INTERNATIONAL STANDARD

ISO 26162-1

First edition
2019-11

Management of terminology resources — Terminology databases —

Part 1: Design

Gestion des ressources terminologiques — Bases de données terminologiques —

Partie 1: Conception

Reference number
ISO 26162-1:2019(E)

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword .... **iv**
Introduction .... **v**
1 Scope **1**
2 Normative references **1**
3 Terms and definitions **1**
3.1 Concepts 1
3.2 Terminology databases 2
4 Terminology database design **4**
4.1 General 4
4.2 Terminological metamodel 5
4.3 Data categories 7
4.3.1 General 7
4.3.2 Types of data categories 8
4.3.3 Shared resources 9
4.3.4 Concept relations 10
4.4 Concept entries 11
4.4.1 Concept orientation 11
4.4.2 Language 11
4.4.3 Dependency and repeatability of data categories 12
4.4.4 Data granularity 12
4.4.5 Data elementarity 13
4.4.6 Data-modeling variation 13
4.5 Roles 14
Annex A (informative) **Terminology database excerpt based on the terminological metamodel — Example** **15**
Bibliography **19**

Foreword

ISO (the International Organization for Standardization) is a worldwide federation of national standards bodies (ISO member bodies). The work of preparing International Standards is normally carried out through ISO technical committees. Each member body interested in a subject for which a technical committee has been established has the right to be represented on that committee. International organizations, governmental and non-governmental, in liaison with ISO, also take part in the work. ISO collaborates closely with the International Electrotechnical Commission (IEC) on all matters of electrotechnical standardization.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of ISO documents should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Technical Committee ISO/TC 37, *Language and terminology*, Subcommittee SC 3, *Management of terminology resources*.

This first edition of ISO 26162-1, together with ISO 26162-2, cancels and replaces ISO 26162:2012, which has been technically revised.

The main changes compared to the previous edition are as follows:

— the document has been split into parts. The first part is focusing on the design of terminology database design, the second part on the development of terminology management systems;

— all references to generic software design principles and specific use cases have been removed.

A list of all parts of the ISO 26162 series can be found on the ISO website.

Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

Terminologies are the totality of concepts in given subject fields represented by terms and other designations and described by using additional terminological data. In general, these data are organized in structured terminology databases and are usually manipulated in specific software applications called terminology management systems. Terminology databases usually vary with regard to their underlying data models and consist of different sets of data categories, while terminology management systems generally differ depending on their functionality and the platform they are designed for.

The ISO 26162 series gives guidance on designing terminology databases and on essential terminology management system features. The series can also be used to evaluate the conformance and suitability of terminology databases and terminology management systems.

Management of terminology resources — Terminology databases —

Part 1: Design

1 Scope

This document specifies general, i.e. implementation- and use-case-independent terminology database design principles to enable maximum efficiency and quality in terminology work. Thus, this document supports creating, processing, and using high quality terminology. The intended audiences of this document are terminologists, translators, interpreters, technical communicators, language planners, subject field experts, and terminology management system developers.

This document describes a maximum approach, i.e. terminology database design for distributed, multilingual terminology management. It can also be used for designing smaller solutions.

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 1087, *Terminology work — Vocabulary*

ISO 12620, *Management of terminology resources — Data category specifications*

ISO 16642:2017, *Computer applications in terminology — Terminological markup framework*

ISO 23185, *Assessment and benchmarking of terminological resources — General concepts, principles and requirements*

3 Terms and definitions

For the purposes of this document, the terms and definitions given in ISO 1087 and the following apply.

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

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

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

3.1 Concepts

3.1.1
object
anything perceivable or conceivable

Note 1 to entry: Objects can be material (e.g. an engine, a sheet of paper, a diamond), immaterial (e.g. a conversion ratio, a project plan) or imagined (e.g. a unicorn, a scientific hypothesis).

Note 2 to entry: Objects can undergo changes which cause conceptual or designation change.

[SOURCE: ISO 1087:2019, 3.1.1, modified — Note 2 to entry added.]

