accessed June 2026, https://opus4.kobv.de/opus4-bib-info/frontdoor/index/index/year/2015/docId/1677.

63. Creative Commons Corporation, Attribution 4.0 International Public License, accessed October 2025, https://creativecommons.org/licenses/by/4.0/.

64. T. Rodani, E. Osmenaj, A. Cazzaniga, M. Panighel, C. Africh, and S. Cozzini, "Towards the FAIRification of Scanning Tunneling Microscopy Images," *Data Intelligence* 5 (2023): 27.

65. S. G. Higgins, A. A. Nogiwa-Valdez, and M. M. Stevens, "Considerations for Implementing Electronic Laboratory Notebooks in an Academic Research Environment," *Nature Protocols* 17 (2022): 179.

66. N. Brandt, L. Griem, C. Herrmann, et al., *Data Science Journal* 20 (2021): 8.

67. eLabFTW, accessed June 2026, https://github.com/elabftw/elabftw/.

68. F. Kirchner, C. Eschke, A.-L. Höhme, et al., "Herbie - The Semantic Laboratory Notebook & Research Database," 2026, https://doi.org/10.5281/zenodo.12205429.

69. T. Mathia, P. Pawlus, and M. Wieczorowski, "Recent Trends in Surface Metrology," *Wear* 271 (2011): 494–508.

70. Joint Committee for Guides in Metrology, International Vocabulary of Metrology (2012), https://doi.org/10.59161/JCGM200-2012.

71. R. Thelen, "Atomic Force Microscopy Dataset," 2025, https://doi.org/10.5281/zenodo.17672031.

72. R. Thelen, "Confocal Scanning Microscopy Dataset," 2025, https://doi.org/10.5281/zenodo.17672155.

73. R. Thelen, "Vertical Scanning Interferometry Dataset," 2025, https://doi.org/10.5281/zenodo.17672093.

74. R. Thelen, "Sample Metadata Document," 2025, accessed June 2026, https://matwerk.datamanager.kit.edu/api/v2/metadata/ce7a3aa3-3f6f-45db-a21e-304d92451aa9.

75. NFDI-MatWerk Instance, Dashboard, accessed November 2025, https://metarepo.nffa.eu.

76. R. Aversa, "General Sample Schema, NFDI-MatWerk Metadata Repository," 2025, accessed June 2026, https://matwerk.datamanager.kit.edu/api/v2/schemas/general_sample.

77. F. Kirchner, D. C. F. Wieland, S. Irvine, et al., "An Ontology-Based Description of Nano Computed Tomography Measurements in Electronic Laboratory Notebooks," *Scientific Data* 13 (2026): 432.

78. ISO 23081–1:2017, "Information and Documentation – Records Management Processes – Metadata for Records. Part 1: Principles," 2017, accessed June 2026, https://www.iso.org/standard/73172.html.

79. ISO 23081–2:2021, "Information and Documentation – Metadata for Managing Records. Part 2: Conceptual and Implementation Issues," 2021, accessed June 2026, https://www.iso.org/standard/81600.html.

80. JSON Schema, accessed November 2025, https://json-schema.org/.

81. R. Aversa, "Material Schema, NFDI-MatWerk Metadata Repository," 2025, accessed November 2025, https://matwerk.datamanager.kit.edu/api/v2/schemas/material_schema.

82. R. Aversa, "Material Schema, NFFA-Europe MetaRepo," 2025, accessed November 2025, https://metarepo.nffa.eu/api/v1/schemas/material_schema.

83. R. Aversa, "Sample Schema, NFDI-MatWerk Metadata Repository," 2025, accessed November 2025, https://matwerk.datamanager.kit.edu/api/v2/schemas/sample_schema.

84. R. Aversa, "Sample Schema, NFFA-Europe MetaRepo," 2025, accessed November 2025, https://metarepo.nffa.eu/api/v1/schemas/sample_schema.

85. R. Aversa, "Model Schema, NFDI-MatWerk Metadata Repository," 2025, accessed November 2025, https://matwerk.datamanager.kit.edu/api/v2/schemas/model_schema.

86. R. Aversa, "Model Schema, NFFA-Europe MetaRepo," 2025, accessed November 2025, https://metarepo.nffa.eu/api/v1/schemas/model_schema.

