



Terminology for Large Organizations

TBX-BASIC

Version 4, © 2025

DRAFT

Table of Contents

1. Document Information	4
2. Background	5
3. About TBX-Basic	5
4. Mandatory data categories	5
5. Data category descriptions	6
5.1 Context	7
5.2 Created by	7
5.3 Creation date	7
5.4 Cross-reference	8
5.5 Customer	8
5.6 Definition	8
5.7 External cross-reference	8
5.8 Figure	8
5.9 Gender	9
5.10 Geographical usage	9
5.11 Language	9
5.12 Last modified by	10
5.13 Last modified date	10
5.14 Note	10
5.15 Part of speech	11
5.16 Project	11
5.17 Source of context	11
5.18 Source of definition	12
5.19 Source of term	12
5.20 Subject field	12
5.21 Term	12
5.22 Term location	12
5.23 Term type	13
5.24 Usage status	13
6. Additional information about data categories	13
6.1 Term type picklist values	14
6.2 Definitions and contexts	15

6.3 Data categories used for categorizing purposes	16
Multiple subject fields can be assigned to a concept.	16
7. The Structure of a TBX-Basic Entry	16
7.1 Concept section	16
7.2 Language section.....	17
7.3 Term section	17
7.4 Backmatter	18
8. Inline markup	18
9. Compliance.....	19
10. Recommended standards and guidelines	19
Appendix A - Data categories by section of the entry	21
Concept section data categories.....	21
Language section data categories	21
Term section data categories.....	21
Appendix B – Properties of a valid TBX-Basic document instance.....	22
Appendix C – Files for Validation	24

1. Document Information

Owner

TerminOrgs

www.terminorgs.net

info@terminorgs.net

For the most recent version of the TBX-Basic specification, go to:

www.terminorgs.net/tbx-basic.html

Copyright

© Copyright Terminology for Large Organizations (TerminOrgs) 2025. All rights reserved. This publication may not be distributed for commercial purposes, reproduced, modified or translated without the explicit prior written permission of the copyright owner. For permission, contact info@terminorgs.net. The TerminOrgs logo is the property of Terminology for Large Organizations (TerminOrgs). These materials are provided by TerminOrgs for informational purposes only, without representation or warranty of any kind, and Terminology for Large Organizations (TerminOrgs) shall not be liable for errors or omissions with respect to the materials.

Typographical conventions

bold Used for the names of data categories

italics Used for sample terms

monospace font Used for XML elements and attributes

Glossary

data category (DC)

A type of data field, such as **definition** or **part of speech**

dialect

XML markup language that validates according to the core structure of TBX and allows exactly those data categories at those levels specified by a particular data category module or set of data category modules. TBX-Basic is a dialect.

module

list of permissible data categories and constraints on them that are used in the design of a TBX-compliant terminological data collection

terminology resource

A file, database, or other collection that contains terms and information about terms; frequently called a “glossary” in the localization industry.

2. Background

The first version of TBX-Basic was developed in 2009 by the Terminology Special Interest Group (Term SIG) of the Localization Industry Standards Association (LISA). In 2011, LISA ceased operations. The Term SIG continues to operate as an independent entity named TerminOrgs. TerminOrgs is the organization responsible for the TBX-Basic specification.

TerminOrgs published a second version of TBX-Basic in 2014. The current version is the third, published in 2024.

TBX-Basic is compliant with *ISO 30042 (2019): TermBase eXchange (TBX)*, the ISO standard for terminology data. That version of TBX is also referred to as : TBX Version 3. TBX refers to a framework consisting of two interacting components: a core structure and a formalism for defining data category modules.

3. About TBX-Basic

TBX-Basic is a TBX *dialect* that is designed for the translation industry and other industries that may want to develop a terminology database that meets basic needs. It is based on information from surveys and studies that were conducted by the LISA Term SIG about the types of terminology data that the localization industry requires.

