CEN/WS OntoWF

Date: 2025-11-13

Draft CWA WSDON001:2025

Workshop: Terminology definition for domain ontologies in materials science

Terminology definition for domain ontologies in materials science

ICS: 01.020, 01.120, 35.240.30

CCMC will prepare and attach the official title page.

Contents

Page

Foreword

This CEN Workshop Agreement (CWA DON:2025) has been developed in accordance with the CEN-CENELEC Guide 29 “CEN/CENELEC Workshop Agreements – A rapid way to standardization” and with the relevant provisions of CEN/CENELEC Internal Regulations - Part 2. It was approved by the Workshop CEN “Terminology definition for domain ontologies in materials science”, the secretariat of which is held by DIN, consisting of representatives of interested parties on 2025-05-16, the constitution of which was supported by CEN following the public call for participation made on 2025-04-14. However, this CEN Workshop Agreement does not necessarily include all relevant stakeholders.

The draft text of this CEN Workshop Agreement was provided to CEN for publication on 2025-11-14.

Results incorporated in this CWA received funding from the European Union’s Horizon Europe research and innovation programme under grant agreement No 101091496.

Attention is drawn to the possibility that some elements of this document may be subject to patent rights. CEN-CENELEC policy on patent rights is described in CEN-CENELEC Guide 8 “Guidelines for Implementation of the Common IPR Policy on Patent”. CEN shall not be held responsible for identifying any or all such patent rights.

Although the Workshop parties have made every effort to ensure the reliability and accuracy of technical and non-technical descriptions, the Workshop is not able to guarantee, explicitly or implicitly, the correctness of this document. Anyone who applies this CEN Workshop Agreement shall be aware that neither the Workshop nor CEN can be held liable for damages or losses of any kind whatsoever. The use of this CEN Workshop Agreement does not relieve users of their responsibility for their own actions, and they apply this document at their own risk. The CEN Workshop Agreement should not be construed as legal advice authoritatively endorsed by CEN/CENELEC.

Introduction

Semantic technologies — including ontologies and knowledge graphs — are increasingly applied to manage materials and process data in line with the FAIR principles (Findability, Accessibility, Interoperability, Reusability — see Annex A for further information). In materials science and manufacturing, large volumes of heterogeneous data often remain difficult to access and integrate, limiting comprehensive analysis and reuse.

Ontologies provide the conceptual foundation for semantic technologies. However, developing ontologies solely through a top-down approach can be time-consuming and cumbersome, as experts must start from general concepts and iteratively specialise them for specific domains. In contrast, a bottom-up approach starts from specific applications and data, identifies the application-specific terminology, and maps the data to well-defined concepts.

While attractive in practice, bottom-up ontology development raises practical questions about how to define terminology consistently — covering naming conventions, concise and differentiating definitions, sources and provenance of descriptions, multilingual labels, relations between concepts, and quality assurance.

This CEN Workshop Agreement addresses these challenges by providing a guideline and repeatable workflow for identifying, defining, and maintaining terminology for application ontologies in materials science and engineering (MSE). The guideline aims to support consistent, scalable, and interoperable ontology creation driven by real use cases and datasets, and to complement existing terminology standards and ontology best practices.

1 Scope

This CWA defines a workflow how to define, review, and implement domain-specific terminology of an application in materials science and engineering (MSE) to support domain ontology creation.

The CWA covers:

— roles and responsibilities for terminology work;

— a stepwise process from scoping and term harvesting to publication and implementation;

— a structured term record with mandatory data categories (e.g. definition type, relations, sources);

— quality insurance and review gates; and

— implementation in terminology databases.

This CWA is applicable to industrial and research use cases across MSE sub-domains and languages. It is intended for terminology curators, domain experts, ontology engineers, and data stewards involved in terminology-driven ontology development.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

— ISO 704, *Terminology work — Principles and methods*

— ISO 860, *Terminology work — Harmonization of concepts and terms*

— ISO 12616, *Terminology work in support of multilingual communication - Part 1: Fundamentals of translation-oriented terminography*

— ISO 30042, *Management of terminology resources - TermBase eXchange (TBX)*

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminological databases for use in standardisation at the following addresses:

— ISO Online browsing platform: available at http://www.iso.org/obp/

— IEC Electropedia: available at http://www.electropedia.org/

3.1

concept

unit of knowledge created by a unique combination of characteristics, which are not necessarily bound to natural languages but can be influenced by social and cultural backgrounds

Note 1 to entry: This dynamic nature leads to concepts also being regarded as units of thinking.

[SOURCE: ISO 26162-1:2020, 3.1.2]

3.2
designation
representation of a *concept* (3.1)

[SOURCE: ISO 26162-1:2020, 3.1.3, modified – Note 1 to entry and Note 2 to entry removed]

3.3
term
designation (3.2) through linguistic means

EXAMPLE : "laser printer", "planet", "pacemaker", "chemical compound", "¾ time", "Influenza A virus", "oil painting"

Note 1 to entry: Terms can be partially or wholly verbal.

