International Standard

ISO 16642

Management of terminology resources — Terminological markup framework

Gestion des ressources terminologiques — Plate-forme pour le balisage de terminologies informatisées

Third edition
2025-12

Reference number
ISO 16642:2025(en)

© ISO 2025



COPYRIGHT PROTECTED DOCUMENT

© ISO 2025
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 Modular approach ... **4**
5 Generic model for describing terminological data ... **6**
5.1 Principles ... 6
5.2 Metamodel ... 7
5.3 Example ... 9
6 Requirements for compliance to TMF ... **9**
7 Interchange and interoperability ... **10**
8 Representing languages ... **10**
9 Defining a TML ... **11**
9.1 Requirements ... 11
9.2 Defining interoperability conditions ... 11
10 Implementing an XML-serialized TML ... **11**
10.1 General ... 11
10.2 Implementing the metamodel ... 11
10.3 Anchoring data categories ... 12
10.3.1 General ... 12
10.3.2 Styles and vocabulary ... 12
10.4 Constraints on datatypes ... 13
10.5 Implementing annotations ... 13
10.6 Implementing brackets ... 13
Annex A (informative) **Conformance of terminological data to TMF: example scenario** ... **14**
Bibliography ... **18**

© ISO 2025 – All rights reserved

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 document 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).

ISO draws attention to the possibility that the implementation of this document may involve the use of (a) patent(s). ISO takes no position concerning the evidence, validity or applicability of any claimed patent rights in respect thereof. As of the date of publication of this document, ISO had not received notice of (a) patent(s) which may be required to implement this document. However, implementers are cautioned that this may not represent the latest information, which may be obtained from the patent database available at www.iso.org/patents. ISO shall not be held responsible for identifying any or all such patent rights.

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 third edition cancels and replaces the second edition (ISO 16642:2017), which has been technically revised.

The main changes are as follows:

— The scope is no longer restricted to representing terminological data in XML format. Terminological markup languages can be serialized in any formats, but they are exemplified in this document in XML.

— DatCatInfo PIDs have been updated.

— Terms and definitions have been updated according to the latest International Standards.

— Examples have been updated to reflect ISO 30042:2019.

— Annex A has been significantly revised in order to show two TMF-compliant XML-based serialization examples of the same terminological data.

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.

© ISO 2025 – All rights reserved

Introduction

Terminological data are collected, managed and stored in a wide variety of systems, typically database management systems, ranging from personal computer applications for individual users to large terminology database systems operated by major companies and governmental agencies. Terminology databases comprise various types of information, called “data categories”, and can adopt different structural models. However, terminological data often need to be shared and reused in a number of applications, and this sharing is facilitated when the data adhere to a common model. To facilitate co-operation and to prevent duplicate work, it is important to develop standards and guidelines for creating and using terminological data collections (TDCs) as well as for sharing and exchanging data.

This document presents a modular approach for analysing existing TDCs and designing new ones. It also provides a framework for defining terminological markup languages (TMLs) that are interoperable.

This document refers to DatCatInfo, an example of an available data category repository (DCR). DatCatInfo is an online database of information about the types of data that can be included in TDCs and other language resources. It is available at www.datcatinfo.net.

© ISO 2025 – All rights reserved

Management of terminology resources — Terminological markup framework

1 Scope

This document specifies a framework for representing data recorded in terminological data collections (TDCs). This framework includes a metamodel and methods for describing specific terminological markup languages (TMLs), exemplified in this document in eXtensible Markup Language (XML). The mechanisms for implementing constraints in a TML are defined, but not the specific constraints for individual TMLs.

This document is designed to support the development and use of computer applications for terminological data and the exchange of such data between different applications. This document also defines the conditions that allow the data expressed in one TML to be mapped onto another TML.

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 and terminology science — Vocabulary*

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

ISO 26162 (all parts), *Management of terminology resources — Terminology databases*

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

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
basic information unit
information unit (3.13) attached to a *component* (3.3) of the *metamodel* (3.15) that can be expressed by means of a single *data category* (3.7)

EXAMPLE Term, note.

© ISO 2025 – All rights reserved

3.2
complementary information
CI
information supplementary to that described in *concept entries* (3.5) and shared across the *terminological data collection* (3.23)

EXAMPLE Domain hierarchies, bibliographic references, references to text corpora.