TBX-Basic allows a limited set of data categories and adheres to a basic entry structure. Its purpose is to formalize the markup that is used in relatively simple terminology resources in an XML standard, in order to increase the resources' structural stability and interoperability.

TBX-Basic is intended to be the primary dialect of terminology exchange. TBX-Basic was designed to efficiently store a large quantity of terminology glossaries in a straightforward XML format. It can be used to handle monolingual, bilingual, or multilingual glossaries.

In ISO 30042 (2019), TBX (V3) is not a format or standard for any specific terminology database or model. The previous concept of a “TBX-Default” terminology markup language has been abandoned. Rather, TBX V3 is a framework for creating dialects based on some standard principles and requirements. A dialect is formed on the telescoping principle of inheritance of modules. The most basic module (and dialect) is TBX-Min. TBX-Basic “includes” TBX-Min and adds a few additional data categories.

The core structure of all TBX-compliant dialects is fixed by the TBX-Core schema (see Appendix C).

4. Mandatory data categories

There are only two mandatory data categories in TBX-Basic: **term** (<term>...</term>), and **language** (specified as the value of the xml:lang attribute on the <langSec> element).

The data categories **definition**, **context**, **part of speech**, and **subject field** should be included in a terminology entry whenever possible. A very important data category is **part of speech**. It is required for the following purposes:

- To differentiate homonyms. For instance, *port* is actually two terms in English: a noun, and a verb, each of which should be recorded in its own entry. Without a part of speech value in the entry, it can be difficult to determine which term the entry represents, and therefore, how to translate it.
- To permit automated processing. The part of speech is required for automated tasks such as importing a set of entries into an existing termbase, applying grammatical filters to facilitate search and export of data, and providing the terminology as a resource to other applications such as spell checking applications.
- To enable interchange. When there is no part of speech value, it becomes necessary to discuss many of the entries with the originator in order to disambiguate their content.

See section 9 *Compliance* for further guidelines about the requirements for the part of speech.

5. Data category descriptions

The data categories allowed in TBX-Basic are described in this section, arranged in alphabetical order. The descriptions contain the following parts:

Identifier

The identifier of the element taken from the data category repository, DatCatInfo, which is available at <https://datcatinfo.net/>.

XML markup

The XML representation of the data category, as defined by ISO 30042:2019. For simplicity, the closing tags of elements are not shown; for instance, the term element is shown as `<term>` and not `<term>...</term>`. In some cases sample text is included for demonstration purposes.

Content type

The type of content that is allowed in the data category element. The content types are defined as follows:

- `plainText` - refers to data that can only contain text (PCDATA)
- `basicText` - allows `<hi>` elements for restricted inline formatting
- `noteText` - allows `<hi>`, and additionally `<foreign>`, `<sc>`, `<ec>`, and `<ph>` elements for additional highlighting, formatting, and linking requirements
- `picklist` - allows a limited set of predefined values

For more information about the elements allowed in `basicText` and `noteText`, see the section *Inline markup*.

Section

The section(s) (sometimes referred to as levels) that the data category can appear in. The sections are as follows, are represented in XML as shown, and must appear in terminological entries in this order:

1. concept section - `<conceptEntry>` (formerly `<termEntry>`)
2. language section - `<langSec>` (formerly `<langSet>`)

3. term section - <termSec> (formerly <tig>)

Comment

A comment on the data category, usually about its interpretation and implementation in TBX-Basic.

5.1 Context

Identifier	DC-149
XML markup	<descrip type="context">
Content type	noteText
Section	term
Comment	A sample sentence that contains the term. See <i>7.2 Definitions and contexts</i> for guidelines.

5.2 Created by

Identifier	DC-162
XML markup	<transacGrp> <transac type="transactionType">creation</transac> <transacNote type="responsibility" target="CA5365">John Harris</transacNote> <date>2008-05-12</date> </transacGrp>
Content type	plainText
Section	concept, language, term
Comment	A <code>transacGrp</code> element contains one <code>transac</code> element plus either one <code>transacNote</code> element, or one <code>date</code> element, or both.