[SOURCE: ISO 26162-1:2020, 3.1.4]

3.4
object
anything perceivable or conceivable, which may be material, immaterial, or imagined

EXAMPLE : an engine, a conversion ratio, or a unicorn

Note 1 to entry: In terminology work, a designation does not link directly to the object; the connection from designation to object is mediated by the concept (semiotic triangle).

[SOURCE: ISO 26162-1 :2020, 3.1, modified – Note 1 to entry added]

3.5
definition

representation of a concept by an expression that describes and differentiates it from related concepts

[SOURCE: ISO 1087:2019, 3.3.1]

4 Guidelines for terminology definition

4.1 Guiding principles

4.1.1 Terminology work for MSE domain

Terminology science aims to analyse, define, and systematically organise concepts and their relations within a domain, and to manage their designations (terms, names, symbols). Its goal is to enable precise, consistent, and interoperable communication across communities and languages. Therefore, the terminology itself shall follow the FAIR principles under consideration of organisational restrictions like internal confidentiality. For the terminology development, established methodologies from standards and domain conventions shall be followed. Due to the fact that concepts are the essential elements for terminology and ontology science, each terminological entry shall contain data for only one concept. The overall goal of the terminology work in this CWA is to lay the foundation for the expansion of the terminology to taxonomies and more expressive ontologies. Annex B outlines the role that concepts play within ontologies and knowledge graphs.

4.1.2 General perspective

There is no direct correlation between names and objects. The relationship between object, name, and concept can be illustrated by the semiotic triangle. The semiotic triangle is a conceptual model that represents the absence of a direct relationship between the object and its name.

Additionally, concept development comes before designation. Work shall model the concept (unit of knowledge) first, designations (terms, names, symbols) are then assigned. Concept mediates between object and designation.

4.1.3 Community, scope, and target area

Terminology reflects communities of practice. Conflicting variants may coexist if scope, usage and status are explicitly recorded.

4.1.4 Resources

The concept and the full terminology entry shall be defined with high precision and the potential of acceptance in the aimed domain. Therefore, suitable sources are based on domain standards, following terminology standards, established textbooks and scientific publications.

4.1.5 Workflow

In order to achieve stakeholder acceptance and to ensure the quality of the terminology entries and the sustainable use of terminology, it is recommended to implement a workflow for review and approval process, i.e. a quality management process that incorporates consensus building.

4.2 Specifying Concepts

4.2.1 Characteristic

A characteristic is an abstraction of properties of objects that is used to form and delimit concepts. Characteristics are themselves conceptual units. They are fundamental for determining concept relations and for writing definitions.

Example: resistance to shear, shape retaining

4.2.2 Concept

Unit of knowledge formed by a unique combination of characteristics. Specified intentionally by stating the immediate superordinate concept and at least one delimiting characteristic and mediating between objects and their designations.

Example: solid material - state of matter that retains definite shape and volume and exhibits a shear resistance at specified conditions

4.3 Specifying Labels

A label is a designation (terms, proper names, symbols) string assigned to a specified concept for human-readable naming and retrieval. Exactly one preferred label shall be provided per language. Admitted synonyms shall be added where needed and deprecated/obsolete forms shall be marked explicitly. Labels shall be assigned immediately after the concept is specified. They make the concept findable and usable across communities and languages. Use concise noun phrases in singular, without articles; apply sentence case (proper names and acronyms as customary). Avoid embedding constraints (units, ranges, “and/or”), equations, or hidden assumptions in labels.

4.4 Definition of the meaning of a concept

A definition of the meaning of a concept shall follow the rules and regulations laid out in ISO 704. For definitions in the field of materials sciences these rules and regulations should however be expanded and / or adapted by the following:

- A definition shall provide the essential characteristics of a concept in its proper place in the concept system and thus differentiates it from related concepts.
- A definition shall take the needs and the background of the target audience into consideration. In this sense it has to be clear, concise and consistent. It is recommendable to perform a test for completeness: In taking a sentence from the domain and substituting the term with its definition, the whole sentence should still make sense.
- A definition alone can be insufficient. Developing terminology resources for non-experts often requires amplification, for example by adding other types of information (e.g. explanations, notes or examples) or by the inclusion of a representation in other media (e.g. equations or graphic illustrations).
- A definition is a statement in the form of an incomplete sentence without a full stop. The definition begins with a noun or other part of speech stating the immediate superordinate concept associated with the concept being defined. This noun is followed by delimiting characteristics that distinguish the concept from its coordinate concept(s). An article (generally indefinite) is implied but not written at the beginning of a definition. The domain, subject or a special usage can be indicated in angle brackets (<>) at the beginning of the definition. If a symbol is consensual for the concept being defined, it can be positioned at the beginning.
- Definitions are the result of conceptual research and analysis. If definitions are authored, this shall be done in line with the concept system of the culture and subject field. If elements from other sources are reused or adapted, the reference, authorship, and version shall be explicitly mentioned.

EXAMPLE 1

Stress
<materials testing – metallic materials>
R
at any moment during the test, force divided by the original cross-sectional area, *S*o, of the test piece

Note 1 to entry: All references to stress in this document are to engineering stress.