3.3
component
elementary description unit of a *metamodel* (3.15) to which *data categories* (3.7) can be associated to form a data model

3.4
compound information unit
information unit (3.13) attached to a *component* (3.3) of the *metamodel* (3.15) that can be expressed by means of several grouped *data categories* (3.7)

EXAMPLE IDLangTgtDtyp, transacGrp.

3.5
concept entry
CE
terminological entry
entry
part of a *terminological data collection* (3.23) which contains the terminological data related to one concept

3.6
conceptual domain
permissible content of a *data category* (3.7)

EXAMPLE In a terminology database, the data category /part of speech/ can have a conceptual domain consisting of the values /noun/, /verb/, /adjective/, /adverb/.

Note 1 to entry: The permissible content can be closed, as in the example, or subject to formal restrictions such as dates, or free text such as the conceptual domain of /definition/. Although the latter type is not formally restricted, it is nevertheless subject to adherence to the requirements of its *data category specification* (3.10), i.e. it contains a true definition and not a note, example or some other piece of information.

[SOURCE: ISO 12620-1:2022, 3.1]

3.7
data category
DC
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/).

3.8
data category repository
DCR
digital collection of *data category specifications* (3.10)

EXAMPLE DatCatInfo, a DCR for language resources (see Reference [6]).

Note 1 to entry: Data category repositories are used as references when specifying language resources.

[SOURCE: ISO 12620-1:2022, 3.6]

© ISO 2025 – All rights reserved

3.9
data category selection
DCS
DC selection
set of *data category specifications* (3.10) selected from a *data category repository* (3.8)

Note 1 to entry: A data category selection can represent the *data categories* (3.7) used within a research discipline or a specific application or project.

[SOURCE: ISO 12620-1:2022, 3.7, modified — "chosen" replaced by "selected"; "DCS" made the second preferred term.]

3.10
data category specification
DC specification
complete descriptive record of a *data category* (3.7)

[SOURCE: ISO 12620-1:2022, 3.5]

3.11
expansion tree
structured group of serialized elements that implement a level of the *metamodel* (3.15) in a given *terminological markup language* (3.24)

3.12
global information
GI
technical and administrative information applying to an entire *terminological data collection* (3.23)

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

3.13
information unit
IU
elementary piece of information attached to a *component* (3.3) of the *metamodel* (3.15)

3.14
language section
LS
part of a *concept entry* (3.5) containing information related to one language

3.15
metamodel
model that specifies one or more other models

[SOURCE: ISO/IEC TR 19583-24:2025, 3.2.19]

3.16
object language
language being described

3.17
PID
persistent identifier
unique identifier that ensures permanent access for a digital object by providing access to it independently of its physical location or current ownership

[SOURCE: ISO 24619:2011, 3.2.4, modified — Note 1 to entry deleted.]

© ISO 2025 – All rights reserved

3.18
serialization
process of translating data structures or object states into a format that can be stored or transmitted and reconstructed later

[SOURCE: ISO/TS 23494-1:2023, 3.16]

3.19
serialization format
data storage format for storing, transmitting, and reconstructing data structures or object states

[SOURCE: ISO/IEC TR 19583-24:2025, 3.2.41, modified — "object state" replaced by "object states".]

3.20
style
specification for the implementation of a *data category* (3.7) in any *serialization format* (3.19)

3.21
term component section
TCS
part of a *term section* (3.22) containing linguistic information about the parts of a term

[SOURCE: ISO 26162-1:2019, 3.2.10, modified — "components" replaced by "parts".]

3.22
term section
TS
part of a *language section* (3.14) containing information about a term

[SOURCE: ISO 26162-1:2019, 3.2.9]

3.23
terminological data collection
TDC
resource consisting of *concept entries* (3.5) with associated metadata and documentary information

EXAMPLE A TBX document instance, ISO 1087:2019.

3.24
terminological markup language
TML
serialization format (3.19) for representing a *terminological data collection* (3.23) conforming to the constraints of the *metamodel* (3.15)

3.25
vocabulary
<data modelling> set of strings used to implement a *data category* (3.7) according to a *style* (3.20)

3.26
working language
language used to describe objects

4 Modular approach

Terminological markup framework (TMF) consists of two levels of abstraction:

— The first (and most abstract) level is the metamodel level. The metamodel level supports analysis, design and exchange at a very general level, i.e. it is independent of any specific implementation or software. The metamodel shall be shared by all TDCs that are compliant with TMF.

— The second level is the data model level, which adds the necessary data categories for representing specific TDCs.

© ISO 2025 – All rights reserved