
**Management of terminology
resources — TermBase eXchange (TBX)**

Gestion des ressources terminologiques — TermBase eXchange (TBX)

iTeh Standards
(<https://standards.itih.ai>)
Document Preview

ISO 30042:2019

<https://standards.itih.ai/catalog/standards/iso/a45e92fb-de7f-435f-bad9-ef4e753d31e4/iso-30042-2019>



iTeh Standards
(<https://standards.iteh.ai>)
Document Preview

ISO 30042:2019

<https://standards.iteh.ai/catalog/standards/iso/a45e92fb-de7f-435f-bad9-cf4e753d31e4/iso-30042-2019>



COPYRIGHT PROTECTED DOCUMENT

© ISO 2019

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
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

Page

Foreword	v
Introduction	vi
1 Scope	1
2 Normative references	1
3 Terms and definitions	1
4 Fundamental principles and concepts	5
5 Data categories	6
5.1 General	6
5.2 Data categories specified in the core structure module	6
5.3 Data categories specified in the data category modules	6
6 DCA and DCT styles	6
7 Dialects	7
7.1 General	7
7.2 Dialect naming	7
7.3 Example of a dialect	7
7.4 Requirements for a dialect to be TBX-compliant	8
7.5 Validating a TBX document instance	9
7.6 Requirements for compliant TBX agents	10
8 The core structure	11
8.1 General	11
8.2 Metamodel	11
8.3 Position of elements within a concept entry	12
8.3.1 Elements that may appear at multiple levels	12
8.3.2 Elements that occur only at the term level	14
8.4 Typology of elements	14
8.4.1 Elements that play a classification or grouping role	14
8.4.2 Elements that represent data categories	15
8.4.3 Inline markup elements	15
8.5 Attributes	18
8.5.1 type	18
8.5.2 xml:lang	18
8.5.3 id and target	18
8.5.4 module	18
8.6 Types of text	18
8.7 Character sets and encoding	19
9 Defining data category modules	19
9.1 General	19
9.2 Naming the module	19
9.3 Selecting data categories	19
9.4 Defining data category properties	20
9.5 Defining data category constraints	20
9.6 Using modules	20
10 Referencing objects	20
10.1 General	20
10.2 Referencing a file that is embedded in the back matter	21
10.3 Referencing a file from the back matter	21
10.4 Referencing a file directly in the entry	21
10.5 Referencing an external source	22
10.6 Referencing and documenting a bibliographic source	22
10.7 Referencing and documenting information about a person or organization	23

10.8 Referencing original data from noteText entities	23
Annex A (normative) Descriptions of the core-structure elements and attributes.....	25
Annex B (informative) Data category names	37
Annex C (informative) RelaxNG schemas for the core structure and TBX Module Description (TBXMD).....	39
Bibliography	43

iTeh Standards
(<https://standards.iteh.ai>)
Document Preview

ISO 30042:2019

<https://standards.iteh.ai/catalog/standards/iso/a45e92fb-de7f-435f-bad9-ef4e753d31e4/iso-30042-2019>

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 on 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 the following URL: 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 second edition cancels and replaces the first edition (ISO 30042:2008), which has been technically revised.

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

- industry-defined dialects consisting of data category selections corresponding to the needs of specific communities have been introduced;
- the XCS formalism has been removed and replaced with the requirement that the dialect be described and its name be declared on the root element of every TBX document instance;
- the DTD for the core structure has been replaced with a schema language-neutral definition;
- this document, containing the essential core and normative content, has been separated from ancillary content produced and distributed publicly by stakeholders;
- a simplified DCT (Data Category as Tag) style has been added alongside the traditional TBX style of DCA (Data Category as Attribute);
- xml namespaces have been introduced as a means for declaring the data categories used in a given TBX dialect (for DCT style).

NOTE Additional details about these and other changes are available on the TBX Info website^[15].

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

This document defines a framework for representing structured terminological data, referred to as TermBase eXchange (TBX). Within this framework, a variety of industry standards, known as dialects, for specific types of terminology interchange scenarios and terminological data collections can be defined.

TBX is designed to support various types of processes involving terminological data, including analysis, descriptive representation, dissemination, and exchange in various computer environments. The primary purpose of TBX is the exchange of terminological data. For example, it facilitates:

- integrating or converting terminological data from multiple sources;
- comparing the contents of various terminological data collections;
- querying multiple terminological data collections by passing data through a common intermediate format on a batch or dynamic basis;
- placing data on an online site for download or public feedback;
- making terminology available dynamically in networked applications through a web service.

A TBX-compliant dialect can facilitate the exchange of terminological data between users, which include people such as translators and writers, as well as applications and systems, such as computer assisted translation tools and controlled authoring software. Therefore, it can be used for both human-oriented and machine-oriented terminological data processing. In this manner, it can enable the flow of terminological information between technologies and systems throughout the information production cycle, both inside an organization and with outside service providers.

TBX document instances of the same defined TBX dialect are interoperable and exchangeable with minimal loss or minimal need for negotiation, because they:

- adhere to the core structure;
- use, or have access to, the same data categories; and
- comply with the same dialect-specific constraints as other instances of the same dialect.

TBX document instances developed according to ISO 30042:2008 can be converted to comply with the current version of TBX by identifying a dialect with which the document instance complies and implementing the other changes in accordance with this document. A converter is available on the TBX Info website for such purposes^[15].

NOTE Supplemental resources are available to assist implementers and users of TBX dialects on the TBX Info website^[15].

TBX is limited in its ability to represent presentational markup (such as bold or italics). However, presentational markup can be autogenerated from descriptive markup in a TBX document instance.

Management of terminology resources — TermBase eXchange (TBX)

1 Scope

This document explains fundamental concepts and describes the metamodel, data categories, and XML styles: DCA (Data Category as Attribute) and DCT (Data Category as Tag). It also specifies the methodology for defining TBX dialects.

The audience for this document is anyone wishing to create a new dialect compliant with TBX. This document can also be used to analyze and to understand a terminological data collection or to design a new terminology database that complies with international standards and best practices. Typical users are programmers, software developers, terminologists, analysts, and other language professionals. Intended application areas include translation and authoring.

The TBX-Core dialect is described in detail in this document. All other industry-supported dialects are out of the scope of this document.

NOTE TBX dialects are defined by industry stakeholders. Any materials needed to implement currently shared dialects are publicly available as self-contained industry specifications (see for instance the TBX Info website^[15]).

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 8601-1, *Date and time — Representations for information interchange — Part 1: Basic rules*

ISO 8601-2, *Date and time — Representations for information interchange — Part 2: Extensions*

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

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

ISO 21720, *XLIFF (XML Localisation interchange file format)*

ISO/IEC 10646, *Information technology — Universal Coded Character Set (UCS)*

3 Terms and definitions

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

ISO and IEC maintain terminology databases for use in standardization 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

attribute class

group of one or more related attributes