Terminorgs

Terminology for Large Organizations

TBX-BASIC

Version 4, © 2025

Table of Contents

1. Document Information

Owner and date

TerminOrgs

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

Last updated: November 9, 2025

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

Copyright

Typographical conventions

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

Glossary

data category (DC)

> A type of data field, such as **definition** or **part of speech**. The data categories used in TBX-Basic are fully described in an online data category repository, DatCatInfo (see section 4).

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[1]. TBX-Basic is a dialect.

module

> A list of permissible data categories and constraints on them that are used in the design of a TBX-compliant terminology resource.

terminology resource

> A file, database, or other data collection that contains terms and information about terms.

[1] ISO 30042:2019.

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. The current version is the fourth, published in 2025.

TBX-Basic is compliant with *ISO 30042 (2019): TermBase eXchange (TBX)*, the ISO standard for terminology data.

3. About TBX-Basic

TBX-Basic is a TBX-compliant *dialect*[2] intended to represent and even guide the design of terminology databases (termbases) that have a rather basic structure and a limited number of data categories (DCs), such as those developed by translation companies. It is not intended for the more complex termbases found in large governments, NGOs, and commercial enterprises.

Its purpose is to formalize the XML markup that can be used to represent relatively simple terminology resources, in order to increase structural stability and interoperability.

TBX-Basic is also intended to be a dialect for terminology exchange. It can be used to handle monolingual, bilingual, or multilingual data.

The core structure of all TBX-compliant dialects, including TBX-Basic, is fixed by the TBX-Core schema (see Appendix C). Thus, TBX-Basic inherits the structure and elements that are defined in the TBX-Core schema, such as `<term>` and `<date>`, the structural elements such as `<conceptEntry>`, `<langSec>` and `<termSec>`, inline textual elements such as `<hi>`, and other parts found e.g. in the header and backmatter. The core structure also constrains the order and location of certain elements. To these fundamental parts that are common to all TBX dialects, TBX-Basic specifies additional constraints such as the permissible values of elements, which elements from the core are excluded, and others. The TBX-Basic specific constraints are validated through an additional schema file (see Appendix C).

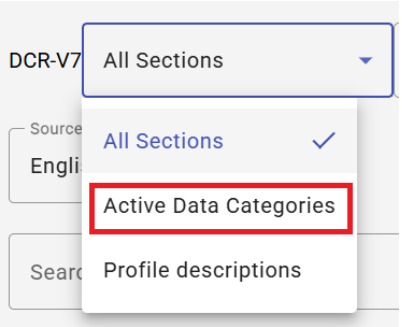
4. DatCatInfo

The DCs used in TBX-Basic are described in this document, but they are also described, some in more detail, in an online data category repository, DatCatInfo, which is available at https://datcatinfo.net.

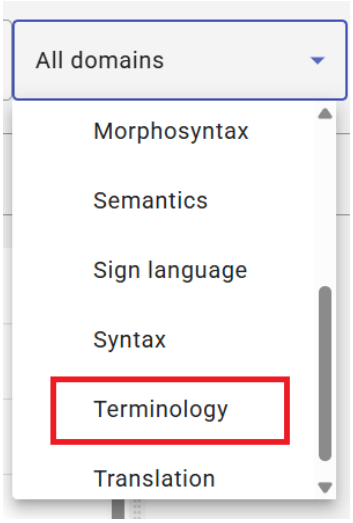
The easiest way to access the TBX-Basic DCs in DatCatInfo is by using the following filter settings.

Under **All Sections**, select **Active Data Categories**:

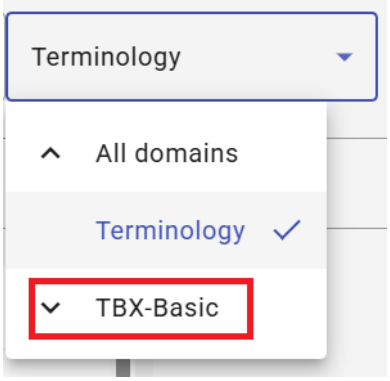
[2] For more information about TBX, dialects, and modules, refer to: www.tbxinfo.net



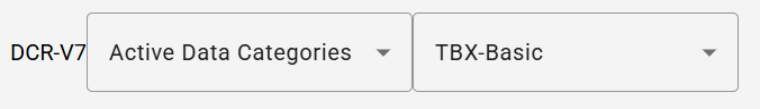
Under **All Domains**, select **Terminology**.