5.3 Creation date

Identifier	DC-393
XML markup	<transacGrp> <transac type="transactionType">creation</transac> <transacNote type="responsibility" target="CA5365">John Harris</transacNote> <date>2008-05-12</date> </transacGrp>
Content type	date format (see the Comment)
Section	concept, language, term
Comment	The required format from ISO 8601 is: YYYY-MM-DD, where YYYY is the year, MM is the month, DD is the day. A <code>transacGrp</code> element contains one <code>transac</code> element plus either one <code>transacNote</code> element, or one <code>date</code> element, or both.

5.4 Cross-reference

Identifier	DC-216
XML markup	<code><ref type="crossReference" target="element_id"></code>
Content type	plainText
Section	concept, term
Comment	Pointer to another entry, or to a term in another entry, in the same TBX-Basic file.

5.5 Customer

Identifier	DC-165
XML markup	<code><admin type="customerSubset"></code>
Content type	plainText
Section	concept, term
Comment	Used to identify terms that are required for specific customers.

5.6 Definition

Identifier	DC-168
XML markup	<code><descrip type="definition"></code>
Content type	noteText
Section	concept, language
Comment	<i>do we want any comment here?</i>

5.7 External cross-reference

Identifier	DC-226
XML markup	<code><xref type="externalCrossReference" target="external_ID"></code>
Content type	plainText
Section	concept, term
Comment	Pointer to an external reference for information about the term or concept.

5.8 Figure

Identifier	DC-2920
XML markup	<code><xref type="xGraphic" target="file_location"></code>

Content type	plainText
Section	concept
Comment	Reference (URI, URL, or local file path) to a graphic file that is external to the TBX-Basic file. The reference to the graphic file is recorded as the value of the target attribute of the <code><xref...></code> element. The type value <code>xGraphic</code> identifies this external reference as pointing to a graphic file. The content of the element is the name or description of the file for display purposes. For example: <pre><xref type="xGraphic" target="bat.jpg">cricket bat</xref></pre>

5.9 Gender

Identifier	DC-245
XML markup	<code><termNote type="grammaticalGender"></code>
Content type	picklist permissible values: <ul style="list-style-type: none"> • masculine (DC-246) • feminine (DC-247) • neuter (DC-248) • other (DC-249)
Section	term
Comment	Pointer to an external reference for information about the term or concept.

5.10 Geographical usage

Identifier	DC-243
XML markup	<code><termNote type="geographicalUsage"></code>
Content type	plainText
Section	term
Comment	An indicator of any particular geographical region that the term's usage is restricted to or characteristic of. It is best practice to implement this data category as a picklist. If the picklist values correspond to countries, use the ISO 3166 country codes. If they correspond to locales, use the codes from IETF RFC 4646 or its successor, as identified in IETF BCP 47.

5.11 Language

Identifier	DC-279
XML markup	<code><langSec xml:lang="xx-XX"></code>
Content type	n/a - this element only allows child elements, it does not allow any content

Section	language
Comment	Mandatory attribute for the language section. The language code must be taken from ISO 639-1, ISO 639-2, or ISO 639-3, as recommended in BCP-47: www.rfc-editor.org/rfc/bcp/bcp47.txt

5.12 Last modified by

Identifier	DC-451 and DC-4337
XML markup	<pre><transacGrp> <transac type="transactionType">modification</transac> <transacNote type="responsibility" target="CA5365">John Harris</transacNote> <date>2008-05-12</date> </transacGrp></pre>
Content type	plainText
Section	concept, language, term
Comment	A <code>transacGrp</code> element contains one <code>transac</code> element plus either one <code>transacNote</code> element, or one <code>date</code> element, or both.