[SOURCE: ISO 6892-1:2019: Metallic materials — Tensile testing — Part 1: Method of test at room temperature, Item 3.10]

EXAMPLE 2

electrical resistance
<materials testing – paints and varnishes>
R
ratio of the potential difference along a conductor and the current through the conductor

$$R = \frac{U}{I} \tag{1}$$

where

U is the potential difference;
I is the current.

The unit of electrical resistance is the ohm (Ω), given by:

The electrical resistance depends on the material of the conductor, its dimensions (length and cross-section) and its temperature.

Note 1 to entry: Resistance is given by Ohm's law shown in Formula (1)

[SOURCE: ISO 15091:2019: Paints and varnishes — Determination of electrical conductivity and resistance, Item 3.1]

EXAMPLE 3

grain-oriented electrical steel
<steel products>
steel alloyed with silicon that is anisotropic in that it possesses a metallurgical structure that gives preferential magnetic properties in the direction of rolling

Note 1 to entry: These steels are supplied with an insulating coating on both surfaces.

[SOURCE: ISO 6929:2013: Steel products — Vocabulary, Item 1.3.3.2]

4.5 Sources of information

4.5.1 Description of sources

When looking at existing ontologies and terminologies from institutions such as ISO, DIN and BS, these have to be defined, precise, traceable and consistent by regulation. This is crucial for unambiguous communication.

Sources can be categorised as either primary or secondary. Primary sources are the standards themselves, which are developed through formal consensus and are widely accepted by the community. Examples can be found in Section 4.4. where the three mentioned examples of concepts are described and the source of a standard is attributed. In contrast, secondary sources are documents with narrower dissemination or lower consensus, such as technical or industry-recognised dictionaries, peer-reviewed literature and academic papers, authorities' handbooks, and regulatory definitions.

Sources shall be attributed as described in Section 4.4, including the identifier, version number, name of the source and, where possible, the section number. When working with multiple languages, at least one source per language should be specified (see Section 4.6 for further details). Multiple sources in the same language may also be cited to justify the entry structure and to cross-validate the definition (concordant sources strengthen the verification).

When external sources are used, special care regarding copyright and IP is necessary, see the following Section 4.5.2.

4.5.2 Handling copyrights and references

ISO, DIN and IEC standards are protected by copyright. If text is reproduced verbatim, written permission must be obtained from the copyright holder (usually ISO, IEC, or DIN). This has to also include a copyright notice and acknowledgement. In the event of paraphrasing, the source has to be cited.

And as stated in Section 4.6.1, each source used and considered during each task in the terminology building process shall be referenced and the recommended practice for dealing with copyright issues (listed below) shall be followed.

- Exact definitions - Copyright applies
- Equivalences mapping - Classed as conceptual work - This is typically not copyrightable, and a safe method.
- Add metadata from sources - references required - place in source field.
- Publish ontology - if in the public domain, reuse needs to be considered - Avoid redistributing existing text, try to share only your paraphrased or structured data.

The copyright and information sources shall be defined in the concept definition phase (Section 4) but shall also be reviewed in the review and approval process (Section 5).

4.6 Languages

4.6.1 Multi-language use

Depending on the application, it can be necessary that terms, labels and descriptions are to be provided in multiple languages. Following ISO 12616-1:2021, "terminology work is not to be confused or equated with the process of translation, which is concerned with rendering source language content into target language content. The translation process typically involves two languages. Terminology work can be monolingual, bilingual, or multilingual. It is focused on concepts and their designations in a specific subject field. While terminology work aims at understanding and describing concepts in a way that is independent of individual contexts of communication, translation is always dependent on contextual and situational meaning."

When defining concepts in multiple languages, concept orientation should be preserved. According to ISO 12616-1:2021, this means that "in contrast to an entry in lexical resources, such as traditional dictionaries, a concept entry shall contain information about one concept". Terms and labels should be defined following their conceptual equivalence and not via translation. Furthermore, ISO 12616-1:2021 states that "in translation-oriented terminology work, one definition in a language that is understood by all involved translators is often sufficient. When definitions for two or more cultures are necessary, it is generally not recommended to translate definitions. Instead, the terminology worker should research the concept and its designations in each language or culture. It can be necessary to translate definitions in specific settings, e.g. in international standardisation, multilingual legislation, or in-house standardisation of company-specific terminology, or when introducing new terms and loan terms to a target language where no definition exists."

4.6.2 Handling synonyms

A synonym is a term in one language that designates the same concept as another term in that language, including domain-specific variants used across different materials science subfields, see ISO 860. All terms are to be documented in one concept-oriented entry with a stable concept identifier. For each language, a section can be added for terms to be used. One preferred term per language should be defined (if one exists), and any admitted or deprecated synonyms in the same entry should be documented. Each term should carry at least the term string, a language tag (e.g., en-GB, de-DE, fr-FR), usage status (preferred/alternative), and part of the domain within materials science where it is used like that.

