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

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.