
**Systems to manage terminology,
knowledge and content — TermBase
eXchange (TBX)**

*Systèmes de gestion de la terminologie, de la connaissance et du
contenu — TermBase eXchange (TBX)*



PDF disclaimer

This PDF file may contain embedded typefaces. In accordance with Adobe's licensing policy, this file may be printed or viewed but shall not be edited unless the typefaces which are embedded are licensed to and installed on the computer performing the editing. In downloading this file, parties accept therein the responsibility of not infringing Adobe's licensing policy. The ISO Central Secretariat accepts no liability in this area.

Adobe is a trademark of Adobe Systems Incorporated.

Details of the software products used to create this PDF file can be found in the General Info relative to the file; the PDF-creation parameters were optimized for printing. Every care has been taken to ensure that the file is suitable for use by ISO member bodies. In the unlikely event that a problem relating to it is found, please inform the Central Secretariat at the address given below.

This document is a preview generated by EVS



COPYRIGHT PROTECTED DOCUMENT

© ISO 2008

All rights reserved. Unless otherwise specified, no part of this publication may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying and microfilm, without permission in writing from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
Case postale 56 • CH-1211 Geneva 20
Tel. + 41 22 749 01 11
Fax + 41 22 749 09 47
E-mail copyright@iso.org
Web www.iso.org

Published in Switzerland

Contents

1 Scope.....	1
2 Normative references.....	1
3 Terms and definitions.....	2
4 Relationship to other standards.....	5
5 Applications of TBX.....	5
6 Fundamental principles.....	5
6.1 General.....	5
6.2 Principles relating to grouping and representing data-categories.....	6
7 Requirements for TBX files.....	7
7.1 Compliance requirements.....	7
7.2 Examples of non-compliance.....	8
7.3 Implementation levels.....	8
8 The core-structure module.....	8
8.1 Introduction.....	8
8.2 Hierarchy.....	9
8.3 Components of a terminological entry.....	10
8.4 Elements that can appear at multiple levels of the entry.....	10
8.5 Elements that occur only at the term level or lower.....	11
8.6 Handling of text.....	12
8.7 Meta data elements.....	14
8.8 Attributes.....	15
8.9 Character set issues.....	16
8.10 Language.....	16
9 The default data-category constraints.....	16
9.1 Introduction.....	16
9.2 Data-categories built into the core structure DTD of TBX.....	17
9.3 Data-categories specialized from meta data-categories through the default XCS file.....	17
10 Examples.....	21
10.1 Example of a typical TBX file.....	21
10.2 Examples of encoding TBX elements.....	22
10.3 Examples of TBX entries.....	23
11 Referencing objects.....	25
11.1 General information about referencing.....	25
11.2 Referencing a file that is embedded in the back matter of a TBX file.....	26
11.3 Referencing a file from the back matter.....	26
11.4 Referencing a file directly in the entry.....	27
11.5 Referencing an external source.....	27
11.6 Referencing and documenting a bibliographic source.....	27
11.7 Referencing and documenting information about a responsible person or organization.....	28
11.8 Referencing an external concept system, classification system, or thesaurus.....	29
11.9 Referencing a TBX entry from within a corpus.....	29
12 Creating customized TBX TMLs.....	29
12.1 General information about TMLs.....	29
12.2 Example of an XCS file for a user-defined TBX TML.....	30
12.3 Creating customized picklist display names.....	31
Annex A (Normative) DTD for the core structure module.....	33
Annex B (Normative) DTD for the data-category constraints (XCS file).....	38
Annex C (Normative) Default XCS file.....	40
C.1 Introduction.....	40
C.2 XCS file for the default data-categories and constraints.....	40
Annex D (Normative) Descriptions of the core structure elements and attributes and the default data-categories.....	48
Annex D.1 General information about the descriptions.....	48
D.2 Macros.....	48

D.3 Attribute classes.....	49
D.4 Elements.....	50
D.5 Default data-categories.....	61
Annex E (Normative) Descriptions of elements and attributes for the XCS file.....	73
E.1 Introduction.....	73
E.2 Attribute classes.....	73
E.3 Elements.....	73
Annex F (Informative) Integrated schema and other TBX resources.....	81
Annex G (Informative) TBX-Basic.....	82
Annex H (Informative) Summary of changes.....	83
Annex I (Informative) Indexes.....	88
I.1 Core-module DTD.....	88
I.2 XCS DTD.....	89
I.3 Terminological data-categories.....	89
Bibliography.....	91

This document is a preview generated by EVS

Foreword

The International Organization for Standardization (ISO) 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.

International Standards are drafted in accordance with the rules given in the ISO/IEC Directives, Part 2.

The main task of ISO technical committees is to prepare International Standards. Draft International Standards adopted by the technical committees are circulated to the member bodies for voting. Publication as an International Standard requires approval by at least 75 % of the member bodies casting a vote.

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.

ISO 30042 was prepared by LISA OSCAR and was adopted, under a special "fast-track procedure", by Technical Committee ISO/TC 37, *Terminology and other language and content resources*, Subcommittee SC 3, *Systems to manage terminology, knowledge and content*, in parallel with its approval by the ISO member bodies.

The Localization Industry Standards Association (LISA - www.lisa.org) is the standards organization for the globalization industry. Within LISA, the OSCAR (Open Standards for Container/content Allowing Reuse) Special Interest Group develops XML-based standards for automated language-processing in the areas of globalization, internationalization, localization, and translation, including standards for translation memory, terminology, text memory, word/character counts, and other related areas. The main task of the OSCAR Special Interest Group is to develop standards to facilitate and automate the globalization of products and services in a way that supports local language and culture conventions. Publication as an OSCAR standard requires approval by the OSCAR steering committee. An earlier version of TBX was developed and published by LISA in 2002.