5.13 Last modified date

Identifier	DC-304 and DC-2526
XML markup	<pre><transacGrp> <transac type="transactionType">modification</transac> <transacNote type="responsibility" target="CA5365">John Harris</transacNote> <date>2008-05-12</date> </transacGrp></pre>
Content type	plainText
Section	concept, language, term
Comment	The required format from ISO 8601 is: YYYY-MM-DD, where YYYY is the year, MM is the month, DD is the day. A <code>transacGrp</code> element contains one <code>transac</code> element plus either one <code>transacNote</code> element, or one <code>date</code> element, or both.

5.14 Note

Identifier	DC-382
XML markup	<note>
Content type	noteText
Section	concept, language, term

Comment	Any kind of note, such as a usage note, explanation, or instruction.
---------	----------------------------------------------------------------------

5.15 Part of speech

Identifier	DC-396
XML markup	<code><termNote type="partOfSpeech"></code>
Content type	picklist permissible values: <ul style="list-style-type: none"> • noun (DC-1333) • verb (DC-1424) • adjective (DC-1230) • adverb (DC-1232) • properNoun (DC-384) • other (DC-4336)
Section	term
Comment	Mandatory under certain conditions. See section 9 <i>Compliance</i> for details.

5.16 Project

Identifier	DC-406
XML markup	<code><admin type="projectSubset"></code>
Content type	plainText
Section	concept, term
Comment	Used to identify terms that are required for or specific to particular jobs or projects.

5.17 Source of context

Identifier	DC-471
XML markup	<code><admin type="source"></code>
Content type	noteText
Section	term
Comment	It is best practice to document the source of the context. This element shall occur in a <code><descripGrp></code> element so that it can be associated with a context, as in the following example: <pre><descripGrp> <descrip type="context">This is a sample context.</descrip> <admin type="source">New York Times</admin> </descripGrp></pre>

5.18 Source of definition

Identifier	DC-471
XML markup	<code><admin type="source"></code>
Content type	noteText
Section	concept, language
Comment	<p>It is best practice to document the source of the definition. This element shall occur in a <code><descripGrp></code> element so that it can be associated with a definition, as in the following example:</p> <pre><descripGrp> <descrip type="definition">This is a sample definition.</descrip> <admin type="source">Webster's Dictionary</admin> </descripGrp></pre>

5.19 Source of term

Identifier	DC-471
XML markup	<code><admin type="source"></code>
Content type	noteText
Section	term
Comment	

5.20 Subject field

Identifier	DC-489
XML markup	<code><descrip type="subjectField"></code>
Content type	plainText (see Comment)
Section	concept
Comment	It is best practice to implement this data category as a picklist. You can specify the picklist values in a multi-level hierarchy. See section 7.3.

5.21 Term

Identifier	DC-508
XML markup	<code><term></code>
Content type	basicText
Section	term
Comment	Mandatory data category; there must be at least one in an entry.

5.22 Term location

Identifier	DC-1823
XML markup	<code><termNote type="termLocation"></code>
Content type	plainText
Section	term
Comment	Refers to a location in the source materials—such as a software application user interface, product packaging, or an industrial process—where the term frequently occurs. It is best practice to implement this data category as a picklist. A list of possible picklist values for user interface locations is available in DatCatInfo (datcatinfo.net).

5.23 Term type

Identifier	DC-2677
XML markup	<code><termNote type="termType"></code>
Content type	picklist permissible values: <ul style="list-style-type: none"> • fullForm (DC-321) • acronym (DC-334) • abbreviation (DC-331) • shortForm (DC-332) • variant (DC-330) • phrase (DC-339)
Section	term
Comment	See section 7.1 for guidelines.

5.24 Usage status

Identifier	DC-70
XML markup	<code><termNote type="administrativeStatus"></code>
Content type	picklist permissible values: <ul style="list-style-type: none"> • preferred (DC-72) • admitted (DC-73) • notRecommended (DC-74) • obsolete (DC-75)
Section	term
Comment	Used for controlled authoring and controlled translation purposes, to mark term usage preferences. Only one value is permitted for each term.