87. R. Aversa, "Raw Data Schema, NFDI-MatWerk Metadata Repository," 2025, accessed November 2025, https://matwerk.datamanager.kit.edu/api/v2/schemas/raw_data_schema.

88. R. Aversa, "Raw Data Schema, NFFA-Europe MetaRepo," 2025, accessed November 2025, https://metarepo.nffa.eu/api/v1/schemas/raw_data_schema.

89. R. Aversa, "Analyzed Data Schema, NFDI-MatWerk Metadata Repository," 2025, accessed November 2025, https://matwerk.datamanager.kit.edu/api/v2/schemas/analyzed_data_schema.

90. R. Aversa, "Analyzed Data Schema, NFFA-Europe MetaRepo," 2025, accessed November 2025, https://metarepo.nffa.eu/api/v1/schemas/analyzed_data_schema.

91. R. Aversa, "Dataset Schema, NFDI-MatWerk Metadata Repository," 2025, accessed November 2025, https://matwerk.datamanager.kit.edu/api/v2/schemas/dataset_schema.

92. R. Aversa, "Dataset Schema, NFFA-Europe MetaRepo," 2025, accessed November 2025, https://metarepo.nffa.eu/api/v1/schemas/dataset.

93. NFFA-Europe instance, Dashboard, accessed November 2025, https://matwerk.datamanager.kit.edu/.

94. A. Prabhune, H. Ansari, A. Keshav, R. Stotzka, M. Gertz, and J. Hesser, "MetaStore: A Metadata Framework for Scientific Data Repositories," in *2016 IEEE International Conference on Big Data (Big Data)* (IEEE, 2016): 3026–3035.

95. A. Azócar Guzmán, O. Brendike-Mannix, and V. Hofmann, "Electron Microscopy Glossary, Repository-Code," 2024, accessed October 2025, https://codebase.helmholtz.cloud/em.

96. JSON for Linked Data, 2024, accessed October 2025, https://json-ld.org.

97. eXact Lab, Metadata Editor, accessed November 2025, https://metadata-editor.gitlab.io/documentation/.

98. D. Valentinis, "Correlation between Phase Stiffness and Condensation Energy across the Non-Fermi to Fermi-Liquid Crossover in the Yukawa-Sachdev-Ye-Kitaev Model on a Lattice-DATA," (2024), https://doi.org/10.35097/1775.

99. D. Valentinis, "BCS to Incoherent Superconductivity Crossover in the Yukawa-Sachdev-Ye-Kitaev Model on a Lattice- DATA," (2023), https://doi.org/10.35097/1772.

100. D. Valentinis, G. A. Inkof, and J. Schmalian, "Correlation Between Phase Stiffness and Condensation Energy Across the Mon-{Fermi} to {Fermi}-Liquid Crossover in the {Yukawa-Sachdev-Ye-Kitaev} Model on a Lattice," *Physical Review Research* 5 (2023): 043007.

101. D. Valentinis, G. A. Inkof, and J. Schmalian, "BCS to Incoherent Superconductivity Crossover in the Yukawa-Sachdev-Ye-Kitaev Model on a Lattice," *Physical Review B* 108 (2023): L140501.

102. D. Valentinis, "Model Metadata Document," 2026, accessed April 2026, https://matwerk.datamanager.kit.edu/api/v2/metadata/f596d4a7-04ef-4013-a7c5-ffc0846845f8.

103. D. Valentinis, "Dataset Metadata Document," 2025, accessed April 2026, https://matwerk.datamanager.kit.edu/api/v2/metadata/da5bf365-eeed-44c8-aa19-0bb3765fb201.

104. D. Valentinis, "Dataset Metadata Document," 2025, accessed April 2026, https://matwerk.datamanager.kit.edu/api/v2/metadata/fbe6f701-775b-4aa0-90f4-1b7b73ad6b50.

105. T. Zio, M. Dirindin, C. D. Giorgio, et al., "Two-Dimensional Diboron Trioxide Crystal Composed by Boroxol Groups," *Science (New York, N.Y.)* 390, no. 6768 (2025): 95–99, https://doi.org/10.1126/science.adv2582.

106. T. Zio, M. Dirindin, C. D. Giorgiod, et al. Two-Dimensional Diboron Trioxide Crystal Composed by Boroxol Groups, (Materials Cloud, 2025), https://doi.org/10.24435/materialscloud:e1-bf.

