
**Management of terminology
resources —
Terminology databases —
Part 3:
Content**

iTeh STANDARD PREVIEW
(standards.iteh.ai)

*Gestion des ressources terminologiques — Bases de données
terminologiques —
Partie 3: Contenu*

ISO 26162-3:2023

<https://standards.iteh.ai/catalog/standards/sist/774f6e85-75dd-4f84-9e7a-8c94fc88586f/iso-26162-3-2023>



iTeh STANDARD PREVIEW
(standards.iteh.ai)

ISO 26162-3:2023

<https://standards.iteh.ai/catalog/standards/sist/774f6e85-75dd-4f84-9e7a-8c94fc88586f/iso-26162-3-2023>



COPYRIGHT PROTECTED DOCUMENT

© ISO 2023

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
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
4 Identifying terms	3
4.1 Requirements for term selection	3
4.2 Unithood and termhood	4
4.3 Corpora and term extraction	5
5 Collecting terminological data	6
5.1 Data requirements	6
5.1.1 Evaluation procedure	6
5.1.2 Quality data	6
5.1.3 Purpose of the termbase	6
5.1.4 Data correctness	6
5.1.5 Fitness for use	7
5.2 Data model	8
5.3 Data categories	8
5.4 Data portability	8
6 Validating concept entry quality	8
6.1 General validation criteria	8
6.2 Error typology and system design	8
6.3 Error types	9
6.3.1 Termbase specification and maintenance	9
6.3.2 Error types associated with the inclusion of concept entries	9
6.3.3 Error types associated with automatically generated content	9
6.3.4 Error types associated with open data categories	10
6.3.5 Error types associated with closed data categories	10
Annex A (informative) Evaluation models	11
Annex B (informative) Sample concept entry	13
Annex C (informative) Sample error typology	17
Bibliography	21

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*.

A list of all parts in 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

Managers, educators and terminology database maintenance authorities conduct both periodic and continuous evaluation of terminology databases containing concept entries for a number of purposes:

- quality assurance-related validation of terminological data collections in business, government and non-governmental organizations;
- formative assessment and summative evaluation and feedback in training and educational environments.

ISO 26162-1 and ISO 26162-2 specify design principles and software considerations for modelling terminology databases (termbases). ISO 26162-1 establishes the general principles of termbase design as outlined in core ISO/TC 37 standards, such as ISO 704, which, among other topics, treats general principles for concept entry content and structure, term identification, basic principles for modelling concept systems and a range of other areas associated with terminology work. ISO 26162-1 also encourages conformity to the terminological metamodel as outlined in ISO 16642. It describes the role that data categories play in modelling terminological data and sets down basic principles for ensuring and evaluating the quality of data stored in termbases, such as data granularity, elementarity and modelling variance. These criteria comprise fundamental benchmarks against which to measure the quality and reliability of terminological data. ISO 26162-2 relates the principles outlined in ISO 26162-1 to the implementation of database design with respect to software and user interface considerations, together with pragmatic workflow implementations in terminology management environments.

This document provides guidance for defining procedures for ensuring high-quality content in terminological data collections designed to meet documentation needs in a range of environments involving, for instance, translation, interpreting and technical communication. Conformity to this document can strengthen processes designed to support a quality management system, such as ISO 9001, and the related auditing procedures in a translation, interpreting or technical communication environment. An error typology is presented in the framework of an overall evaluation model, with generic (non-standardized) options for creating a concept entry evaluation model, depending on the needs of users and of the sponsoring organization.

[Annexes A](#) to [C](#) provide pragmatic advice on error evaluation practice. [Annex A](#) describes the creation of scoring models reflecting the error typology described in [Clause 6](#), allowing for design variations depending on organization needs. For instance, a given scoring model can form the basis for a score card used for students and trainees, which is likely to be different from a score card used for a major enterprise or a national term bank.

[Annex B](#) presents a sample term entry. [Annex C](#) presents a sample evaluation model that can be adopted or adapted to meet the needs of terminologists, individuals working as freelancers or in companies, governmental organizations and NGOs. The values in this evaluation model can be used to create a scoring method, with the understanding that actual scoring practice is likely to vary according to specifications and user needs.

Management of terminology resources — Terminology databases —

Part 3: Content

1 Scope

This document specifies content-related aspects of terminology database maintenance. It gives guidance on the content of terminological data collections, with emphasis on data quality evaluation.

This document gives guidance for modellers of concept entries who need to ensure interoperability and high-quality content. It aims to ensure that terminological data collections themselves meet high standards for design conformity with standards such as ISO 12620-1 and ISO 16642, data accuracy and performance. It outlines principles for assuring data quality (see ISO 9001) and evaluating terminological data collections for purposes of continuous improvement. This approach contrasts that of ISO 23185:2009, which focuses on the usability of existing terminology resources.

This document does not apply to the management of text corpora or to term extraction tools.