6. Additional information about data categories

6.1 Term type picklist values

The **term type** data category is optional. When a term has no term type value, it is assumed to be an ordinary entry term, i.e. it is not an abbreviation or a variant or a full form, etc., of another term. The values should be used as described below.

fullForm

The complete representation of a term for which there is an abbreviated form.

abbreviation

An abbreviated form that is formed by omitting letters from a full form. Example:

- full form: corporation
- abbreviation: corp

shortForm

An abbreviated form that includes fewer words than the full form. Example:

- full form: *log-likelihood relationship measure*
- short form: *log-likelihood measure*

acronym

An abbreviated form that is made up of the initial letters of the components of the full form or from the syllables of the full form. Example:

- full form: *access control list*
- acronym: *ACL*

variant

An alternative form of a term other than an abbreviated form. Variants can include words that have an alternative spelling, punctuation, capitalization, word formation, or even a numeric representation. Example:

- term: *soft switch*
- variant: *softswitch*

phrase

Any group of two or more words that are frequently expressed together and that consist of more than one concept. The individual words in a phrase usually function in more than one grammatical category (part of speech) within the syntax of a sentence. Although phrases comprise more than one concept, they are often stored in single concept entries in terminological databases to address a need to record their translations for end-users. Examples: *send feedback*, *work offline*

Note: a noun that comprises more than one word but represents one concept is NOT a **phrase**. This is the case for the majority of multi-word terms. For example, the terms *airport runway light* and *voice over Internet Protocol* are not phrases, even though they comprise more than one word.

6.2 Definitions and contexts

A definition describes the meaning of a term. A context is a sample sentence that contains the term. In TBX-Basic, either a definition or a context sentence shall be present in all entries.

6.2.1 Definitions

ISO 704:2022 - Terminology work – Principles and methods provides guidelines for writing definitions.

Following the ISO model, a definition is a sentence that explains the meaning of a term by a) identifying the class to which the term belongs and b) describing the characteristics that distinguish this term from other terms in this class. For example:

- term: *pacemaker*
- definition: implantable medical device that treats abnormal heart rhythms

6.2.2 Context sentences

Context sentences must be "authentic," that is, they must be retrieved from an actual document or other source; the terminologist should never create or translate a context sentence. Context sentences should also be retrieved from original (not translated) documents, and they should be a complete sentence. Contexts that provide useful semantic and/or linguistic information should be prioritized.

6.3 Data categories used for categorizing purposes

The majority of organizations that manage terminology also collect categorical information, such as subject fields and product or service categories. TBX-Basic provides three data categories for this purpose: subject field, customer, and project.

Well-organized and consistently-applied categorical information can play an important role in clarifying the meaning of a term, where the term is used, or how the term is differentiated from other terms. Categorical information, combined with grammatical and contextual information, can help to narrow the scope of a term. In this sense, it has great practical value in the translation process.

Categorical information can also be used for filtering and sorting.

At least one field with categorical information per concept (entry) is recommended for termbases. A widely used categorical data field is **subject field** (sometimes called "domain"). ISO TC37 defines *subject field* as "a field of special knowledge". Depending on the terminology management software, it may be possible to specify subject field values in a multi-level hierarchy (for example: energy -> renewable energy -> solar power). For guidance on how to represent hierarchical subject field taxonomies in TBX, see *ISO/TS 24634:2021 TBX-compliant representation of concept relations and subject fields*.

Typically, categorical values are chosen from a picklist.

Multiple subject fields can be assigned to a concept.

7. The Structure of a TBX-Basic Entry

Regarding entry structure, TBX-Basic complies with TMF (*ISO 16642: Terminology Markup Framework*). However, certain additional restrictions have been placed on the entry structure.

7.1 Concept section

This section contains elements whose immediate parent is or can be `<conceptEntry>`. All are repeatable. With the exception of `langSec`, all are optional. The `<langSec>` element(s) must come after the others, if used.