4.6.3 Handling ambiguities

If ambiguities occur anyway, the relevant concepts should be identified and revised. To separate concepts, preferred labels can be added/extended and the domains should be noted in which these concepts are valid. Additionally, updates made to such concepts should be marked, for example, using references like "replaced-by".

In general, ambiguities can be avoided if the review process (see Section 5) is conducted thoroughly.

4.7 Arrangement of terminology entries

For documentation purposes and further working with terminology, information about concepts shall be collected in one form. Table 1 shows such a form in accordance with ISO 10241-1.

Table 1 — Structure proposal of a terminology entry and description

Field	Description
Entry number	Unique identifier for the concept in the document
Term/Concept ID	Term/Concept ID used in this context and the language used
Pref label	Preferred Label
Alt label	Alternative Label
Definition	The concise description for the concept according to Section 4.4
Source	Meta data referring to the sources of the definition - where adapted or adopted
Language	Look at the languages by the parties for this concept and include these in the format
Notes/Examples	Additional illustrations or explanations to improve understanding of the concept.

5 Workflows for review and approval

5.1 Review and approval - general

5.1.1 Introduction

After a concept has been defined (see Section 4), it shall undergo a comprehensive review process before it can be published and used by a broader community. The review process is important as it ensures that the terminology is unique and well-defined, thereby minimising ambiguity and promoting consistency across the field. The review workflow consists of four stages and three different paths, see Figure 1. The paths depend on whether a concept already exists in recognised terminology databases and, if so, whether the existing concept definition aligns with the definition of the proposed concept. These are:

— Path A: Reuse of an existing concept (if the concept exists and fits)

— Path B: Adaptation of an existing concept (if the concept exists but does not fit perfectly)

— Path C: Creation of a new terminological entry (if the concept does not exist or existing concepts do not fit)

Before the correct path can be identified in the second stage, the first stage (due diligence) shall be accomplished. After a path has been identified and followed, the concept will undergo review and approval in the third and fourth stages. Figure 1 shows the complete review workflow as it is suggested by this CWA. The different stages are described in detail in the following. Most of the rules originate from ISO 704 and have been modified for the use in this CWA.