3.1.2
concept
unit of knowledge created by a unique combination of characteristics

Note 1 to entry: Concepts are not necessarily bound to particular natural languages. They are, however, influenced by the social or cultural background which often leads to different categorizations.

Note 2 to entry: Due to their dynamic nature, concepts are also defined as units of thinking (see ISO 704:2009, 5.1 and DIN 2342:2011-08, 4.1).

[SOURCE: ISO 1087:2019, 3.2.7, modified — former Note 2 to entry removed and replaced by a new Note 2 to entry.]

3.1.3
designation
designator
representation of a *concept* (3.1.2) by a sign which denotes it in a domain or subject

Note 1 to entry: A designation can be linguistic or non-linguistic. It can consist of various types of characters, but also punctuation marks such as hyphens and parentheses, governed by domain-, subject-, or language-specific conventions.

Note 2 to entry: A designation may be a *term* (3.1.4) including appellations, a proper name, or a symbol.

[SOURCE: ISO 1087:2019, 3.4.1]

3.1.4
term
designation (3.1.3) that represents a general concept by linguistic means

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

Note 1 to entry: Terms may be partly or wholly verbal.

[SOURCE: ISO 1087:2019, 3.4.2]

3.2 Terminology databases

3.2.1
terminology database
termbase
database comprising a *terminological data collection* (3.2.4)

[SOURCE: ISO 30042:2019, 3.28, modified — admitted term "terminology database" made preferred term and preferred term "termbase" made admitted term.]

3.2.2
data model
graphical and/or lexical representation of data, specifying their properties, structure, and inter-relationships

[SOURCE: ISO/IEC 11179-1:2015, 3.2.7]

3.2.3
terminological metamodel
data model (3.2.2) that describes the basis for designing and implementing *terminological data collections* (3.2.4)

3.2.4
terminological data collection
TDC
resource consisting of *concept entries* (3.2.7) with associated metadata and documentary information

[SOURCE: ISO 16642:2017, 3.21, modified — "terminological entries" replaced by "concept entries".]

3.2.5
global information
GI
technical and administrative information applying to the entire *terminological data collection* (3.2.4)

EXAMPLE The title of the terminological data collection, revision history, owner or copyright information.

[SOURCE: ISO 16642:2017, 3.11, modified — "Note 1 to entry" replaced by "EXAMPLE"; "For example," removed in the example.]

3.2.6
complementary information
CI
information supplementary to that described in *concept entries* (3.2.7) and shared across the *terminological data collection* (3.2.4)

EXAMPLE Domain hierarchies, institution descriptions, bibliographic references, and references to text corpora.

[SOURCE: ISO 16642:2017, 3.2, modified — "terminological entries" replaced by "concept entries" within definition; "Note 1 to entry" replaced by "EXAMPLE"; "are typical examples of complementary information" removed in the example.]

3.2.7
concept entry
CE
terminological entry
part of a *terminological data collection* (3.2.4) which contains the terminological data related to one *concept* (3.1.2)

[SOURCE: ISO 16642:2017, 3.22, modified — "concept entry" and acronym "CE" added as preferred terms; preferred term "terminological entry" made admitted term; preferred term "TE" removed; Note 1 to entry removed.]

3.2.8
language section
LS
part of a *concept entry* (3.2.7) containing information related to one language

[SOURCE: ISO 16642:2017, 3.13, modified — "terminological entry" replaced by "concept entry"; Note 1 to entry removed.]

3.2.9
term section
TS
part of a *language section* (3.2.8) containing information about a *term* (3.1.4)

[SOURCE: ISO 16642:2017, 3.20, modified — "giving" replaced by "containing".]

3.2.10
term component section
TCS
part of a *term section* (3.2.9) containing linguistic information about the components of a *term* (3.1.4)

[SOURCE: ISO 16642:2017, 3.19, modified — "giving" replaced by "containing.]

3.2.11
data category
class of data items that are closely related from a formal or semantic point of view

EXAMPLE /part of speech/, /subject field/, /definition/.

Note 1 to entry: A data category can be viewed as a generalization of the notion of a field in a database.

Note 2 to entry: In running text, such as in this document, data category names are enclosed in forward slashes (e.g. /part of speech/).