Under **Terminology**, a sub-value will now appear. Select **TBX-Basic**.



Your selections should now look as shown below.



5. Mandatory and recommended data categories

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

The **definition**, **context**, **part of speech**, and **subject field** should be included in a terminology entry whenever possible. The **part of speech** is useful 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 software.
- 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 for further guidelines about the requirements for the part of speech.

6. Data category descriptions

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

Identifier

The identifier of the DC taken from DatCatInfo.

XML markup

The XML representation of the DC, 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>`.

Content type

The type of content that is allowed in the DC. 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 XML elements allowed in basicText and noteText, see section 8.

Section

The section(s) (sometimes referred to as levels) that the DC can appear in. The sections are represented in XML as shown below, 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>`)[3]

[3] In ISO 30042:2019, the elements <termEntry>, <langSet>, and <tig> were renamed to <conceptEntry>, <langSec>, and <termSec> respectively.

6.1 Context

Identifier	DC-0149
XML markup	`<descrip type="context">`
Content type	noteText
Section	term
Comment	A sample sentence that contains the term.

6.2 Cross-reference

Identifier	DC-0164
XML markup	`<ref type="crossReference" target="element_id">`
Content type	plainText
Section	concept, term
Comment	Pointer to another entry, or to a term in another entry, in the same TBX-Basic file or resource.

6.3 Customer

Identifier	DC-0165
XML markup	`<admin type="customerSubset">`
Content type	plainText
Section	concept, term
Comment	Used to identify terms that are required for specific customers.

6.4 Date

Identifier	DC-4335
XML markup	`<date>2008-05-12</date>`
Content type	ISO 8601 date format: yyyy-mm-dd
Section	concept, language, term

Comment	The date must be associated with a specific event, such as creation or modification of data in a concept section, language section, or term section. It can also be associated with the person responsible for the event. Sample markup:
	`<transacGrp>`
	`<transac type="transactionType">modification</transac>`
	`<transacNote type="responsibility">John Harris</transacNote>`
	`<date>2025-04-14</date>`
	`</transacGrp>`

6.5 Definition

Identifier	DC-0168
XML markup	`<descrip type="definition">`
Content type	noteText
Section	concept, language
Comment	Used to describe the meaning of the concept. ISO 704 provides guidelines on creating definitions. To include the source of the definition, the element is nested in a <descripGrp> with an <admin type=”source”>

6.6 External cross-reference

Identifier	DC-0226
XML markup	`<xref type="externalCrossReference" target="external_ID">`
Content type	plainText
Section	concept, term
Comment	Pointer to an external reference for information about the term or concept.

6.7 Figure

Identifier	DC-2920
XML markup	`<xref type="xGraphic" target="file_location">`
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 `<xref...>` element. The `type` value `xGraphic` 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: `<xref type="xGraphic" target="bat.jpg">cricket bat</xref>`

6.8 Gender

Identifier	DC-0245
XML markup	`<termNote type="grammaticalGender">`
Content type	picklist permissible values: • masculine (DC-0246) • feminine (DC-0247) • neuter (DC-0248) • other (DC-0249)
Section	term

6.9 Geographical usage

Identifier	DC-0243
XML markup	`<termNote type="geographicalUsage">`
Content type	plainText
Section	term
Comment	An indicator of a geographical region that the term's usage is restricted to or characteristic of. It is best practice to implement this DC 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 Language

Identifier	DC-0279
XML markup	`<langSec xml:lang="xx-XX">`
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

6.11 Note

Identifier	DC-0382
XML markup	`<note>`
Content type	noteText
Section	concept, language, term
Comment	Any kind of note, such as a usage note, explanation, or instruction.

