Terminology for Large Organizations

TBX-BASIC

Version 3.1, © 2014

For the most recent version of the TBX specification, please go to www.terminorgs.net/tbx-basic.html

Table of contents

1 Document information....3

2 Background....4

3 Introduction....4

4 Where to get more information and resources....5

5 Mandatory data categories....5

6 Data category descriptions....5

6.1 Context....6

6.2 Created by....7

6.3 Creation date....7

6.4 Cross-reference....7

6.5 Customer....8

6.6 Definition....8

6.7 External cross-reference....8

6.8 Gender....8

6.9 Geographical usage....9

6.10 Figure....9

6.11 Language....10

6.12 Last modified by....10

6.13 Last modified date....10

6.14 Note....11

6.15 Part of speech....11

6.16 Project....12

6.17 Subject field....12

6.18 Source of context....12

6.19 Source of definition....13

6.20 Source of term....13

6.21 Term....14
6.22 Term location....14
6.23 Term type....14
6.24 Usage status....15
7 Additional information about data categories....16
7.1 Term type picklist values....16
7.2 Definitions and contexts....18
7.3 Data categories used for categorizing purposes....20
8 The structure of a TBX-Basic entry....21
8.1 Concept level....21
8.2 Language level....22
8.3 Term level....22
8.4 Backmatter....23
8.5 Structural differences between TBX-Default and TBX-Basic....23
9 Compliance....24
9.1 Conditions for compliance....24
9.2 Compliance-checking TBX-Basic files....24
10 Recommended standards and guidelines....27
Appendix A – Term location picklist values for Windows user interface objects....28
Appendix B – Data categories by levels of entry....29

1 Document information

Owner

TerminOrgs
- www.terminorgs.net
- info@terminorgs.net

Version

Version 3.1

This document was updated on September 12, 2014. In this update, ISO TC37 Data Category Registry identifiers for the following data categories were corrected:

6.12 Last modified by

6.13 Last modified date

6.23 Term type

Copyright

Typographical conventions

bold	Used for the names of data categories
italics	Used for sample terms
`monospace font`	Used for the XML elements and attributes

Terms

data category

A type of data field, such as **definition**.

terminology resource

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

terminological markup language (TML)

An XML markup language for terminology resources. TBX-Basic is a TML.

termbase

A terminology database.

2 Background

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.

TBX-Basic is compliant with ISO 30042, *TermBase eXchange (TBX)*, the ISO standard for terminology data.

3 Introduction

TBX-Basic is a terminological markup language (TML) that is a lighter version of TBX-Default, the TML that is defined in ISO 30042. TBX-Basic is designed for the localization industry and 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[1].

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.

---

[1] Terminology Survey Results (2001) www.lisa.org/LISA-Terminology-Sur.460.0.html

4 Where to get more information and resources

Information about TBX and TBX-Basic, as well as tools and resources, such as files for compliance checking, are available at www.terminorgs.net.

5 Mandatory data categories

There are only two mandatory data categories in TBX-Basic: **term**, and **language**.

Several of the remaining data categories, including **definition**, **context**, **part of speech**, and **subject field** are very important and should be included in a terminology whenever possible. The most important non-mandatory data category is **part of speech**.

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

6 Data category descriptions

The data category names used in this section have been adopted for TBX-Basic for convenience purposes and some may not therefore correspond to the official names in other resources such as ISO 30042 and the ISO TC37 Data Category Registry. The official name can be obtained by opening the Identifier URL.

The data category descriptions in this section contain the following components:

Identifier The identifier of the element taken from the ISO TC37 Data Category Registry, which is available at www.isocat.org.

XML representation The mandatory XML representation of the data category, as defined by TBX. For simplicity, the closing tags of elements are not shown; for instance, the term element is shown as `<term>` and not `<term>...</term>`.

Content type The type of content that is allowed in the data category element. Most of the content types are from TBX-Default. However, to facilitate data exchange, TBX-Basic provides some content types that are more precise. For instance, in TBX-Default, **part of speech** is plainText, but, in TBX-Basic, it is picklist. The content types are as follows:

Type	Description
plainText	Refers to data that can contain only text (PCDATA)
basicText	Allows `<hi>` elements for restricted inline formatting
noteText	Allows additional embedded elements such as for highlighting, formatting, and linking
picklist	Allows a limited set of predefined values

Level The levels that the data category can appear in. The levels are as follows, are represented in XML as shown, and must be defined in this order:

Level	XML representation
Concept	termEntry
Language	langSet
Term	tig

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

6.1 Context

Identifier www.isocat.org/datcat/DC-149

XML representation `<descrip type="context">`

Content type noteText

Level Term

Comment A sample sentence that contains the term. See *7.2 Definitions and contexts* for guidelines.