1. `descrip` - used for subject fields and definitions. Subject fields occur only at the concept level. Definitions can occur at other levels, as described in the following sections.
2. `descripGrp` - used to document a definition and its source. Contains one `descrip` and one `admin` element.
3. `admin` - used for the name of a customer or project that the entry is associated with, or for the source of a definition.

4. `transacGrp` - used for administrative information, such as the date that the entry was created and the name of the person who created it. Contains: one `transac`, and either one or both of `transacNote`, and `date`.
5. `note` - any concept-level note information.
6. `ref` - used for an internal cross-reference, with the target attribute pointing to the concept ID of another entry.
7. `xref` - used for an external reference, such as a URL of a web site, or to point to an external graphic file.

`langSec` - a container element for a language section. At least one is required in a `conceptEntry`.

7.2 Language section

This section contains elements whose immediate parent is or can be `<langSec>`. All are repeatable. With the exception of `<termSec>`, all are optional. The `<termSec>` element(s) must come after the others.

1. `descrip` - used for a definition documented at the language level. This position therefore allows for definitions in different languages.
2. `descripGrp` - used to document a definition and its source. Contains one `descrip` and one `admin` element.
3. `transacGrp` - used for administrative information about the language section, such as the date that the language section was created and the name of the person who created it. Contains one `transac`, and either one or both of `transacNote`, and `date`.
4. `note` - any language-level note
5. `termSec` - a container element for one term and associated term-level elements. Each `termSec` contains information about one term. At least one is required in a `langSec`.

7.3 Term section

This section contains elements whose immediate parent is or can be `<termSec>`. With the exception of `<term>` which is required but only one per `<termSec>`, all are optional and repeatable. The order must be `<term>` followed by zero or more `<termNote>` elements, followed by the remaining in any order.

1. `term` - contains the term. One (and only one) is required.
2. `termNote` - contains information about the term, such as the part of speech, or term type.
3. `descrip` - used at this level only for the context sentence.
4. `descripGrp` - used to document a context and its source. Contains: one `descrip` and one `admin` element.
5. `admin` - used to document the source of the term or of a context, or the name of a customer or project that the term is associated with. If used for the source of a context, both are nested in a `<descripGrp>`.

6. `transacGrp` - used for administrative information about the term, such as the date that the term was added to the entry and the name of the person who added it. Contains: one `transac`, and either one or both of `transacNote`, and `date`.
7. `note` - any note about the term.
8. `ref` - used for an internal cross-reference, with the `target` attribute pointing to the ID of a term in another entry.
9. `xref` - used for an external reference providing term-related information, such as the URL of a website.

7.4 Backmatter

The backmatter of a TBX-Basic file is used to record the names and contact information for people who are responsible for creating or updating the terminology entries. The following is a sample of the markup allowed in the backmatter. The values of the `type` attribute for the `<item>` elements are taken from the vCard standard; more values are available, such as to record telephone numbers.

```
<back>
  <refObjectSec type="respPerson">
    <refObject id="US5001">
      <item type="fn">Jane Doe</item>
      <item type="email">jane_doe@mymail.com</item>
      <item type="role">approver</item>
    </refObject>
    <refObject id="US5002">
      <item type="fn">John Smith</item>
      <item type="email">john_smith@mymail.com</item>
      <item type="role">inputter</item>
    </refObject>
  </refObjectSec>
</back>
```

8. Inline markup

The use of markup for presentational formatting, such as bold, underline, italics, and so forth, is uncommon and in fact not recommended in terminology resources at least in their native format, which is usually a database or an XML file (e.g. TBX or other proprietary XML-based formats). Such formatting may be found in tertiary resources intended for human presentation, such as glossaries in MSWord. However, there can be a need to represent minimal core markup such as for mathematical expressions and formulae, and for links to other terms (entailed terms) in text fields such as definitions. TBX supports four elements for inline markup, based on function not style, to accommodate such needs: `<hi>`, `<sc>`, `<ec>`, and `<ph>`.