6.12 Part of speech

Identifier	DC-0396
XML markup	`<termNote type="partOfSpeech">`
Content type	picklist permissible values: • noun (DC-1333) • verb (DC-1424) • adjective (DC-1230) • adverb (DC-1232) • properNoun (DC-0384) • other (DC-4336)
Section	term
Comment	Mandatory under certain conditions. See section 9.

6.13 Project

Identifier	DC-0406
XML markup	`<admin type="projectSubset">`
Content type	plainText
Section	concept, term
Comment	Used to identify terms that are required for or specific to particular jobs or projects.

6.14 Responsibility

Identifier	DC-0451
XML markup	`<transacNote type="responsibility">`
Content type	plainText
Section	concept, language, term

Comment	Used to indicate the name of a person or entity responsible for the creation or modification of a concept section, language section, or term section.
	Must occur inside a `<transacGrp>` element and be accompanied by one `<transac type="...">` element to indicate either creation or modification. May be accompanied by a date to indicate the date of the transaction. For example:
	`<transacGrp>`
	`<transac type="transactionType">modification</transac>`
	`<transacNote type="responsibility">John Harris`
	`</transacNote>`
	`<date>2025-04-14</date>`
	`</transacGrp>`

6.15 Source of context

Identifier	DC-0471
XML markup	`<admin type="source">`
Content type	noteText
Section	term

Comment	It is best practice to document the source of the context.
	This element occurs in a `<descripGrp>` element so that it can be associated with a context, as in the following example:
	`<descripGrp>`
	`<descrip type="context">Sample context.</descrip>`
	`<admin type="source">New York Times</admin>`
	`</descripGrp>`

6.16 Source of definition

Identifier	DC-0471
XML markup	`<admin type="source">`
Content type	noteText
Section	concept, language

Comment	It is best practice to document the source of the definition.
	This element occurs in a `<descripGrp>` element so that it can be associated with a definition, as in the following example:
	`<descripGrp>`
	`<descrip type="definition">Sample definition.</descrip>`
	`<admin type="source">Webster's Dictionary</admin>`
	`</descripGrp>`

6.17 Source of term

Identifier	DC-0471
XML markup	`<admin type="source">`
Content type	noteText
Section	term
Comment	When this element occurs immediately subordinate to the `<termSec>` element, it represents the source of the term. When nested in a `<descripGrp>` it represents the source of the context or definition. See 6.15 and 6.16.

6.18 Subject field

Identifier	DC-0489
XML markup	`<descrip type="subjectField">`
Content type	plainText (see Comment)
Section	concept
Comment	It is best practice to implement this DC as a picklist.

6.19 Term

Identifier	DC-0508
XML markup	`<term>`
Content type	basicText
Section	term
Comment	Mandatory DC; there must be at least one in an entry.

6.20 Term location

Identifier	DC-1823
XML markup	`<termNote type="termLocation">`
Content type	plainText (see Comment)
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 DC as a picklist. A list of picklist values is available in DatCatInfo.

6.21 Term type

Identifier	DC-2677
XML markup	`<termNote type="termType">`
Content type	picklist permissible values: • fullForm (DC-0321) • acronym (DC-0334) • abbreviation (DC-0331 • shortForm (DC-0332) • variant (DC-0330) • phrase (DC-0339)
Section	term
Comment	Refer to DatCatInfo for guidelines on the use of the listed values.