6.2 Created by

Identifier www.isocat.org/datcat/DC-162

XML representation

```
<transacGrp>
      <transac type="transactionType">creation</transac>
      <transacNote type="responsibility" target="CA5365">John
      Harris</transacNote>
      <date>2008-05-12</date>
</transacGrp>
```

Content type plainText

Level Concept, Language, Term

Comment A `transacGrp` element can contain either one `transacNote` element, or one `date` element, or both.

6.3 Creation date

Identifier www.isocat.org/datcat/DC-393

XML representation

```
<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 below)

Level 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 `transacGrp` element can contain either one `transacNote` element, or one `date` element, or both.

6.4 Cross-reference

Identifier	www.isocat.org/datcat/DC-216
XML representation	`<ref type="crossReference" target="element_id">`
Content type	plainText
Level	Concept, Term
Comment	Pointer to another entry, or to a term in another entry, in the same TBX-Basic file.

6.5 Customer

Identifier	www.isocat.org/datcat/DC-165
XML representation	`<admin type="customerSubset">`
Content type	plainText
Level	Concept, Term
Comment	Used to identify terms that are required for specific customers.

6.6 Definition

Identifier	www.isocat.org/datcat/DC-168
XML representation	`<descrip type="definition">`
Content type	noteText
Level	Concept, Language

6.7 External cross-reference

Identifier	www.isocat.org/datcat/DC-226
XML representation	`<xref type="externalCrossReference" target="external_id">`

Content type plainText

Level Concept, Term

Comment Pointer to an external reference for information about the term or concept.

6.8 Gender

Identifier www.isocat.org/datcat/DC-245

XML representation `<termNote type="grammaticalGender">`

Level Term

Content type Picklist, with permissible values as follows:

- `masculine`
- `feminine`
- `neuter`
- `other`

6.9 Geographical usage

Identifier www.isocat.org/datcat/DC-243

XML representation `<termNote type="geographicalUsage">`

Content type plainText

Level Term

Comment 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.

6.10 Figure

Identifier www.isocat.org/datcat/DC-2920

XML representation `<xref type="xGraphic" target="file_location">`

Content type plainText

Level 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 `<xref...>` element. The type value `xGraphic` identifies this external reference as pointing to a graphic file. The content of the `<xref...>` element is the name or description of the file for display purposes. For example: `<xref type="xGraphic" target="bat.jpg">cricket bat</xref>`

6.11 Language

Identifier	www.isocat.org/datcat/DC-279
XML representation	`<langSet xml:lang="xx-XX">`
Content type	Language code
Level	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

6.12 Last modified by

Identifier	www.isocat.org/datcat/DC-451 and www.isocat.org/datcat/DC-4337

XML representation	`<transacGrp>`
	`  <transac type="transactionType">modification</transac>`
	`  <transacNote type="responsibility" target="CA5365">John Harris</transacNote>`
	`  <date>2008-05-12</date>`
	`</transacGrp>`

Content type	plainText
Level	Concept, Language, Term
Comment	A `transacGrp` element can contain either one `transacNote` element, or one `date` element, or both.

6.13 Last modified date

Identifier www.isocat.org/datcat/DC-305, and www.isocat.org/datcat/DC-2526

XML representation