[SOURCE: ISO 12620:2019, 3.2, modified — preferred term "DC" removed".]

3.2.12
repeatability
principle whereby a *data category* (3.2.11) can be repeated within a database definition and whereby it can also be combined with other data categories

3.2.13
concept orientation
principle whereby a *concept entry* (3.2.7) describes a single *concept* (3.1.2)

Note 1 to entry: When two or more different concepts are represented by the same designation (in the same language), this designation is considered a homograph. Such concepts are documented in separate concept entries.

3.2.14
term autonomy
principle whereby all *terms* (3.1.4) in a *concept entry* (3.2.7) are considered independent sub-units and can be described using the same set of *data categories* (3.2.11)

3.2.15
data granularity
degree of data precision

EXAMPLE The set of individual data categories /part of speech/, /grammatical gender/, and /grammatical number/ provides for greater data granularity than does the single data category /grammar/.

3.2.16
data elementarity
principle whereby a data field contains only one data element

EXAMPLE For example, including both a full form and an abbreviation of a term in the same data field would be a violation of data elementarity.

3.2.17
data-modeling variation
variation in *data models* (3.2.2) describing the same information

4 Terminology database design

4.1 General

Terminology database design requires a deep understanding of terminology theory and terminology work. In this sense, and to achieve high quality results, the following shall be used:

— established terms and definitions as specified in ISO 1087;

— principles and methods as specified in ISO 704;

— data-modeling criteria as specified in ISO 16642 and ISO 12620;

— usability metrics as specified in ISO 23185.

Terminology databases have a logical structure that is reflected in a fundamental hierarchical data model (as described in 4.2) containing various levels at which data categories (see 4.3) can be anchored. This data-modeling approach provides the necessary flexibility, since the design of a terminology database is always subject to specific work profiles (terminology work, technical communication, translation, etc.) and to organizational needs (freelancers, translation agencies, company or organization in-house departments, etc.). Thus, in the very early design process, a long-term and detailed management plan shall be defined, taking into consideration all possible user groups, as well as organizational and technical issues in order to avoid the need for substantial, time consuming and costly changes after concluding the design process.

4.2 Terminological metamodel

Terminology databases shall comply with the terminological metamodel (or a subset thereof) defined in ISO 16642:2017 (see Figure 1 and Figure 2). The essence of the metamodel constitutes the principle of concept orientation (see 4.4.1), i.e.:

— in a terminology database, an entry (concept entry = CE) describes a concept that can be further described using:

— additional sublevels for instantiating languages and language-specific information (language section = LS), further subdivided by:

— terms and term-specific information (term section = TS);

— individual words of a multiword term or one of the components of a single-word term, such as a morpheme (term component section = TCS).

Furthermore, the metamodel provides high-level containers that allow for documenting:

— global information (GI) that applies to the complete terminology database (name of the terminology database, institution or individual originating the file, copyright information, history, etc.);

— complementary information (CI) such as complete bibliographical or administrative information, binary data, picklist values or references to text corpora that are referenced from concept entries.

The above-mentioned levels of the terminological metamodel can be schematized as shown in Figure 1.