6.22 Transaction type

Identifier	DC-1689
XML markup	`<transac type="transactionType">`
Content type	picklist permissible values: • creation (DC-0303) • modification (DC-0305)
Section	concept, language, term

Comment	Used to indicate the creation or modification of data in a concept section, language section, or term section. Must occur inside a `<transacGrp>` element. Can be optionally accompanied by a `<date>` (see 5.4) and **Responsibility** (see 6.14) to indicate the date of the transaction and the person or entity responsible. Sample markup:
	`<transacGrp>`
	`<transac type="transactionType">modification</transac>`
	`<transacNote type="responsibility">John Harris</transacNote>`
	`<date>2025-04-14</date>`
	`</transacGrp>`

6.23 Usage status

Identifier	DC-0070
XML markup	`<termNote type="usageStatus">`
Content type	picklist permissible values: • preferred (DC-0072) • admitted (DC-0073 • deprecated (DC-0074)
Section	term
Comment	Used for controlled authoring and controlled translation purposes, to mark term usage preferences. Only one value is permitted for each term.

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.

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` element plus either one `transacNote` element, or one `date` element, or both.
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 website, or to point to an external graphic file.
8. `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. admin - used for the source of the definition, nested in the descripGrp.
4. 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.
5. note - any language-level note.
6. 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 constrained by the TBX-Core module. It 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 permissible markup. 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, constrained in TBX-Core, based on function not style: <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> and <ec> 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 <br> 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_schema.xsd` or `core_schema.rng`, and `xml.xsd`), and the TBX-Basic constraints (`basic_schema.sch`). (See Appendix C.)
- It uses only the DCs that are defined in this document as described herein in terms of the nature and type of data.
- The DCs 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 used in any form of automated processing (such as a computer-assisted translation tool or controlled authoring system), 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 (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 - Data categories by section of the entry

This appendix lists the TBX-Basic DCs according to the sections of the entry in which they occur. Strictly speaking, all DCs are optional except the **Term** (at least one term must occur in the entry), and the language attribute in the `<langSec>` tag. However, it is recommended to adhere to the guidelines provided in this document with respect to the use of the **Part of speech**, **Subject field**, **Definition**, and **Context** (for example, section 5 and section 9).

Note that the **Definition** and the **Source of definition** are listed in both the Concept section and the Language section; either is possible, however it is recommended to choose only one level in a given termbase. 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 DCs	**Language section DCs**	**Term section DCs**
Subject field	Definition	Term
Definition	Source (of definition)	Source of term
Source (of definition)	Note	Part of speech
Customer	Transaction type	Gender
Project	Date	Term type
Note	Responsibility	Geographical usage
Cross-reference		Term location
External cross-reference		Usage status
Figure		Context
Transaction type		Source (of context)
Date		Customer
Responsibility		Project
		Note
		Cross-reference
		External cross-reference
		Transaction type
		Date
		Responsibility

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-2">
```

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).

<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>, <adminGrp>, <note>, <ref>, and <xref>.

The following elements are NOT allowed in TBX-Basic:

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

Appendix C – Files needed for validation

Three files are needed to validate a TBX-Basic document instance:

1. A file to validate the core schema
2. A file to validate the language data
3. A file to validate the TBX-Basic specific constraints

The files needed foe (1) and (2) are available from the ISO website: https://standards.iso.org/iso/tr/24633/-2/

- `core_schema.xsd` and `core_schema.rng` : Only one is required; they are technically equivalent and differ only by the schema language. Choose the file that corresponds to your preferred validation method, RelaxNG or XML Schema.
- `xml.xsd` : This file is required for language identification and validation.

The file needed for (3) is available from the TerminOrgs web site: https://www.terminorgs.net/TBX-Basic.html

- `basic_schema.sch`

Appendix D – Sample files

The following sample files are also available from the TerminOrgs web site:

A valid TBX-Basic file:

- `TBX-Basic-sample-valid.tbx`

Files presenting various types of errors:

- `TBX-Basic-sample-badPOS.tbx`
- `TBX-Basic-sample-bad-termtype.tbx`
- `TBX-Basic-sample-crossRef-language-level.tbx`
- `TBX-Basic-sample-def-termlevel.tbx`
- `TBX-Basic-sample-POS-wrong-level.tbx`
- `TBX-Basic-sample-responsibility-not-nested.tbx`
- `TBX-Basic-sample-termtype-language-level.tbx`
- `TBX-Basic-sample-wrong-order-termsec.tbx`