```
<transacGrp>
      <transac type="transactionType">modification</transac>
      <transacNote type="responsibility" target="CA5365">John
      Harris</transacNote>
      <date>2008-05-12</date>
</transacGrp>
```

Content type date format (see comment below)

Level 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 `transacGrp` element can contain either one `transacNote` element, or one `date` element, or both.

6.14 Note

Identifier www.isocat.org/datcat/DC-382

XML representation `<note>`

Content type noteText

Level Concept, Language, Term

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

6.15 Part of speech

Identifier www.isocat.org/datcat/DC-396

XML representation `<termNote type="partOfSpeech">`

Content type picklist

Permissible values and their ISOcat PIDs are as follows:

Value	ISOcat PID
`noun`	www.isocat.org/datcat/DC-1333
`verb`	www.isocat.org/datcat/DC-1424
`adjective`	www.isocat.org/datcat/DC-1230
`adverb`	www.isocat.org/datcat/DC-1232
`properNoun`	www.isocat.org/datcat/DC-384
`other`	www.isocat.org/datcat/DC-4336

Level Term

Comment Mandatory under certain conditions. See *9 Compliance* for details.

In TBX-Default, the data type for **part of speech** is plainText. TBX-Basic's use of picklist is in compliance with TBX-Default because picklist is more constrained than plainText.

The `other` value can be used for terms of the `phrase` type.

6.16 Project

Identifier www.isocat.org/datcat/DC-406

XML representation `<admin type="projectSubset">`

Content type plainText

Level Concept, Term

Comment Used to identify terms that are required for or specific to particular jobs or projects.

6.17 Subject field

Identifier www.isocat.org/datcat/DC-489

XML representation `<descrip type="subjectField">`

Content type plainText

Level 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. The way you define the hierarchy depends on the terminology application that you use.

6.18 Source of context

Identifier www.isocat.org/datcat/DC-471

XML representation `<admin type="source">`

Content type noteText

Level Term

Comment It is best practice to document the source of the context. In the user interface, you can use the less ambiguous label "Source of context."

This element shall occur in a `<descripGrp>` element so that it can be associated with a context, as in the following example:

```
<descripGrp>
      <descrip type="context">This is a sample context.</descrip>
      <admin type="source">New York Times</admin>
</descripGrp>
```

6.19 Source of definition

Identifier www.isocat.org/datcat/DC-471

XML representation `<admin type="source">`

Content type noteText

Level Concept, Language

Comment It is best practice to document the source of the definition. In the user interface, you can use the less ambiguous label "Source of definition."

This element shall occur in a `<descripGrp>` element so that it can be associated with a definition, as in the following example:

```
<descripGrp>
      <descrip type="definition">This is a sample
      definition.</descrip>
      <admin type="source">Webster's Dictionary</admin>
</descripGrp>
```

6.20 Source of term

Identifier www.isocat.org/datcat/DC-471

XML representation `<admin type="source">`

Content type noteText

Level Term

Comment In the user interface, you can use the less ambiguous label "Source of term."

6.21 Term

Identifier www.isocat.org/datcat/DC-508

XML representation `<term>`

Content type basicText

Level Term

Comment Mandatory data category; there must be at least one in an entry.

6.22 Term location

Identifier www.isocat.org/datcat/DC-1823

XML representation `<termNote type="termLocation">`

Content type plainText

Level Term

Comment Refers to a location in the corpus—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. See *Appendix A – Term location picklist values for Windows user interface objects*.

6.23 Term type

Identifier www.isocat.org/datcat/DC-2677

XML representation `<termNote type="termType">`

Content type picklist

Permissible values and their ISOcat PIDs are as follows:

Value	ISOcat PID
`fullForm`	www.isocat.org/datcat/DC-321
`acronym`	www.isocat.org/datcat/DC-334
`abbreviation`	www.isocat.org/datcat/DC-331
`shortForm`	www.isocat.org/datcat/DC-332
`variant`	www.isocat.org/datcat/DC-330
`phrase`	www.isocat.org/datcat/DC-339

Level Term

6.24 Usage status

Identifier www.isocat.org/datcat/DC-70

XML representation `<termNote type="administrativeStatus">`

Content type

picklist, with values as shown in the following table

In the TBX-Basic file, you can use the permissible values in the first column of the following table. However, in XCS files, you must use the values (which are taken from TBX-Default) in the second column.

TBX-Basic permissible values	XCS required values and ISOcat PID	Description
`preferred`	`preferredTerm-admn-sts` www.isocat.org/datcat/DC-72	The term that, among a set of synonymous terms, is most recommended for use.
`admitted`	`admittedTerm-admn-sts` www.isocat.org/datcat/DC-73	The term is acceptable for use.
`notRecommended`	`deprecatedTerm-admn-sts` www.isocat.org/datcat/DC-74	The term should not be used.
`obsolete`	`supersededTerm-admn-sts` www.isocat.org/datcat/DC-75	The term is no longer used, usually because a more modern term has replaced it.

Level

Term

Comment

Used for controlled authoring and controlled translation purposes, to mark whether a term is approved or not recommended for use. Only one value is permitted for each term.

7 Additional information about data categories

7.1 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 that is not an abbreviation or a variant of another term or an abbreviation of another full form term.

Picklist value	Description	
`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 longer form.	
	Full form	**Abbreviation**
	corporation	corp.
	lower bound	LB
`shortForm`	An abbreviated form that includes fewer words than the full form.	
	Full form	**Short form**
	Intergovernmental Group of Twenty-four on Monetary Affairs	Group of Twenty-four
	log-likelihood relationship measure	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.	
	Full form	**Acronym**
	access control list	ACL
	Extensible Markup Language	XML

<table>
<tr><th>Picklist value</th><th>Description</th></tr>
<tr><td><code>variant</code></td><td>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.<table><tr><th>Term</th><th>Variant</th></tr><tr><td>Yahoo</td><td>Yahoo!</td></tr><tr><td>soft switch</td><td>softswitch</td></tr></table></td></tr>
<tr><td><code>phrase</code></td><td>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</td></tr>
</table>

Note that there is no term type value of “synonym” or “translation”. This is because all terms in a concept entry within the same language section are synonyms, and all terms in a concept entry in different language sections are assumed to be possible translations of each other.

Use the following examples as guidelines for assigning **termType** values:

- The term *blue box*, which is a recycling box in Canada, does not have any abbreviated forms, and it is not an alternate form of another term. Therefore, you do not need to specify a **termType** value.
- The term *cell phone* is a short form of the more correct form *cellular phone*. Therefore, *cellular phone* should have a **termType** value of `fullForm` and *cell phone* should have a **termType** value of `abbreviation`.
- The term *application program interface* and *application programming interface* are interchangeably used. However, the more common form of the term is *application programming interface*. The term *application program interface* could, therefore, be considered a type of spelling variant. You do not need to specify a **termType** value for *application programming interface*, but for *application program interface*, you should specify a **termType** value of `variant`.

7.2 Definitions and contexts

It is very useful to translators to have some type of conceptual description in the entry, so that they can be sure of the meaning of the term and whether or not the suggested translation is suitable for the text they are translating. In TBX-Basic, two data categories are provided for this purpose: **Definition** and **Context**.

A definition describes the meaning of a term. A context is a sample sentence that contains the term. Definitions provide meaningful explanatory information, whereas sometimes context sentences do not. It is recommended to include definitions in terminological entries; however, they are time-consuming to prepare. If you do not have the time to prepare definitions, include a context sentence in the entry. Context sentences can be automatically extracted by some term extraction software and some translation alignment software. A definition *or* a context sentence should be provided whenever possible. Thus, for TBX-Basic, either a definition or a context sentence shall be present in all entries.

7.2.1 Definitions

The ISO 704:2009 standard, *Terminology work–Principles and methods*, provides comprehensive 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:

pacemaker

implantable medical device that treats abnormal heart rhythms

7.2.2 Context sentences

Not all context sentences are equally valuable. Some are very useful, while others are not worth recording at all because they provide no useful information about the term.

Context sentences must always be "authentic," that is, they must be retrieved from an actual document or other communication medium; the terminologist should never create a context sentence. Context sentences should also be retrieved from original (not translated) documents, and they should be a complete sentence.

Context sentences that are automatically extracted from a corpus should be reviewed manually before they are imported into a terminology database. Automated extractors usually do not have any selection criteria for determining which contexts are better than others. Furthermore, some automatically extracted contexts may not be full sentences, and may even just constitute the term itself, such as a simple user interface label. These poor context sentences should be replaced by better ones.