The `<hi>` element delimits a section of text for various processing purposes, such as to mark a mathematical expression, or to mark an entailed term in a definition or another text field. It is also allowed within `<term>` elements to mark portions of terms, such as a character in subscript or superscript, or a component within the term that is italicized or otherwise requires different formatting. The following example shows an entailed term in the definition of the term “canopy”.

```
<descrip type="definition">the expanding, umbrella-like part of a
<hi type="entailedTerm" target="CID1234">parachute</hi></descrip>
```

The `<sc>` (formerly `<bpt>`) and `<ec>` (formerly `<ept>`) elements are used as a pair to encapsulate markup inherited from another format such as an XLIFF file. The `<ph>` element is used to represent a native standalone code or a sequence of such codes (such as an HTML `
` tag). These three elements are adopted from *ISO 21720 (XLIFF Version 2.0)*. Refer to that standard or *ISO 30042:2019 (TBX)* for more information and markup samples.

Aside from these four markup elements, there is also the `<foreign>` element which is used to mark a segment of text that is in a language different from that of the surrounding text. The optional `xml:lang` attribute may be used to identify the language of the text contained in this element.

9. Compliance

A terminology resource is compliant with TBX-Basic if it meets all the following conditions:

- It validates against the TBX Core Structure (`core_structure.xsd` or `core_structure.rng`) and the additional TBX-Basic constraints (`basic_schema.sch`). (See Appendix C.)
- It uses only the data categories that are defined in this document, and uses them in the manner described in this document in terms of the nature and type of data.
- The data categories are inserted at the correct level of the entry structure as specified in this document.
- It respects the usage guidelines and best practices outlined in this document.
- One of the following conditions has been met:
 - If the resource is to be submitted to any form of automated processing, each term section (`<termSec>`) has a **Part of speech** explicitly indicated.
 - If the resource is only for human consultation, the **Part of speech** may be omitted if either a **Definition** or a **Context** is provided.

10. Recommended standards and guidelines

In addition to ISO 30042 Management of terminology resources — TermBase eXchange (TBX), some of the most relevant terminology standards published by the International Organization for Standardization (ISO) are listed below:

- ISO 704 Terminology work – Principles and methods
- ISO 860 Terminology work – Harmonization of concepts and terms
- ISO 16642 Computer applications in terminology – Terminological markup framework
- ISO/TR 24633-1 Management of terminology resources — Companion to TermBase eXchange (TBX) — Part 1: General
- ISO/TR 24633-2 Management of terminology resources — Companion to TermBase eXchange (TBX) — Part 2: RNG schema for TBX core
- ISO/TR 24633-3 Management of terminology resources — Companion to TermBase eXchange (TBX) — Part 3: RDF representation for TBX core
- ISO/TS 24364 Management of terminology resources — TermBase eXchange (TBX)-compliant representation of concept relations and subject fields

- ISO 26162-1 Management of terminology resources — Terminology databases — Part 1: Design
- ISO 26162-2 Management of terminology resources — Terminology databases — Part 2: Software
- ISO 26162-3 Management of terminology resources — Terminology databases — Part 3: Content
- ISO 26162-4 Management of terminology resources — Terminology databases — Part 4: Quality

Also, TerminOrgs publishes a *Terminology Starter Guide*, available at www.terminorgs.net.

Appendix A - Data categories by section of the entry

This appendix lists the TBX-Basic data categories according to the sections of the entry in which they occur, in the recommended order of their occurrence. Strictly speaking, all data categories are optional except the term (at least one term must occur in the entry), as well as the specification of the language attribute in its parent `<langSec>` tag. However, it is recommended to adhere to the guidelines provided earlier in this document with respect to the use of the part-of-speech, subject field, definition, and context.

Refer to section 6 *Data category descriptions* for further details such as the `DatCatInfo` identifier and the XML markup. Note that the Definition and the Source of definition has been listed in both the Concept section and the Language section; either is possible. Typically they occur in the Concept section, however, some termbases include them in the Language section to allow for definitions in different languages.