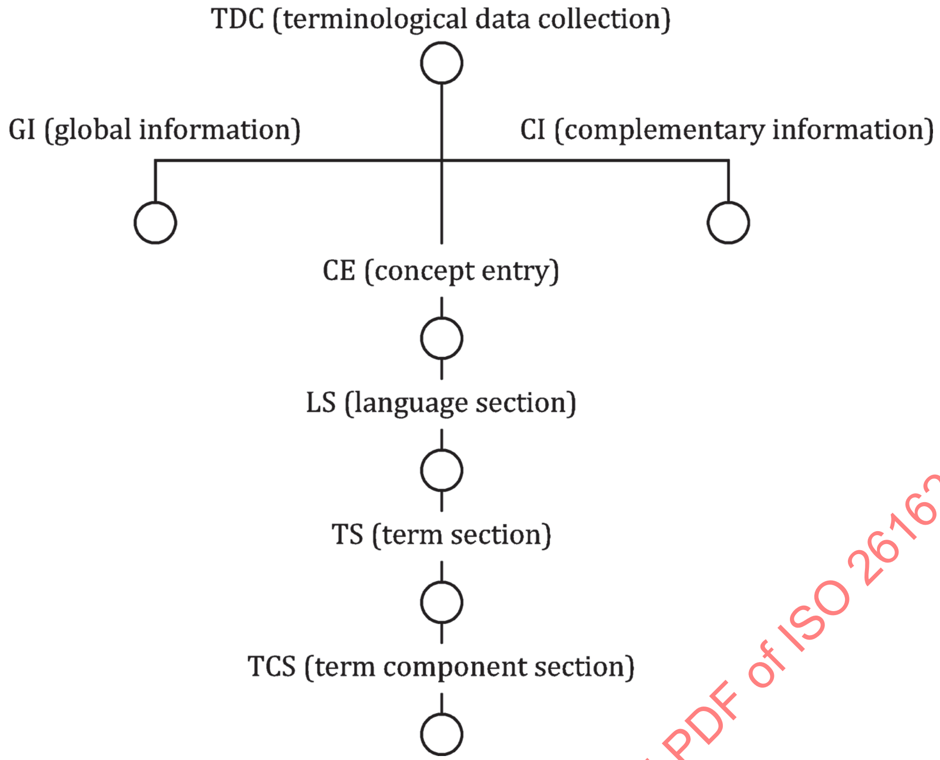
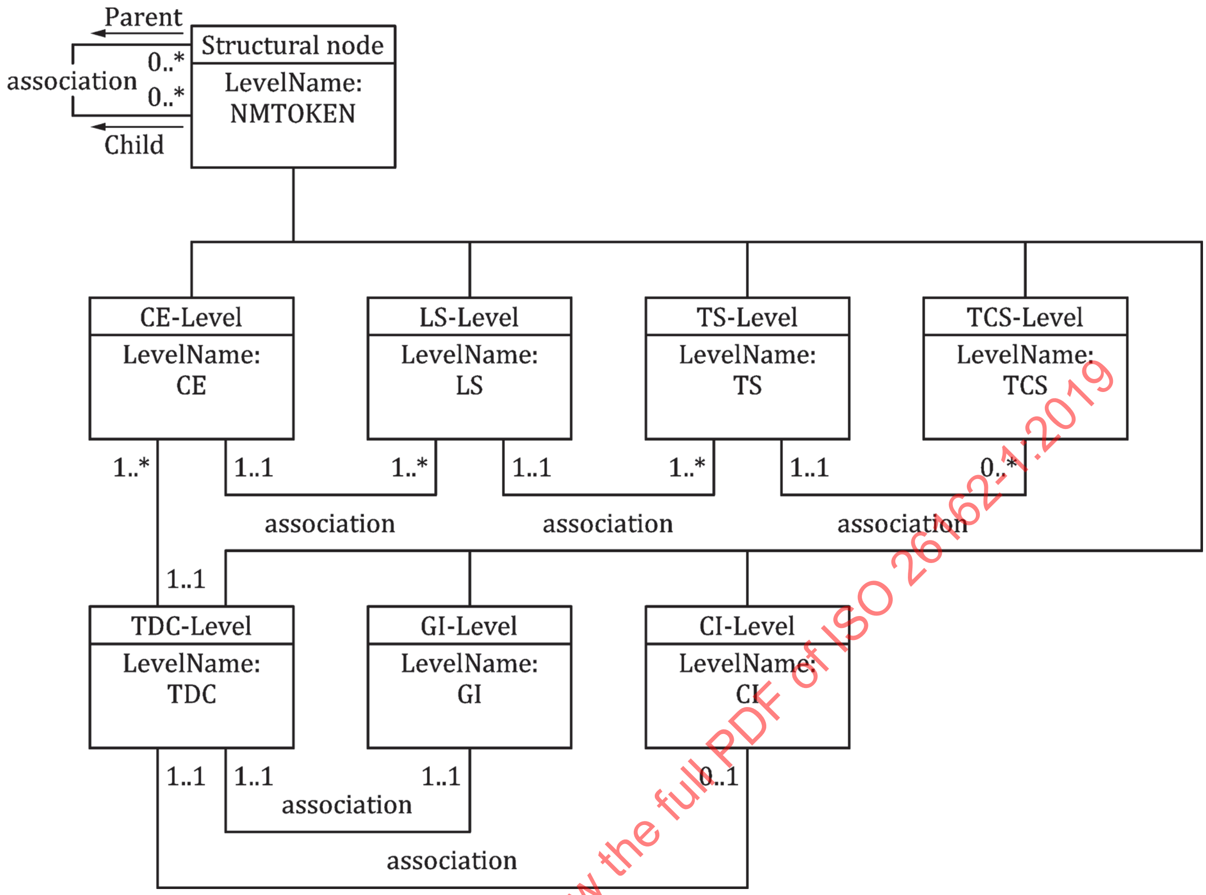