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 1087, *Terminology work and terminology science — Vocabulary*

ISO 26162-1, *Management of terminology resources — Terminology databases — Part 1: Design*

ISO 26162-2:2019, *Management of terminology resources — Terminology databases — Part 2: Software*

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 terminology databases for use in standardization at the following addresses:

- ISO Online browsing platform: available at <https://www.iso.org/obp>
- IEC Electropedia: available at <https://www.electropedia.org/>

3.1

concept entry entry

terminological entry

part of a *terminological data collection* (3.14) which contains the terminological data related to one concept

Note 1 to entry: A concept entry can contain information treating one or more languages.

Note 2 to entry: In this document, the term *entry* is used as a short form for *concept entry*.

[SOURCE: ISO 30042:2019, 3.5, modified — “entry” made second preferred term, notes to entry added.]

3.2 evaluation model

model for analysing data in a *concept entry* (3.1) according to terminology principles and specified data requirements consistent with the purpose of the *termbase* (3.11)

3.3 core structure

common structure and *data categories* (3.6) that are used in all TermBase eXchange (TBX) *dialects* (3.5)

Note 1 to entry: The core structure is compliant with ISO 16642 (TMF).

[SOURCE: ISO 30042:2019, 3.6]

3.4 error typology

systematic list of error types

3.5 TBX dialect

eXtensible Markup Language (XML) that validates according to the *core structure* (3.3) of TermBase eXchange (TBX) and allows exactly those *data categories* (3.6) at those levels specified by a precisely defined configuration of data categories

Note 1 to entry: See ISO 30042 for more detail.

[SOURCE: ISO 30042:2019, 3.12, modified — Simplified for this document, Note 1 to entry replaced.]

3.6 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 30042:2019, 3.8]

3.7 data integrity

conformance of data values to a specified set of rules

[SOURCE: ISO/IEC TR 10032:2003, 2.23]

3.8 lexical unit

unit of language, belonging to the lexicon of a given language

[SOURCE: ISO 1951:2007, 3.8, modified — Reference to a dictionary removed.]

3.9 quality assurance

QA
set of planned and systematic activities necessary to provide confidence that a *concept entry* (3.1) satisfies acceptance criteria based on terminology principles and specified data requirements

3.10 specification

document that sets out detailed requirements to be satisfied by a *terminological data collection* (3.14)

Note 1 to entry: Specifications can include procedures for checking conformity to these requirements.

3.11**termbase****terminology database**

database comprising a *terminological data collection* (3.14)

[SOURCE: ISO 30042:2019, 3.28, modified — “terminology database” is no longer an admitted term, but a second preferred term.]

3.12**termbase quality evaluator**

person who is qualified as a terminologist or subject-matter specialist who conducts a quality evaluation of a *terminological data collection* (3.14)

3.13**termhood**

degree to which a *lexical unit* (3.8) is recognized as a term

EXAMPLE “bulk carrier ship” has stronger termhood than “ship” alone. “Mouse” has termhood in computer applications, whereas it does not in general language.

Note 1 to entry: Termhood applies to both simple terms (consisting of a single word) and complex terms (consisting of more than one word or lexical unit), and to other designations, such as proper names and appellations, as well as formulas and symbols.

3.14**terminological data collection**

resource consisting of *concept entries* (3.1) with associated metadata and documentary information

EXAMPLE A TBX document instance, ISO 1087.

[SOURCE: ISO 30042:2019, 3.29, modified — Second preferred term “TDC” removed.]

3.15**unithood**

degree to which a given sequence of words has sufficient collocational strength to form a stable *lexical unit* (3.8)

EXAMPLE “art deco table” has stronger unithood than “modern table”.

Note 1 to entry: Because unithood derives from the collocational relationship of words making up a given string, it only applies to multi-word terms.

4 Identifying terms**4.1 Requirements for term selection**

Concept entries should meet the needs of their intended audience and purpose as well as organization-specific requirements, including requirements for terms. ISO 704 discusses principles for assigning and analysing designations.

While it is not possible to set universal requirements for term selection and for the content of concept entries, an important goal of a termbase in many commercial environments is prescriptive in nature: directing users away from terms that are problematic for one reason or another, and towards the use of preferred terms. Thus, term selection involves both identifying the organization’s preferred terms and documenting the corresponding synonymous terms that are to be avoided. In this context, implementers should document deprecated and obsoleted terms along with preferred terms, as well as non-central terms that nonetheless recur in critical enterprise content.

Common considerations of including a term in the termbase include:

- Human safety: Important terms that, when used incorrectly, could affect the safety of customers or employees. When the wrong terms are used, the product can be used incorrectly, inadvertently damaging the product, or in some cases, posing a risk to the user. The same concerns apply for terms used in documents related to government regulations that affect citizen access to critical healthcare, interaction with the court system, occupational safety and health, and other resources and human services.