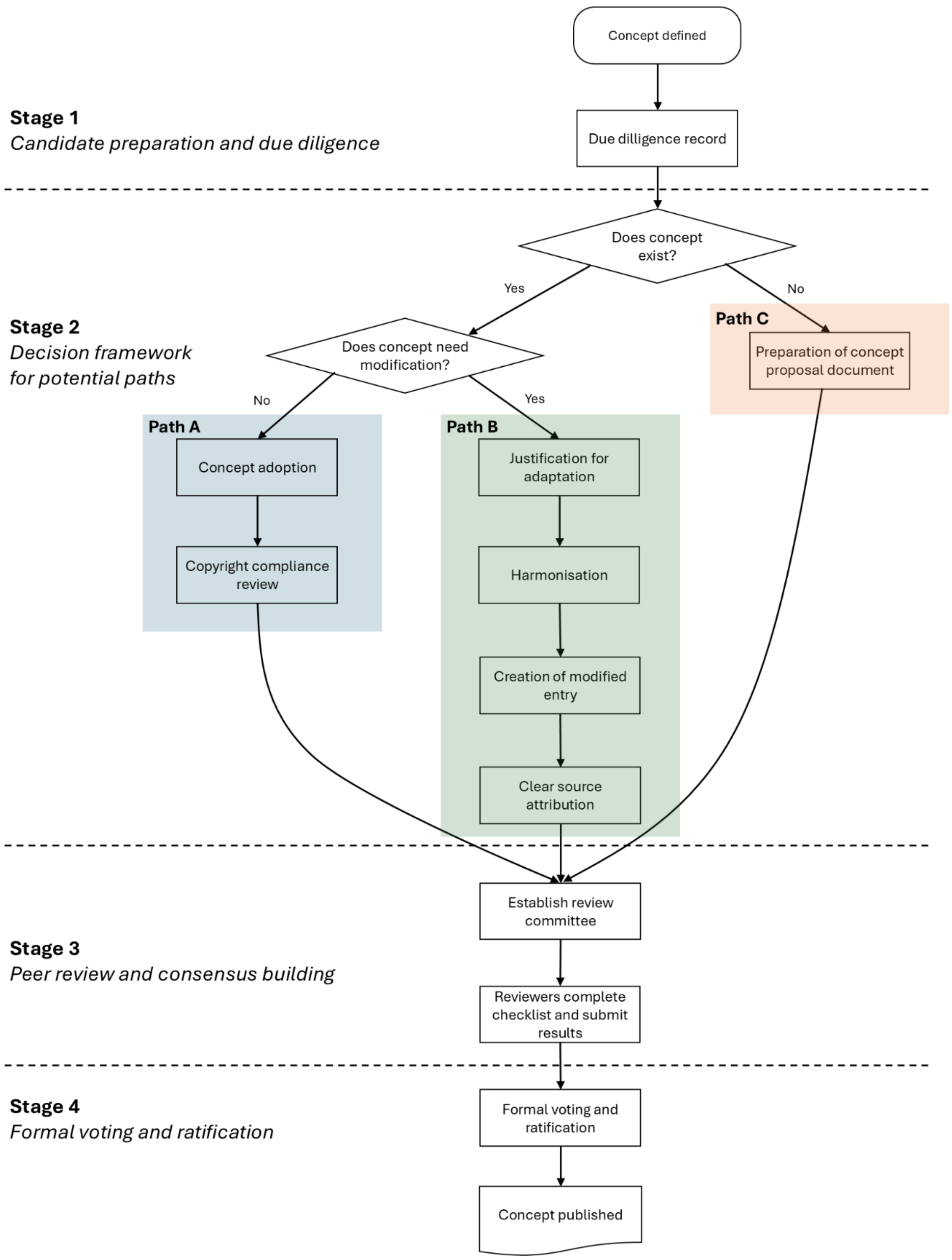

Figure 1— Workflow diagram for terminology review and approval

5.1.2 Stage 1: Candidate preparation and due diligence

After a concept has been defined (see Section 4), a systematic search in recognised terminology databases has to be conducted and documented (due diligence). The documentation shall include the specific keywords and search queries used for each database and a summary of the findings. If relevant concepts were found, they must be listed along with their source and an analysis of why they are or are not suitable for the present purpose. If no relevant concepts were found, this must also be explicitly stated. Table 2 shows a documentation form for due diligence.

Table 2 — Documentation form for due diligence

Administrative Information	
Date of Search	
Conducted by	*(Name and Organisation, address)*
Search details	
Database names	*(List of recognised terminology databases)*
Keywords used	*(List of keywords)*
Search queries	*(Exact queries conducted using the keywords above)*
Access date	*(Date when search was conducted)*
Summary of findings	
Term 1	*(Concept, source, definition, analysis of suitability)*
Term 2	*(Concept, source, definition, analysis of suitability)*
…	
Conclusion	
Were relevant terms found?	[ ] Yes [ ] No

5.1.3 Stage 2: Decision framework for potential paths

Depending on the outcome of the due diligence process (stage 1), Path A, B, or C shall be followed. The framework proposed prioritises the reuse of existing standards and imposes increasing levels of scrutiny for adaptation and the creation of new concepts. A detailed description of the potential paths and the decision to be made follows.

5.1.3.1 Path A: Reuse of an existing concept

If the due diligence identifies a concept in an existing database whose definition aligns with the proposed concept, the procedure for the reuse of an existing concept has to be followed. Consequently, the definition, and any accompanying explanatory information shall be adopted verbatim, without any modification. Additionally, the source of the existing term is to be cited. Moreover, the copyright compliance has to be reviewed, see Section 5.1.6.

5.1.3.2 Path B: Adaptation of an existing concept

This path is taken when a closely related concept exists, but its definition is either too broad, too narrow, or lacks a specific characteristic required. The following steps shall be taken (cf. ISO 860):

1. Justification for adaptation: A robust and explicit justification documenting is required to be provided to justify why the existing concept is not sufficient.

2. Harmonisation analysis: An analysis is to be conducted, which involves comparing the intentions of the source concept and the proposed concept to identify the differences and justify modifications. This analysis is governed by the concept harmonisation as specified in ISO 860.
3. Creation of modified entry: A new terminological entry is created. The definition is modified to reflect the harmonised concept.
4. Clear source attribution: The modified entry shall cite the source of the existing term and shall clearly indicate that it is an adaptation.

5.1.3.3 Path C: Creation of a new terminological entry

This path shall be followed when the due diligence process (stage 1) has conclusively shown that no suitable concept exists for reuse or adaptation. The path involves introducing a novel concept into a system. Therefore, a concept proposal sheet has to be prepared according to the required fields defined in Section 4 of this CWA. Additionally, the concept's intention and place within the system shall be explained.

5.1.4 Stage 3: Peer review and consensus building

All proposals, regardless of their development path (A, B, or C), shall undergo a formal peer review process to ensure quality, consistency, and community consensus.

5.1.4.1 The review committee

A review committee should be established based on the system where the concept will be introduced. The core composition should be domain experts from the primary field, complemented by specialists from related domains to prevent blind spots and reduce conflicts. To strengthen legitimacy and consensus, representatives from relevant communities may also be included. The committee's goal is to ensure rigorous, balanced, and well-supported reviews.

To ensure quality, expert requirements can be, for example:

- Demonstrated confidence and competence in their field
- Evidence of expertise (e.g., authored scientific papers, notable projects, certifications)
- A structured self-assessment, e.g. including a checklist rating key competencies from 1–10
- Commitment to impartiality and disclosure of potential conflicts of interest

5.1.4.2 Review criteria

To ensure that all reviews are objective, consistent, and grounded in established best practices, reviewers shall use a formal checklist. Table 3 shows such a checklist.

Table 3 — Review checklist

Organisational information	
Reviewer	
Date of review	
Term to be reviewed	
Due diligence	
Has a sufficient search in recognised terminology databases been documented?	