TBX and the TBX logo are registered trademarks of LISA, and the TBX logo is subject to terms of use as defined by LISA. LISA maintains copyright on the TBX specification that is available on the LISA Web site, and ISO maintains copyright on the TBX specification that it distributes as ISO 30042. The technical content of these two documents is identical, and is subject to joint maintenance by a team of ISO TC 37 and LISA OSCAR members.

Introduction

This International Standard defines an XML-based framework for representing structured terminological data referred to as TermBase eXchange (TBX). Within this framework, a variety of terminological markup languages (TMLs) can be defined. A TML defined by TBX can facilitate the interchange of terminological data between users, which include people such as translators and writers, and 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. In this manner, it can enable the flow of terminological information throughout the information production cycle, both inside an organization and with outside service providers.

The intended audience for this document consists of two groups: (1) programmers and analysts who wish to develop software applications that process TBX-compliant data files; (2) terminologists and other language specialists who wish to analyse a terminological data collection for representation in TBX or to understand a TBX file.

This version of TBX is an update of a version that was published by the Localization Industry Standards Association (LISA) in 2002. Among other enhancements, the current version provides reference to an integrated schema that includes the core-structure module and the data-category constraints in combined declarations using the Relax NG and Schematron languages. It also provides reference to a TBX-compliant TML called TBX-Basic.

Users of this International Standard should first study the body (clauses 1-12). The suggested use of annexes A-I is described below.

(1) The core-structure module of TBX

All TMLs within the TBX framework have the same core structure. The core-structure module is described in Clause 8. A DTD for the core-structure module is found in Annex A. The elements, attributes, and data types are described in Annex D, and listed alphabetically in Annex I.

(2) The XCS module

TMLs may differ with respect to which data-categories are allowed, and at what levels of a terminological entry these data-categories can occur. These constraints on the core structure, which define a particular TML, are formally represented in an XCS file. A DTD for the XCS module is found in Annex B. The elements and attributes are described in Annex E, and listed alphabetically in Annex I.

(3) The default XCS of TBX

The TBX-default TML is constrained by the default XCS file. The TBX default XCS is described in Clause 9. The default XCS file is provided in Annex C. The data-categories are described in Annex D, and listed alphabetically in Annex I.

(4) Compliance checking of TBX document instances

Once a TBX TML has been defined by an XCS, a TBX document instance can be checked for compliance with that TML. The requirements for compliance are found in Clause 7. One can use a variety of methods and schema definition languages to check compliance. In particular, the Relax NG schema referred to in Annex F can be used to check whether a TBX document instance is compliant with the TBX-default TML. Annex F also indicates where a TBX user can find additional resources for compliance checking. Another TBX TML, called TBX-Basic, is referred to in Annex G.

(5) Changes that have been made to TBX since its submission to ISO in February 2007 are summarized in Annex H.

Summary of annexes:

- A: DTD for core-structure module
- B: DTD for XCS module
- C: Default XCS that defines the TBX-default TML
- D: Descriptions of core structure elements and attributes
- D.5: Descriptions of default data-categories
- E: Descriptions of XCS elements and attributes
- F: Relax NG schema and other resources for compliance checking
- G: Reference to TBX-Basic
- H: Summary of changes to TBX
- I: Indexes (alphabetical lists of elements and data-categories)

Systems to manage terminology, knowledge, and content - TermBase eXchange (TBX)

1 Scope

The TBX framework defined by this International Standard is designed to support various types of processes involving terminological data, including analysis, descriptive representation, dissemination, and interchange (exchange), in various computer environments. The primary purpose of TBX is for interchange of terminological data. It is limited in its ability to represent presentational markup. Intended application areas include translation and authoring.

TBX is modular in order to support the varying types of terminological data, or *data-categories*, that are included in different terminological databases (termbases). TBX includes two modules: a core structure, and a formalism for identifying a set of data-categories and their constraints, both expressed in XML. The term *TBX*, when used alone, refers to the framework consisting of these two interacting modules.

To maximize interoperability of the actual terminological data, TBX also provides a default set of data-categories that are commonly used in terminological databases. However, subsets or supersets of the default set of data-categories can be used within the TBX framework to support specific user requirements.

TBX, when used with its default set of data-categories, qualifies as a *terminological markup language (TML)* as defined in ISO 16642, which will be referred to as the *TBX-default TML* in this International Standard. Likewise, other markup languages that comply with TBX and use a subset of the default set of data-categories are also TMLs, but may go by other names, such as the one referred to in [Annex G \(Informative\) TBX-Basic](#).

2 Normative references

The following referenced documents are indispensable for the application 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 639-1:2002, *Codes for the representation of names of languages – Part 1: Alpha-2 code*

ISO 639-2:1998, *Codes for the representation of names of languages – Part 2: Alpha-3 code*

ISO 639-3:2007, *Codes for the representation of names of languages – Part 3: Alpha-3 code for comprehensive coverage of languages*

ISO/IEC 646:1991, *Information technology – ISO 7-bit coded character set for information interchange*

ISO 3166-1:2006, *Codes for the representation of names of countries and their subdivisions – Part 1: Country codes*

ISO 8601:2004, *Data elements and interchange formats – Information interchange – Representation of dates and times*

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

ISO 12200:1999, *Computer applications in terminology – Machine-readable terminology interchange format (MARTIF) – Negotiated interchange*

ISO 12620, *Computer applications in terminology – Data categories*

ISO 16642:2003, *Computer applications in terminology – Terminological markup framework*