7.2.2.1 Purposes of context sentences

Context sentences serve the following purposes:

- They prove that the term actually exists in real language.
- They can clarify the meaning of the term.
- They can provide additional information about the term that is not in the definition (the who, why, when, where, how).
- They can illustrate how the term is used in discourse (collocations, register, etc.). For instance, a context sentence could alert the translator that the term is colloquial.
- They can provide grammatical information (such as gender), stylistic clues (such as hyphenation or capitalization), and alternate forms (abbreviations and so forth).
- The requirement to include a context sentence for the target language term helps to prevent the terminolgist from simply "translating" the source language term, by requiring him or her to find an equivalent designation of the concept that is actually in use in the target language. This helps to ensure authenticity of the target language term and helps to reduce influence of the source language on the target language.

7.2.2.2 Types of context sentences

The following is a description of the various types of context sentences, arranged in descending order of preference for terminological entries.

Definitional context

A definitional context explains the meaning of a term. It defines the term, but not necessarily the rigid form of a definition. For example:

Connection Manager

The Connection Manager is a utility for managing all of the registered connections to workspaces and repositories.

Encyclopedic context

An encyclopedic context provides some information about the meaning of the term, but not enough to fully understand the concept. For example, rather than telling you what something is, it may tell you where or how it is used. For example:

navigation controller

The navigation controller allows navigation from panel to panel, data mapping across the process, and execution of operations in response to certain events.

Associative context

An associative context provides little or no information about the meaning of the term, but it reveals the subject area of the term by virtue of the other associated words in the context. The following example enables us to determine that the term comes from the field of banking:

transaction posting engine

For a bank teller application, access to the services of these entities (for example, to conduct a withdrawal transaction) requires delivery channels and a transaction posting engine that can handle the many tasks involved with transaction processing.

Metalinguistic context

A metalinguistic context describes some linguistic feature of the term. It is a sentence about the term itself, not about the concept. For example:

Blue box

When Blue Box designates the recycling receptacle, it should be written with a capital B.

Discursive context

A discursive context shows only that the term is actually used. It provides no explanatory or linguistic information. For example:

virtual tester

Use the technique described here to enable multiple virtual testers.

7.3 Data categories used for categorizing purposes

The majority of localization companies that manage terminology also collect categorical information, such as subject fields, product identifiers, and so forth. For translation and localization business processes, this kind of information is important. 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. Many localization companies need to filter their terminology data for use in specific translation environments or for reviewing purposes, and they use categorical fields to do this.

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 a subject field as "a field of special knowledge".[2]

Typically, categorical information like subject fields are chosen from a picklist. The use of a picklist ensures that the values are entered correctly and that they are standardized.

Multiple subject fields can be assigned to a concept, both on the same level (biology + chemistry) or in a hierarchical structure (technical -> heavy machinery -> wheel loaders).

8 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 for simplification purposes. The following sections describe the hierarchical entry structure of TBX-Basic.

8.1 Concept level

This level contains elements whose immediate parent is or can be `<termEntry>`. They are, in this order:

1. `descrip` - used for subject fields and definitions. Subject fields should occur only at the concept level. Definitions can occur at other levels, as described in the following sections.
2. `descripGrp` - used instead of descrip to document a definition and its source. Contains: one descrip and one admin element. If the source of the definition is not required or available, use only a descrip.
3. `admin` - used to document the source of a customer or project that the entry is associated with
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 reference, with the target attribute pointing to the concept ID of another entry.
7. `xref` - used for an external cross-reference, such as a URL, or to point to an external graphic file.

[2] ISO TC37 termbase, available at http://iso.i-term.dk/
Userid and password: TC37 (case sensitive)

8.2 Language level

This level contains elements whose immediate parent is or can be `<langSet>`. They are, in this order:

1. `descrip` - used for a definition that needs to be documented at the language level. This position therefore allows for definitions in different languages.
2. `descripGrp` - used instead of descrip to document a definition and its source. Contains: one descrip and one admin element. If the source of the definition is not required or available, use only a descrip.
3. `transacGrp` - used for administrative information about the language, 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 information
5. `tig` - a nesting element for the term level elements. Each tig contains information about one term.

8.3 Term level

This level contains elements whose immediate parent is or can be `<tig>`. They are, in this order:

1. `term` - contains the term.
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. Do not use this element to record a definition at the term level.
4. `descripGrp` - used instead of descrip to document a context and its source. Contains: one descrip and one admin element. If the source of the context is not required or available, use only a descrip.
5. `admin` - used to document the source of the term, or a customer or project that the term is associated with.
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 or any of the term-related data categories.
8. `ref` - used for an internal reference, with the target attribute pointing to the ID of a term in another entry.
9. `xref` - used for an external cross-reference providing term-related information, such as a URL.

8.4 Backmatter

The backmatter of a TBX-Basic file is significantly simplified compared to TBX-Default. It is used only 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. As with TBX-Default, 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>
   <refObjectList 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>
   </refObjectList>
</back>
```