Is the justification for not reusing an existing term (if applicable) adequate?	
Concept analysis	
Is the definition complete, intensional, and differentiating (per Section 4.5)?	
Is the term's intention and place within the system clearly explained?	
Is the source attributed correctly?	
Is exactly one preferred label provided per required language, with any synonyms marked?	
If multilingual, are labels/definitions compliant with Section 4.6 (concept-orientation, language tags, regional variants where relevant)?	
Are sources and rights complete (citation details; reuse/adaptation indicated; permission if verbatim)?	
Conclusion	
Overall recommendation	[ ] Approved [ ] Approved with minor revision [ ] Major revisions required [ ] Rejected
Comments	

5.1.4.3 Review results

Reviewers submit their completed checklists and any additional comments. It is required to formally respond to every comment and suggested revision. A proposal is considered ready for a formal vote only after the Review Committee agrees that all substantive comments have been adequately addressed.

5.1.5 Stage 4: Formal voting and ratification

Once a proposal has successfully passed the peer review and consensus-building stage, it shall be submitted for a formal vote by all users of the system. The voting process is to be conducted over a defined period. For a vote to be valid, a minimum of 50% has to cast a vote. For a concept proposal to be approved, it is required to receive a two-thirds majority of the votes cast. The results of the vote (number of approve, disapprove, and abstain votes) shall be published for transparency reasons.

After voting, a unique, persistent identifier is assigned to the new terminological entry. The entry in the official terminology resource will then be finalised. This includes recording comprehensive metadata, such as the version number, the date of approval, and a link to the final proposal and voting record. This creates a complete and auditable history for each concept, fulfilling the need for comprehensive documentation. Ultimately, the approved concept is published in the official terminology collection.

5.1.6 Legal aspects

The reuse of terminology from existing standards is a cornerstone of this workflow, but it shall be managed with strict attention to copyright and intellectual property rights. Concept sources like standards are often copyright-protected works. The unauthorised reproduction of their content, including definitions, within another publication can constitute copyright infringement. This presents a conflict between the desire for usability (by including definitions directly in the term entries) and the need for legal compliance. Depending on the source, copying extracts for inclusion in other documents can require correct acknowledgement and may require formal authorisation as well as prior written permission, see Section 4.5.3 for further details.

5.2 Automated review and approval processes

While the core intellectual tasks of concept definition, writing, and consensus-building has to remain human-led, automation can play a significant role in supporting and streamlining the review process. Automation should be viewed not merely as a tool for efficiency, but as part of the quality assurance process. Its primary application is in Stage 1 of the review process. This includes automatic database lookups, standardised due diligence reporting, consistency checking (duplicate entries, consistent use of language, similarity to existing concepts). By automating the most repetitive and time-consuming part of the process, it is ensured that every single review process undergoes an identical and comprehensive level of scrutiny before a concept is presented to human reviewers. This human-in-the-loop model leverages the strengths of both automation (for speed, consistency, and comprehensiveness in data gathering) and human expertise (for nuance, judgment, and consensus-building).

6 Guidelines for technical implementation

6.1 Storing terminology

Storage and access of the terminology should be compatible with various data management practices: databases, basic folder structures, ontologies, etc. It should be generic enough to be supported by at least two different major (electronic) storage platforms, and where possible, based on open standards.

Terminology collections shall be stored in a manner that is compatible with diverse data management practices, including structured databases, ontologies, and simple folder systems. Storage and access methods *should* be generic, based on open standards, supported by multiple software platforms, and should provide access via stable protocols such as server endpoints (e.g., API, SPARQL endpoint). To facilitate storage-platform independence and data exchange, the use of a standardized interchange format, such as ISO 30042 (TermBase eXchange - TBX), is recommended. It is also recommended that terminology developed for research and standardization adheres to the FAIR principles by providing rich, open, and machine-readable metadata, even when commercial practices might otherwise restrict such access.

NOTE For more information on *TBX*, see [1] and on *Linked Open Vocabularies*, see [2].

6.2 Logical structuring

The logical data model for terminology shall be concept-oriented, as specified in Section 4 and in accordance with ISO 704. This model separates the concept from its linguistic and administrative representations. The model shall also support the definition of interrelationships between concepts, including hierarchical (generic/specific) and associative relations. To represent this rich structure for integration with formal ontologies, a semantic model should be used. The W3C OntoLex-Lemon model is recommended as it provides a standard vocabulary for linking terminological entries to formal ontology concepts within an RDF/OWL environment.

NOTE For more information on *OntoLex*, see [3]

6.3 Technical implementation

The technical implementation ensures that the stored and structured terminology is usable in practice. This involves a variety of tools, from simple spreadsheets to dedicated ontology editors and enterprise terminology management systems. Terminology and ontology artefacts should be managed in and retrieved from databases that provide robust search and retrieval. These include national and international terminology portals (e.g., DINterm, ISO Online Browsing Platform), ontology repository platforms such as those in the Ontoportal alliance (e.g., matportal.org), and lookup services (e.g., EBI Ontology Lookup Service). To ensure interoperability between these tools and repositories, implementations shall rely on the standardized formats defined in 6.1 (e.g., TBX) and 6.2 (e.g., OntoLex-based RDF/OWL).