Concept section data categories

1. Subject field
2. Definition
3. Source of definition
4. Customer
5. Project
6. Note
7. Cross-reference
8. External cross-reference
9. Figure
10. Created by
11. Creation date
12. Last modified by
13. Last modified date

Language section data categories

1. Definition
2. Source of definition
3. Note
4. Created by
5. Creation date
6. Last modified by
7. Last modified date

Term section data categories

1. Term
2. Part of speech
3. Gender
4. Term type
5. Geographical usage
6. Term location
7. Usage status
8. Context

9. Source of context
10. Source of term
11. Customer
12. Project
13. Note
14. Cross-reference
15. External cross-reference
16. Created by
17. Creation date
18. Last modified by
19. Last modified date

Appendix B – Properties of a valid TBX-Basic document instance

1. The first three lines define the file as being of type XML, specify the encoding, and point to the schema files for validation. The following is an example. In the example the core schema file is the XSD version. Modify that line for the alternative RNG version if used.

```
<?xml version="1.0" encoding="UTF-8"?>
<?xml-model href="core_schema.xsd" type="application/xml" schematypens="http://www.w3.org/2001/XMLSchema"?>
<?xml-model href="basic_schema.sch" type="application/xml" schematypens="http://purl.oclc.org/dsdl/schematron"?>
```

2. The TBX root element must include declarations of the Core module namespace, the type (name of the dialect) and the style (DCA):

```
<tbx type="TBX-Basic" style="dca" xml:lang="en" xmlns="urn:iso:std:iso:30042:ed:3.0">
```

3. The root element is followed by some file information in the tbxHeader element followed by the text and body elements. The following is an example:

```
<tbxHeader>
  <fileDesc>
    <titleStmt>
      <title>Sample file for TBX-Basic</title>
    </titleStmt>
    <sourceDesc>
      <p>Sample file for TBX-Basic</p>
    </sourceDesc>
  </fileDesc>
</tbxHeader>
<text>
<body>
```

4. The body element contains the concept entries, followed by the optional backmatter (see section 7.4) and closing elements.

```
<body>
  <conceptEntry>
```

```

...
</conceptEntry>
<conceptEntry>
...
</conceptEntry>
...
<back>
...
</back>
</body>
</text>
</tbx>

```

5. Only the XML elements listed below are permitted (shown here in alphabetical order, see Appendix A for a listing by section in the recommended order).

- <admin>
- <back>
- <body>
- <change>
- <conceptEntry>
- <date>
- <descrip>
- <descripGrp>
- <ec>
- <fileDesc>
- <foreign>
- <hi>
- <item>
- <langSec>
- <note>
- <p>
- <ph>
- <publicationStmt>
- <ref>
- <refObject>
- <refObjectSec>
- <revisionDesc>
- <sc>
- <sourceDesc>
- <tbx>
- <tbxHeader>
- <termNote>
- <termSec>
- <text>
- <title>
- <titleStmt>

- <transac>
- <transacGrp>
- <transacNote>
- <xref>

In TBX-Basic, the <descripGrp> element is used only to associate a source to a definition or to a context. The following child elements are not supported inside a <descripGrp>: <descripNote>, <admin>, <adminGrp>, <note>, <ref>, and <xref>.

The following elements, which are part of the TBX markup language, are NOT allowed in TBX-Basic:

- <adminGrp>
- <adminNote>
- <descripNote>
- <itemGrp>
- <itemSet>
- <termGrp>
- <termNoteGrp>

Appendix C – Files for Validation

The following files are provided on the TerminOrgs web site for validation and testing purposes (*add the URLs*)

`core_schema.xsd` and `core_schema.rng`

These two files are used to validate compliance with the TBX core schema. They are equivalent, and only one is required. Chose the file that corresponds to your preferred validation method, RelaxNG or XML Schema.

`basic_schema.sch`

This file is required to validate the additional constraints required in TBX-Basic.

`TBX-Basic-sample.tbx`

A sample TBX-Basic document instance, for demonstration purposes only.