8.5 Structural differences between TBX-Default and TBX-Basic

The core structure of TBX-Basic is simpler than, yet compliant with, the core structure of TBX-Default. Both are expressed as a DTD file. A customized version of the TBX DTD has been provided for TBX-Basic, for compliance-checking purposes. In the custom DTD, the non-supported elements are commented out and brief comments are added to explain the differences. The key differences are as follows:

- TBX-Default supports both <tig> and <ntig> for term information groups. TBX-Basic supports only <tig>.
- Documenting term components (the individual parts of terms) is not supported in TBX-Basic. Therefore, the following elements are not supported: <termComp>, <termCompList>, <termCompGrp>, and <termGrp>.
- TBX-Basic does not support the following grouping elements and their child elements: <adminGrp>, <termNoteGrp>, <itemSet>, and <itemGrp>. Of the grouping elements, only <descripGrp> and <transacGrp> are allowed.
- 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: <descripNote>, <admin>, <adminGrp>, <note>, <ref>, and <xref>.
- In TBX-Basic, the attribute values DCSName and XCSContent are not supported on the paragraph tag in the <encodingDesc> element.

9 Compliance

9.1 Conditions for compliance

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

- It validates against the TBX-Basic Core Structure DTD and the TBX Core Structure DTD.
- It uses only the data categories that are defined in this document and in the TBX-Basic XCS file, 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.
- Each entry contains at least one language section and at least one **Term**.
- One of the following conditions has been met:
  - If the resource is to be submitted to any form of automated processing, each term level (`<tig>`) 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.

9.2 Compliance-checking TBX-Basic files

The files discussed in this section are available at www.terminorgs.net.

The core structure of a TBX-Default file is defined in the DTD `TBXcoreStructv02.dtd`. It is published in Annex A of the TBX specification (ISO 30042). The core structure of TBX-Basic is slightly simpler than the core structure of TBX-Default, and, therefore, has its own DTD: `TBXBasiccoreStructV02.dtd`. This TBX-Basic DTD is a subset of the TBX-Default DTD and, therefore, fully complies with it.

The set of data categories used in a TML is called a Data Category Selection (DCS). The DCS for TBX-Default and TBX-Basic is defined in an XML file called an Extensible Constraints Specification (XCS). The XCS file for TBX-Default is called `TBXXCSV02.xcs`. TBX-Basic, which comprises a subset of the default DCS, requires its own XCS file, called `TBXBasicXCSV02.xcs`. The XCS files are validated by using the DTD `tbxxcsdtd.dtd`.

As stated earlier in this document, there are three structural levels in a TBX-Default or a TBX-Basic entry, and most data categories are permitted only at certain levels. However, for compliance checking, only the levels of `<descrip...>` elements can be formally constrained by the XCS file provided with TBX-Default. Note also that the levels are not formally constrained by the RelaxNG schema that is referred to in *9.2.2 Using the validation files to check your own TBX-Basic file*.

The XCS formalism constitutes a second layer of constraints on top of the TBX Core Structure DTD. This type of two-layer validation is not a standard XML practice, and off-the-shelf XML validators cannot

read an XCS file. Consequently, you cannot use an off-the-shelf XML validator to check that your TBX-Basic file complies with an XCS file. However, an off-the-shelf XML validator can check that the TBX-Basic file complies with the TBX Core Structure DTD. The XCS file is actually intended as a machine- and human-readable record of the data categories used in a TML, primarily for data interchange purposes.

To validate a TBX-Basic file against the TBX Core Structure DTD *and* the XCS file, you must use the TBX-Checker. The TBX-Checker is a validation tool that is specifically designed for checking files in TBX-compliant formats against both the TBX Core Structure DTD *and* the XCS file. TBX-Checker is available free of charge from SourceForge at http://sourceforge.net/projects/tbxutil.

Alternatively, you can express the constraints that are in the TBX Core Structure DTD and in the XCS file in an integrated schema, such as the RelaxNG schema available from TerminOrgs. Then, you can use any off-the-shelf validator that supports your chosen schema language to validate your TBX-Basic file against all the constraints. If you use the RelaxNG schema, you need to use an XML validator, such as oXygen, that supports the RelaxNG and Schematron languages.

9.2.1 Validation limitations of the TBX-Checker

Currently, TBX-Checker cannot validate the level constraints for the `xref`, `ref`, and `admin` elements because these level constraints are not expressible in the XCS file format. TBX-Checker may be updated in the future to address this limitation.