107. L. Talirz, S. Kumbhar, E. Passaro, et al., "Materials Cloud, a Platform for Open Computational Science," *Scientific Data* 7, no. 1 (2020): 299.

108. D. De Angelis, "Material Metadata Document," 2025, accessed November 2025, https://metarepo.nffa.eu/api/v1/metadata/517df414-d97d-43a0-b6bd-4e6f10be768f.

109. D. De Angelis, "Sample Metadata Document," 2025, accessed November 2025, https://metarepo.nffa.eu/api/v1/metadata/000f3c1f-6b6b-4f38-9be9-5f84c5c4b6cd.

110. D. De Angelis, "Dataset Metadata Document," 2025, accessed November 2025, https://metarepo.nffa.eu/api/v1/metadata/0f56383c-85bf-4321-9c81-972a998612fb.

111. Metadata WG, The PRIMA Ontology, accessed March 2026, https://github.com/Materials-Data-Science-and-Informatics/MDMC-NEP-top-level-ontology.

112. A. Z. Ihsan, M. Jalali, R. Aversa, et al., "PRIMA: An Ontology to Capture Provenance Information for Materials Science," In Preparation.

113. M. Jalali, M. Mail, R. Aversa, and C. Kübel, "MSLE: An Ontology for Materials Science Laboratory Equipment–Large-Scale Devices for Materials Characterization," *Materials Today Communications* 35 (2023): 105532.

114. Royal Society of Chemistry, Chemical Methods Ontology, accessed October 2025, https://ontobee.org/ontology/CHMO.

115. A. Azócar Guzmán, "Crystallographic Defect Ontology Suite," 2025, accessed November 2025, https://cdos.readthedocs.io/en/latest/docs/ontologies.html.

116. R. Aversa, *Metadata Management in Scientific Research: An Overview* (iENTRANCE@ENL Advanced School, 2024).

117. NFFA-Europe, "Summer School 2024 in Crete - Data Management and Virtual Access," 2024, accessed June 2026, https://nffa.eu/summer-school-2024/.

118. NFDI-MatWerk, "Summer School 2024 in Bremen - Research Data Management in Materials Science and Engineering," 2024, accessed June 2026, https://www.eusmat.net/research/other-events/nfdi-matwerk-summer-school-2024/.

119. NFDI-MatWerk, "Summer School 2025 in Aachen - Research Data Management in Materials Science and Engineering," 2025, accessed June 2026, https://www.eusmat.net/research/other-events/nfdi-matwerk-spring-school-2025/.

120. NFDI-MatWerk/FAIRmat, "Spring School 2026 in Erlangen - Research Data Management in Materials Science and Engineering," 2026, accessed June 2026, https://www.eusmat.net/research/other-events/nfdi-spring-school-2026/.

121. L.Á. vila Calderón, Y. Shakeel, A. Gedsun, et al., "Management of Reference Data in Materials Science and Engineering Exemplified for Creep Data of a Single-Crystalline Ni-Based Superalloy," *Acta Materialia* 286 (2025): 120735.

122. Helmholtz Metadata Collaboration, accessed June 2026, https://helmholtz-metadaten.de.

123. FAIRmat, accessed June 2026, https://www.fairmat-nfdi.eu/fairmat/.

124. NFDI4ING, accessed June 2026, https://nfdi4ing.de.

125. NFDI4Chem, accessed June 2026, https://nfdi4chem.de.

126. Research Data Alliance, Sample Type Classification Working Group, accessed June 2026, https://www.rd-alliance.org/groups/sample-type-classification-wg.

127. Karlsruhe Institute of Technology - Scientific Computing Center, (Meta)data management Cookbook, accessed June 2026, https://kit-data-manager.github.io/metadata-management-cookbook/.

128. P. Tremouilhac, A. Nguyen, Y.-C. Huang, et al., "Chemotion ELN: An Open Source Electronic Lab Notebook for Chemists in Academia," *Journal of Cheminformatics* 9 (2017): 54.

129. L. Ghiringhelli, C. Carbogno, S. Levchenko, et al., "Towards Efficient Data Exchange and Sharing for Big-Data Driven Materials Science: Metadata and Data Formats," *npj Computational Materials* 3 (2017): 46.

130. M. Scheidgen, L. Himanen, A. N. Ladines, et al., "NOMAD: A Distributed Web-Based Platform for Managing Materials Science Research Data," *Journal of Open Source Software* 8 (2023): 5388.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.