Figure 1 — Terminological metamodel — Schematic view (adapted from ISO 16642:2017)

The metamodel levels, their relationships and cardinalities can also be expressed using UML (Unified Modeling Language) as shown in Figure 2.



Cardinalities

1..1 = Shall occur exactly once.

1..* = Shall occur one or more times.

0..1 = May occur zero or one time.

0..* = May occur zero or more times.

Figure 2 — Terminological metamodel — UML diagram (adapted from ISO 16642:2017)

An example of a terminology database excerpt compliant with the terminological metamodel is given in Annex A.

4.3 Data categories

4.3.1 General

Concept entries are made up of specific units of information, such as terms, definitions and contexts, and each class of these units of information is identified by a data category. Over recent decades terminology practitioners have gathered, standardized, classified, and published relevant data categories that are stored in recognized data category repositories, such as DatCatInfo (see Reference [11]). For instance, /grammatical gender/ is a data category that would typically be used as a field name, and "feminine", "masculine", "neuter", etc. would occur as values of this field. Of course, terminology database designers can create and use their data categories according to their specific needs. However, care shall be taken when creating new data categories or when adapting the names for the data categories proposed by recognized data category repositories to avoid confusion or misinterpretation.

ISO 10241-1:2011 and ISO 12616:2002 include a list of common data categories used to document the following units of information:

— terms (in any desired language);

— term-specific information (such as grammatical attributes, term types, register, status, contexts, etc.);

— language-specific information (such as definitions, notes, etc.);

— concept-specific information (such as subject fields, definitions, examples, notes, graphics, etc.);

— administrative and bibliographical information:

  — identifiers of various sorts, such as to identify products or projects with which the terms are associated;

  — dates, names of people who created or modified the concept entry or parts of it;

  — entry status, for example "submitted", "working", "approved";

  — sources of terms, definitions, contexts, notes, etc.

Designers who need new data categories that are not included in recognized data category repositories shall adhere to the requirements for creating, documenting, harmonizing, and maintaining data category specifications as defined in ISO 12620.

4.3.2 Types of data categories

4.3.2.1 General

Terminology databases typically contain several different types of data categories depending on the kind of information they contain. Data categories can be divided into three categories, namely:

— open and closed data categories;

— mandatory, optional, system-generated, and default data categories;

— read-write, read-only, and hidden data categories.

Careful selection of data category types ensures higher quality content in a terminology database.

4.3.2.2 Open and closed data categories

Open data categories can contain any text. For instance, /term/ is considered open, because the actual term that can be recorded in the corresponding data field is unpredictable. Other examples of open data categories are /definition/ or /context/.

In contrast, closed data categories can only contain one or more of a finite set of permissible values. When documenting terms in the German language, for instance, /grammatical gender/ may only take the values "masculine", "feminine", or "neuter". Apart from /grammatical gender/, typical representatives of closed data categories are /part of speech/, /term type/, /geographical usage/, /administrative status/ or /subject field/.

Values themselves can constitute data categories such as /masculine/, for example, with the Boolean values "yes|no" or "true|false".

The use of picklists ensures that only predefined values can be selected, thus preventing the insertion of inadmissible values or misspelled or variant forms. For instance, left to their own devices, users might type "masculine", "masc.", or simply "m." for a masculine noun. Providing a uniform representation of these values ensures consistency throughout the terminology database, which is important for ensuring the performance of searching, filtering, data exchange, and other data management tasks.