Following the principles of automated support outlined in 5.2, automation should be used to support the terminology lifecycle. A recommended implementation is the integration of AI-assisted services, such as Generative AI with a Retrieval-Augmented Generation (RAG) backend. This service's RAG backend should be populated with authoritative sources, including the portals and repositories described above. By connecting to these services, the AI can automate the "due diligence" search for existing concepts, terms, and ontology elements. This integrated service *can* then be embedded into authoring environments. For example, in NISO STS-aware XML editors (used by SDOs) or in ontology editors and domain-specific software (e.g., LIMS), the service can assist authors by proactively finding existing terms, suggesting definitions, and ensuring consistency at the point of creation.

It is essential that both exchange and storage formats adhere to open, well-documented standards that support machine readability and semantic precision. This would ensure the interoperability, accessibility, and long-term sustainability of terminological resources. Rich metadata, versioning, and multilingual content are of great value for terminology exchange formats. For this reason, using ISO 30042 (TermBase eXchange, TBX) is recommended for structured term exchange between management systems. This provides an XML-based schema for the representation of terminological data and the associated metadata.

Another requirement for terminological data is that it should be convertible to, or natively represented in, RDF or OWL formats. This makes the ontology compatible with the Semantic Web and enables linking of terminologies to formal semantic resources by default, thereby enhancing reusability across systems. Data provenance and versioning, with metadata such as creation dates, modification history, and authorship, should be captured in all representations to maintain traceability and auditability over time.

Terminological databases should employ storage architectures that are both structured and interoperable. Relational database systems are a great option for traditional termbases, while graph-based databases such as RDF triple stores or property graphs are instead recommended for ontology-driven data.

However, independent of the storage solution, data should remain exportable in open, machine-readable formats (such as TBX, RDF/XML, JSON-LD) to guarantee platform independence and facilitate long-term preservation.

NOTE 1 For more information on the *Plattform Material Digital* and its ontology collection, see [4] and [5].

NOTE 2 For further information on the *Industrial Ontologies Foundry (IOF)* and its *Industrial Data Ontology*, see [6] and [7].

NOTE 3 For more information on the *Object Oriented Linked Data Schema (OO-LD)*, see [8].

NOTE 4 The *DIN Terminology Portal* is accessible via DIN’s website, see [9].

Annex A
(informative)

FAIR principles (Findability, Accessibility, Interoperability, Reusability)

A.1 General

The guidelines summarised here are grouped below into the four FAIR principles: Findability, Accessibility, Interoperability, and Reusability. There is significant overlap between these four principles, but this grouping should provide clarity on the purpose of each guideline.

NOTE For more in depth information on the *FAIR principles* see [10] or [11].

A.2 Findability

The need for findability has more to do with terminology management practices than with the terms themselves. A domain-specific term is assumed to be findable, provided it is unique. Since the same (or a very similar) term may be used to assign different concepts across different domains (and different domain ontologies), finding the intended term within the intended domain requires that each term can be differentiated from other, identical or similar terms from other domains.

As for guidelines regarding terminology schema/format, the key requirement is therefore the existence of a global unique identifier (GUID) for all terms within the ontology. This should be a widely accepted cross-company, cross-consortium, cross-industry naming convention, specific to the domain(s) of application, with consideration for how widely the terminology is to be shared.

The logic and convention of the GUID must be well-documented. Consistency may be programmatically enforced in the term generation process itself, when automated. However, in most cases, due to manual interventions, it is going to be more important to rely on processes that define who and when assigns the unique identifier.

A.3 Accessibility

Similarly to findability, this is a terminology management problem.

The format/schema used for storing terms should always allow for changes in storage requirements. In other words, the ability for users to access the terminology should not depend on how or where it is physically stored.

Good metadata should always be stored: date of creation, date of modification, software or tooling used, department/person responsible, etc. This may be quite specific for the domain in question.

A.4 Interoperability

One key aspect of interoperability is human and machine readability. Many formats cannot be read by a human either due to being too complex, lacking documentation to know which fields represent what information, or simply being stored in compressed and/or protected files.

Documentation needs to be clear in what fields correspond to what parameters in the terminology, as it is stored, and how these parameters are defined, and this must be readily available information to any user of the terminology.

For any automation of the terminology, it must be stored in such a way that it is easily usable by a programming language – ideally as a dictionary/object/SQL, as in those cases the parameters are less likely to get confused than in a position-based reference system (e.g. .csv). Explicit variables also allow for easy expansions of the format/schema if needed.

The storage format must either be fixed long-term or a versioning system with documentation must exist. This enables automated tools to be built with the data format as specified.

Another issue is the danger of data loss. Sometimes the data exported from one software is not comprehensive. Sometimes fidelity of data may be lost due to formats used.