EXAMPLE 1 Terms related to pharmaceutical prescriptions, safety equipment on an oil rig, airplane landing gear assembly and maintenance, or communications between schools and parents.

- Company survival: Terms that protect the company or organization. Using the wrong terms could result in liability suits or loss of intellectual property.

EXAMPLE 2 Terms related to regulatory requirements in an industry or patents that a company owns.

- Company identity: Terms in this category represent concepts created by a company. Using them correctly protects the brand and trademarks and strengthens market share. More and more global companies today check possible translations for suitability when deciding on brand-related terms and trade names. It is important to make sure that terms will establish the company as an industry leader and set it apart from the competition. These items would be terms that are related to products or services, or terms that are used with a non-standard meaning.

EXAMPLE 3 Product names and registered company trademarks; components in patented products; names of features or services below the branding level.

- Subject field: Terms used in an industry or domain. These include terms from vertical industries or areas where experts (outside of a given organization) agree on a set of terms.

EXAMPLE 4 Domains such as accounting, core automotive, biology.

- Public service entities: Terms that will be transparent to a given target audience.

EXAMPLE 5 Materials written in an under-resourced language sometimes need to use different terminology from the scholarly or industry-standard terms used by experts.

- Pragmatic and locale-related issues: Any terms that benefit from documentation or standardization, including high-frequency terms or difficult terms in cases where documenting and managing these terms will reduce errors or queries. If products are localized or documents are translated into other languages, this category also covers terms that can pose translation challenges, or which can benefit from standardization in one or more languages. Some languages can benefit from a higher level of terminology documentation than others.

4.2 Unithood and termhood

As shown in [Figure 1](#), “termhood” comprises the intersection among unithood, usage in the sponsoring organization’s corpus and the purpose of the terminological data collection (see Reference [\[16\]](#)).

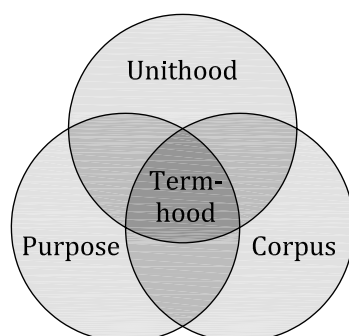


Figure 1 — Termhood

“Unithood” refers to collocational relationships between words and is not applicable to single-word lexical units. Determining whether a given multi-word string comprises a stable, lexical unit that recurs with collocational frequency in the organization’s corpus is important for term selection. Because of their relatively stable syntagmatic structures, multi-word units with strong unithood tend to form key communicative structures demanding consistency in commercial environments. While multi-word terms prevail in commercial and professional environments, this does not exclude the documentation of simple terms that have sufficient prominence (frequency and/or dispersion) in the organization’s corpus and support the purpose of the terminological data collection.

A multi-word lexical unit which meets the criteria for unithood functions as a term if it designates an identifiable concept in the textual and operational context in question. Each concept shall be documented as a separate concept entry in a termbase, adhering to the principle of concept orientation.

EXAMPLE Sometimes one word in one language requires multiple concept entries. For instance, the English word (which in some contexts can function as a term) “river” denotes two concepts as illustrated by the French “fleuve” (a river which flows into a sea or ocean) and “rivière” (a river that flows into another river or lake). The German term “Abstandsbolzen” has two conceptual references as evidenced in English: “distance rivet” and “spacer bolt”. Both of these cases would require separate concept entries in a termbase.

4.3 Corpora and term extraction

In commercial environments, identifying terms is usually a process informed by the textual context in which they occur. It is recommended to use corpora to identify term candidates. Corpus analysis informs the depth or breadth of subject-field coverage without overloading the termbase with unnecessary entries. Corpora may consist of any kind of written materials produced by the sponsoring organization, including marketing materials, product documentation, internal memos, transcripts, bilingual translation memories or other collections of textual materials. The organization-wide corpus provides evidence of the frequency of occurrence for a given term candidate, as well as its dispersion across various types of textual materials. Both frequency and dispersion are criteria for determining whether a candidate meets the criteria for termhood.

EXAMPLE 1 Certain low-frequency lexical units can be important terms for an organization because they appear across multiple types of content (marketing, sales, training, online content, user guides, etc.), or they can nonetheless represent key concepts.

A range of tools and techniques are available to assist in the extraction of lexical units for the purpose of identifying terms (see ISO 26162-2 as well as ISO 12616-1 for additional information on term extraction tools).

NOTE ISO 5078^[3] on terminology extraction is also being developed.

Two key factors to consider for selecting relevant concepts and terms are:

- marketing: key concepts involving brand recognition that distinguish an organization from its competitors;

EXAMPLE 2 Concepts and related terms associated with patented and trademarked products or processes.

- customer satisfaction: areas where terms and concepts have historically caused problems and can reflect past quality assurance (QA) problems or other serious risk criteria.

EXAMPLE 3 Particular topics that have generated a higher than usual number of support calls or customer complaints.