4.3.2.3 Mandatory, optional, system-generated, and default data categories

Data categories can be either mandatory or optional. Typically, a terminology management system does not allow users to save concept entries if a mandatory field is empty. At least one language and one term in that language shall be required for each concept entry, unless the concept has yet to be named or it is represented by a diagram or a node in the concept diagram.

Other typical mandatory fields include information about the subject field or part of speech, both of which can be essential for differentiating homographs. However, designating certain data categories to be mandatory can sometimes be problematic. For instance, it can take considerable time and effort to find definitions and contexts. There can be even more complex conditions, such as forcing a source for a definition or the grammatical gender in case of nouns for a specific language section.

System-generated data fields are not inserted manually by the user, they are populated by the system. For instance, many terminology management systems automatically assign entry numbers to concept entries, as well as creation and modification dates, and the names of the users who created or modified the concept entry.

Part of terminology database design involves making decisions about the levels within the concept entry that will be documented with such administrative information. It is generally insufficient to record such administrative information only for the concept entry as a whole, because different people can be responsible for different parts of the concept entry, especially for different language sections.

It is also possible to set default values for certain data categories. For instance, if a user is going to document terms associated with a single project or source text, it can be convenient to allow the user to pre-set the values for the corresponding data categories /project subset/ or /source/ so that all concept entries created automatically contain those values.

4.3.2.4 Read-write, read-only, and hidden data categories

In a terminology database, it shall be possible to set different access levels for users, depending on the user's needs and role. A field that is visible and can be modified is a "read-write" field. A field that is visible but cannot be modified is a "read-only" field. Fields that are not needed by some users are usually hidden.

For instance, it can be desirable for a lead terminologist to have read-write access to all fields, especially to the concept entry status field, to set the corresponding value ("proposed", "under review", "approved", "deprecated", etc.). In general, this person is also responsible for assigning authorization levels to other users. Users editing concept entries for specific languages should only have read-write access to the fields in their language sections. Product developers, technical communicators, translators, and service and marketing staff can have read-write access to fields necessary to allow them to provide feedback to the lead terminologists, and read-only rights to the remaining fields. It is also frequently desirable to hide administrative fields such as "date" and "responsibility" to ensure that users have a less cluttered view of the data.

4.3.3 Shared resources

Some data point to resources that reside outside the concept entry, such as figures, audio, video, websites, full sets of bibliographical and personal data, references to text corpora, or other concept entries. These resources are called shared resources, because any one resource can be referenced from many concept entries.

For example, a field for figures is typically available at the concept level to allow for figures that represent the entire concept independently of language. A field for figures can also be provided at the language level to accommodate situations where the objects designated by the concept differ in appearance in different cultures or for images including callouts in a specific language. However, rather than repeating figures or complete personal or bibliographical data in the concept entry itself, it is more economical to create separate entries for full bibliographical or personal data sets and point to them from entries using identifiers. This practice also allows for updating bibliographical or personal references in a more efficient way.

Bibliographic data are used to document the source of definitions, terms, contexts, notes, figures, or websites (see Annex A). Personal data are used to record the users who create or modify a concept entry.

Shared resources can be stored in an external database or in a separate location in the terminology database itself. It shall be possible for users to view shared resources while browsing or searching terminology databases, because this information is often crucial for the evaluation of concept entries.

4.3.4 Concept relations

Relations between concepts shall be indicated where necessary. Hierarchical concept relations, such as superordinate and subordinate, are frequently used in prescriptive terminology databases where terminology standardization and the rigor of definitions is paramount, or for software applications that can use such hierarchical data, for instance, for knowledge management or retrieval. In this case, a separate data category will be required for each type of hierarchical relation. Such data categories shall allow the user to uniquely point to the target of the relation, which can be a concept entry identifier or term within a concept entry. Example 1[1)] illustrates generic relations for the concept *circuit-breaker*.