In summary, for interoperability, these are the key requirements:

- Automation
- Reliability
- Documentation
- No data loss
- Generation from primitives defined

A.5 Reusability

There is a significant overlap between the listed requirements, so reusability is already enabled by some requirements listed above.

The validity of any data should also be traceable. An audit log of some sort should be optionally available to trace when existing terminologies are being changed.

When terminologies are easily traceable, with reference to original and related definitions, this can later be revised to understand terms in common use.

Annex B
(informative)

Usage of concepts within ontologies and knowledge graphs

The digital transformation of materials science generates vast, heterogeneous data from simulations, experiments, calculations, and publications that have to be coherently managed to ensure FAIR (Findable, Accessible, Interoperable, Reusable) data practices. Semantic technologies—particularly ontologies and knowledge graphs—enable this integration by providing unified, machine-actionable representations of materials and processes. Ontologies define the conceptual framework of classes, relations, and constraints, while knowledge graphs instantiate these concepts with real-world entities and relationships. Consistent, harmonised terminology underpins both, ensuring semantic integrity across heterogeneous datasets.

Knowledge graphs (KGs) serve as the operational backbone of semantic data infrastructures, supporting interoperability, analytics, and machine learning. Two principal paradigms are employed: RDF-based graphs, built on standards such as RDF, OWL, and SKOS, which ensure compatibility with Linked Data ecosystems; and property-labelled graphs, which provide node–edge structures with key–value attributes enabling efficient traversal, schema evolution, and integration of complex datasets. RDF models ensure semantic precision and standardisation, whereas property graphs offer performance and exploratory flexibility.

To operationalise ontologies, KGs incorporate versioning and provenance metadata for traceability, support querying through SPARQL, Cypher, or Gremlin, and enable data exchange through standard serialisations (RDF/XML, JSON-LD, CSV). Visualisation interfaces further facilitate exploration and validation of relationships among materials, processes, and properties.

Several domain-specific KGs exemplify these practices. PropNet encodes material-property relationships for correlation discovery and feature generation. NanoMine unifies experimental, compositional, and processing data within a semantic framework (SIO, PROV-O, Schema.org) to ensure interoperability and provenance tracking. MatKG demonstrates large-scale automation using natural language processing and graph embeddings, integrating millions of RDF triples across materials, properties, and synthesis data to support link prediction and semantic discovery.

Implementation of such infrastructures relies on open, standards-compliant technologies. Graph databases such as Neo4j, ArangoDB, and OrientDB provide performant storage and querying capabilities, while semantic frameworks including Apache Jena, Protégé, and the OWL API support ontology creation, reasoning, and export in RDF/OWL formats. Visualisation and dashboard tools (e.g., NeoDash, Bloom) enable interactive inspection of graph content.

By integrating these components—ontology engineering, graph-based instantiation, and interactive access—materials science organisations can publish interoperable, versioned, and semantically consistent resources. These infrastructures enhance the discoverability, interoperability, and reuse of materials data, aligning terminology management with FAIR principles and supporting advanced analytics and AI-assisted materials design.

Bibliography

[1] **Wikipedia.** TermBase eXchange. [Online] 05 November 2025. https://en.wikipedia.org/wiki/TermBase_eXchange.

[2] **Vandenbussche, Pierre-Yves and Vatand, Bernard.** Linked Open Vocabularies (LOV). [Online] 05 November 2025. https://lov.linkeddata.es/dataset/lov/.

[3] **Wikipedia.** OntoLex. [Online] 05 November 2025. https://en.wikipedia.org/wiki/OntoLex.

[4] **Plattform Material Digital.** Internetauftritt der Plattform MaterialDigital-Website. [Online] 05 November 2025. https://www.materialdigital.de/.

[5] —. Material Digital Ontology Collection. *GitHub/materialdigital.* [Online] 5. November 2025. https://github.com/materialdigital/materialdigital1_ontology_collection.

[6] **EDM Council.** IOF Ontology. [Online] 05 November 2025. https://spec.industrialontologies.org/iof/.

[7] **Klüwer, Johan Wilhelm, et al.** Industrial Data Ontology. *PCA Reference Data and Services.* [Online] 05. November 2025. https://rds.posccaesar.org/ontology/lis14/ont/core/.

[8] **Stier, Simon und Gold, Lukas.** The Object Oriented Linked Data Schema. *GitHub/OO-LD/schema.* [Online] 05. November 2025. https://github.com/OO-LD/schema.

[9] **DIN Deutsches Institut für Normung e.V.** DIN-TERMinology Portal. *German Institute for Standardisation Website.* [Online] 05. November 2025. https://www.din.de/en/services/din-term.

[10] **GO FAIR International Support and Coordination Office (GFISCO) .** FAIR Principles. *GO FAIR Website.* [Online] 2025. https://www.go-fair.org/fair-principles/.

[11] *FAIR Principles: Interpretations and Implementation Considerations (J/OL).* **Annika Jacobsen, Ricardo de Miranda Azevedo, Nick Juty, et al.** s.l. : Beijing Zhongke Journal Publising Co. Ltd., 2020, Data Intelligence, Bde. 2 (1-2), S. 10-29.