9.2.2 Using the validation files to check your own TBX-Basic file

Before you can use the files supplied by TerminOrgs to validate your own TBX-Basic file, you need to declare the languages that are used in your TBX-Basic document in those files.

9.2.2.1 If you use TBX-Checker, declare the languages used in the TBX file

Add the languages that are used in your TBX-Basic document to the header of the XCS file, as shown in the following sample code. The sample shows how to document English and French by using a simple two-character code. You can use any language code format that is supported by IETF RFC 4646 or its successor.

```
<languages>
   <langInfo>
      <langCode>en</langCode>
      <langName>English</langName>
   </langInfo>
   <langInfo>
      <langCode>fr</langCode>
      <langName>French</langName>
   </langInfo>
</languages>
```

9.2.2.2 If you use an off-the-shelf XML validator, modify the xml:lang attribute

Modify the definition of the xml:lang attribute to declare which languages are used in the TBX-Basic file. Currently, the definition of the xml:lang attribute indicates that it can contain any kind of text. This section starts on line 415.

```
<define xmlns="http://relaxng.org/ns/structure/1.0"
   name="langSet.localattributes">
      <rng:attribute name="xml:lang">
            <a:documentation
   xmlns:a="http://relaxng.org/ns/compatibility/annotations/1.0">
   Indicates the language of the language section. This attribute is required for
   the langSet element. See also the description on the martif
   element.</a:documentation>
            <text/>
      </rng:attribute>
</define>
```

To declare the languages that are used in your TBX-Basic document, follow the example below. This example shows how to declare English and German using a simple two-character language code. You can use any format that complies with IETF RFC 4646 or its successor.

```
<define xmlns="http://relaxng.org/ns/structure/1.0"
name="langSet.localattributes">
      <rng:attribute name="xml:lang">
            <a:documentation
                  xmlns:a="http://relaxng.org/ns/compatibility/annotations/1.0">
                  Indicates the language of the language section. This attribute
                  is required for the langSet element. See also the description on
                  the martif element.</a:documentation>
            <rng:choice>
                  <rng:value>en</rng:value>
                  <rng:value>de</rng:value>
            </rng:choice>
      </rng:attribute>
</define>
```

10 Recommended standards and guidelines

Since terminological resources can be reused for many purposes, you should follow international standards when designing, developing, and using your termbase.

In addition to ISO 30042 (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 26162 – Design, Implementation and Maintenance of Terminology Management Systems

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

Appendix A – Term location picklist values for Windows user interface objects

The following picklist values are recommended for software user interface locations in a Windows environment.[3] In TBX-Basic, these values should be written in camel case (for example, "menuItem").

- Menu item
- Dialog box
- Group box
- Text box
- Combo box
- Combo box element
- Check box
- Tab
- Push button
- Radio button
- Spin box
- Progress bar
- Slider
- Informative message
- Interactive message
- ToolTip
- Table text
- User defined type

[3] From the Dandelion project led by Klaus-Dirk Schmitz, of Cologne University of Applied Sciences.

Appendix B – Data categories by levels of entry

This appendix lists the TBX-Basic data categories according to the levels of the entry in which they occur, in the recommended order of their occurrence. Strictly speaking, all data categories are optional except the **term** and the specification of the language attribute in its parent language 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**.

In this table, the data category names used are those that appear in *6 Data category descriptions*. These names have been adopted for TBX-Basic for convenience purposes and some may not therefore correspond to the official names in other resources such as ISO 30042 and the ISO TC37 Data Category Registry. Refer to *6 Data category descriptions* for further details such as the unique ISO identifier and the XML representation.

Concept-level data categories

- Subject field
- Definition
- Source of definition
- Customer
- Project
- Created by
- Creation date
- Last modified by
- Last modified date
- Note
- Cross-reference
- External cross-reference
- Figure

Language-level data categories

- Definition
- Source of definition
- Created by

- Creation date
- Last modified by
- Last modified date
- Note

Term-level data categories

- Term
- Term type
- Part of speech
- Gender
- Term location
- Geographical usage
- Usage status
- Context
- Source of context
- Source of term
- Customer
- Project
- Created by
- Creation date
- Last modified by
- Last modified date
- Note
- Cross-reference
- External cross-reference