EXAMPLE 1

Concept entry identifier: 3 (for concept *circuit-breaker*)

  Superordinate concept generic: mechanical switching device

  Subordinate concept generic: air circuit-breaker

  Subordinate concept generic: oil circuit-breaker

  Subordinate concept generic: vacuum circuit-breaker

  Subordinate concept generic: SF_6 circuit-breaker

In Example 1, the concept *circuit-breaker* represented by the term "circuit-breaker" has a superordinate concept *mechanical switching device* and four subordinate concepts (*air circuit-breaker*, *oil circuit-breaker*, *vacuum circuit-breaker*, and *SF_6 circuit-breaker*). The data elements of the data fields "Superordinate concept generic" and "Subordinate concept generic" are linked and point to the corresponding concept entry.

Concept systems can also be constructed based on notational numbers designed to represent the position of a given concept within a concept system as shown in Example 2.

EXAMPLE 2

Concept entry identifier: 3 (for concept *circuit-breaker*)

  Concept position: 1.2

  Superordinate concept generic: mechanical switching device

    Concept position: 1

  Subordinate concept generic: air circuit-breaker

    Concept position: 1.2.1

  Subordinate concept generic: oil circuit-breaker

    Concept position: 1.2.2

  Subordinate concept generic: vacuum circuit-breaker

    Concept position: 1.2.3

1) For the examples in this document, indentation is used to visualize the Terminological Markup Framework metamodel levels and relationships between the instances of terminology databases.

Subordinate concept generic: SF_6 circuit-breaker

Concept position: 1.2.4

Concept systems can also comprise associative relations, such as spatial, sequential or causal relations. There are also non-hierarchical relations where the type of relation is not specified with the data category name as illustrated in Example 3.

EXAMPLE 3

Concept entry identifier: 3 (for concept *circuit-breaker*)

Concept position: 1.2

Related concept: release

Concept relations shall be expressed through designated data categories as shown in the previous examples and shall not be hidden in notes, examples or comments. This approach optimizes the ability for such relations to be machine-processable.

In some terminology databases the relations are embedded within other text fields, such as in definitions. Indeed, terms in definitions are often closely related to the concept being defined, and it is a popular practice to treat these terms as hyperlinks that lead to the related concept entry. Example 4 shows a link to the concept *mechanical switching device* from the definition for the concept *circuit-breaker*.

EXAMPLE 4

Concept entry identifier: 3 (for concept *circuit-breaker*)

Concept position: 1.2

Definition: mechanical switching device, capable of making, carrying and breaking currents under normal circuit conditions and also making, carrying for a specified duration and breaking currents under specified abnormal circuit conditions such as those of short circuit[2)]

4.4 Concept entries

4.4.1 Concept orientation

The structure of concept entries in terminology databases is distinguished from that of conventional alphabetically sorted dictionaries or lexicons in that each concept entry contains information about a single concept, together with all the terms that are used to designate this concept, in as many languages as desired, whereas entries in dictionaries or lexicons comprise all meanings associated with one word.

4.4.2 Language

In multilingual environments, different language sections are used in order to instantiate information that needs to be available in these languages, such as definitions, notes, and terms. In the following example, the definitions and their sources are allocated at the language level.

EXAMPLE

Concept entry identifier: 10 (for concept *vacuum circuit-breaker*)

Concept position: 1.2.3

Subject field: switchgears, controlgears and fuses

Language: German

2) Definition taken from Electropedia, http://www.electropedia.org.

Definition: Leistungsschalter, dessen Kontakte sich innerhalb eines hochevakuierten Gefäßes öffnen und schließen[3)]

Source (of definition): IEV

Term: Vakuumleistungsschalter

Part of speech: noun

Grammatical gender: masculine

Term: Vakuumschalter

Part of speech: noun

Grammatical gender: masculine

Language: French

Definition: disjoncteur dont les contacts s'ouvrent et se ferment dans une enceinte où règne un vide poussé[4)]

Source (of definition): Electropedia

Term: disjoncteur à vide

Part of speech: noun

Grammatical gender: masculine

4.4.3 Dependency and repeatability of data categories

In general, data categories are directly associated to the nodes of the terminological metamodel. However, data categories can be further nested. For example, /definition/, /context/, and /note/ can be supplemented by /source/ to indicate the source of the information, or /term/ can be specified by aggregating a variety of term-specific data categories to describe the properties of the term (see 4.4.2, Example).

Additionally, in certain application scenarios, data fields need to be repeated as often as necessary and be combined with other data fields to adequately record information about concepts or terms in concept entries. As shown in the example in 4.4.2, the data field "term" and the corresponding term-specific data fields shall be repeatable in a language section as many times as needed in order to document the synonyms for a given language. The possibility to provide an equal level of information and description for each term included in a concept entry is called term autonomy.

Depending on the purpose of a terminology database, it can be decided that some data categories or data fields shall not be repeatable in a concept entry or in part of it. For instance, in a prescriptive terminology database, the data field "definition" could be constrained to a single instance in the concept entry, or as a single instance in a language section, whereas in a highly descriptive terminology database, the data field "definition" might be repeatable. Similarly, it is common to have only one subject field that applies for a given concept entry, or only one part of speech value for a given term. However, if various subject fields need to be added to a concept entry, i.e. a primary subject field and a secondary subject field, the first can be designed as a non-repeatable and the second as a repeatable data field.

4.4.4 Data granularity

Concept entries shall provide data categories designed to accommodate the appropriate degree of data granularity. For instance, grammatical information shall be recorded in various specific data fields as illustrated in Example 1.

3) Definition taken from DKE-IEV Wörterbuch, https://www.dke.de/de/services/iev-woerterbuch.

4) Definition taken from Electropedia, http://www.electropedia.org.

EXAMPLE 1

Term: Leistungsschalter

Part of speech: noun

Grammatical gender: masculine

Grammatical number: singular

Combining multiple properties in one general data category, as shown in Example 2, is not recommended.

EXAMPLE 2

Term: Leistungsschalter

Grammar: noun

Grammar: masculine

Grammar: singular

Terminology database designs that feature multiple specific data categories are more "granular" than those that use fewer, more general, data categories. Failing to differentiate subtypes of information in this way can lead to situations where data become difficult to retrieve or exchange.

4.4.5 Data elementarity

Data elementarity refers to the principle whereby a data field shall only hold a single unit of information. For example, every term shall be recorded in a separate data field "term" and not together with its synonyms. Such bad practices are common in published glossaries, but shall be avoided in a terminology database. Automatic processing of much of the data becomes difficult, if not impossible, if this principle is not observed. For instance, it would be impossible to search for or to extract acronyms from the database effectively if they were combined in the same field as their full forms. Similarly, the presence of two or more terms in the same field causes problems for tools that are designed to automatically propose or insert terms from concept entries into documents as they are being created or translated.

Decisions affecting data granularity and data elementarity involve weighing the information retrieval advantages afforded by providing more granular information against the sometimes increased effort required to enter information into separate fields. Care is to be taken during the design stage to determine what kind of information users want to record. Failure to provide the appropriate data fields for each kind of information can result in people using inappropriate fields or entering different kinds of information in a single field. Conversely, requiring users to fill in more fields than they need can decrease productivity during data input, or even induce users to provide placeholders instead of the required information and hence affect usability of the terminology database.

4.4.6 Data-modeling variation

Applying the design principles mentioned in this document can yield various specific data models. Example 1 and Example 2 below illustrate two different ways of expressing the same relation between a concept and one of its subordinate concepts.

EXAMPLE 1

Concept entry identifier: 3 (for concept *circuit-breaker*)

Subordinate concept generic: vacuum circuit-breaker

EXAMPLE 2

Concept entry identifier: 3 (for concept *circuit-breaker*)

Related concept: vacuum circuit-breaker

Relation type: subordinate generic

4.5 Roles

In distributed workflow environments, terminology database designers shall consider including a granular set of dedicated roles that need to be assigned to the users ranging from giving comprehensive rights for the whole terminology database to granting limited rights for only manipulating specific terminology database objects (such as terminology database definitions, filters or export/import profiles) or instances (such as concept entries, language sections or certain terms). However, the roles and rights management is highly use-case specific and needs to be adapted to the circumstances in place.