Module Description Formalism

Contents

[Introduction 1](#_Toc499907567)

[Structure 1](#_Toc499907568)

[Elements: 1](#_Toc499907569)

[<contents> 1](#_Toc499907570)

[<datCatSet> 1](#_Toc499907571)

[<[classification]Spec> 1](#_Toc499907572)

[<description> 2](#_Toc499907573)

[<header> 2](#_Toc499907574)

[<level> 2](#_Toc499907575)

[<levels> 2](#_Toc499907576)

[<tbxmd> 2](#_Toc499907577)

[<termNoteSpec> 3](#_Toc499907578)

[<title> 3](#_Toc499907579)

[<value> 3](#_Toc499907580)

[Sample TBXMD 3](#_Toc499907581)

# Introduction

The TBX Module Description Formalism (TBXMD) is an XML format designed for creating a machine-processable description of a TBX data category module. TBXMD is not intended to replace the prose definition of a module as shown in ISO 30042.

# Structure

The root element of an TBXMD file is [<tbxmd>](#_<tbxmd>).

## Elements:

### <contents>

* Description: If the permitted contents of a data category are picklist values, they will be placed here in [<value>](#_<value>) tags. Any other data category must select from one of the 4 permitted text datatypes as described in ISO 30042.
* Attribute:
  + Datatype:
    - Description: The values permitted here are found in ISO 30042.
    - Allowed Values:
      * “PCDATA”
      * “noteText”
      * “basicText”
      * “picklist”
* Children:
  + If @datatype=”picklist”:
    - [<value>](#_<value>)+

### <datCatSet>

* Description: This element contains the collection of permitted data categories in the module.
* Children:
  + One or More:
    - [<[classification]Spec>](#_<[classification]Spec>) and/or [<termNoteSpec>](#_<termNoteSpec>)

### <[classification]Spec>

* [classification] must be replaced with a classification element. The following specs are therefore possible ([<termNoteSpec>](#_<termSecLevel>) **has unique value constraints and is therefore described elsewhere**):
  + <adminSpec>
  + <adminNoteSpec>
  + <descripSpec>
  + <descripNoteSpec>
  + <transacSpec>
  + <transacNoteSpec>
  + <refSpec>
  + <xrefSpec>
* Description: This element is used to provide all necessary information to describe a single data category.
* Attributes:
  + @name
    - This is the name of the data category being included in the MDF
  + @datcatId
    - This is the URI pointing to the source of the data category. For standard data categories found on datcatinfo.net, this is the PID.
* Children:
  + [<levels>](#_<levels>)
  + [<contents>](#_<contents>)

### <description>

* Description: A description of the module being described.
* Text Content

### <header>

* Description: Container for header information.
* Children:
  + [<title>](#_<title>)
  + [<description>](#_<description>)\*

### <level>

* Description: The values permitted are the three levels of TBX.
* Permitted Values (one per [<level>](#_<langSecLevel>)):
  + “conceptEntry”
  + “langSec”
  + “termSec”

### <levels>

* Description: The levels in which the data category may occur.
* Children:
  + One or More:
    - [<level>](#_<level>)

### <tbxmd>

* Description: The root element of an TBXMD file.
* Attribute
  + @module
    - This is the name of the module. (*e.g.*, “fictional”, “basic”, *etc*.)
  + @ns
    - This is the namespace URI of the module (used in TBX DCT style).
* Children:
  + [<header>](#_<header>)
  + [<datCatSet>](#_<datCatSet>)

### <termNoteSpec>

* Description: This element is used to provide all necessary information to describe a single data category. Because termNotes cannot be at any level other than <termSec>, it is not necessary to declare levels.
* Attributes:
  + @name
    - This is the name of the data category being included in the MDF
  + @datcatId
    - This is the URI pointing to the source of the data category. For standard data categories found on datcatinfo.net, this is the PID.
* Children:
  + [<contents>](#_<contents>)

### <title>

* Description: This element is used to provide a more descriptive name for a module, in contrast to the simple name given in the @module attribute of [<tbxmd>](#_<mdf>)
* Text Content

### <value>

* Description: A [<value>](#_<value>) element contains a specific allowed instance of text for a given data category. For example, if the data category is /partOfSpeech/, one of the [<value>](#_<value>) elements in the [<contents>](#_<contents>) section would be: <value>noun</value>

# Sample TBXMD

In this section we will walk through creating a Description for a very simple module with only three data categories: /subjectField/, /partOfSpeech/, and /definition/. We will call this the “Sample” module. We will use the fake namespace identifier: http://www.example.org/ns/sample.

First, we must create the root element (and the XML declaration) with header information:

<?xml version=”1.0” encoding=”UTF-8”?>

<tbxmd module=”Sample” ns=<http://www.example.org/ns/sample>>

<header>

<title>Sample module</title>

</header>

</tbxmd>

Second, we will add the <datCatSet> element. We will be filling it with our data category entries.

<?xml version=”1.0” encoding=”UTF-8”?>

<tbxmd module=”Sample” ns=<http://www.example.org/ns/sample>>

<header>

<title>Sample module</title>

</header>

<datCatSet>

</datCatSet>

</tbxmd>

Finally, we will create entries for each data category using the various <…Spec> elements (descripSpec, adminSpec, etc.). The specific element depends on the classification which is given for each data category. For the data categories we are using in Sample we will use: <descripSpec> (for /subjectField/ and /definition/) and <termNoteSpec> (for /partOfSpeech/).

Each <…Spec> besides <termNoteSpec> expresses the following properties of a data category:

* Data category name (in camelCase)
* Data category ID (this is the PID which points to the data category on datcatinfo.net)
* Levels (conceptEntry, langSec, and/or termSec)
* Contents
  + Can be: PCDATA (plain text), basicText, noteText, or a specific list of permitted values (picklist)

<termNoteSpec> is almost the same, except that levels are not indicated. This is because termNote data categories by definition can only appear at the termSec level.

This is the <descripSpec> for /subjectField/:

<descripSpec name=”subjectField” datcatId=”<http://www.datcatinfo.net/datcat/DC-489>”>

<levels>

<level>conceptEntry</level>

</levels>

<content datatype=”PCDATA”/>

</descripSpec>

The <termNoteSpec> for /partOfSpeech/ looks similar, except with no <levels> and a specific list of values in <content>:

<termNoteSpec name=”partOfSpeech” datcatId=<http://www.datcatinfo.net/datcat/DC-1345>>

<content datatype=”picklist”>

<value>adjective</value>

<value>noun</value>

<value>other</value>

<value>verb</value>

<value>adverb</value>

</content>

</termNoteSpec>

The full TBXMD should look as follows:

<?xml version=”1.0” encoding=”UTF-8”?>

<tbxmd module=”Sample” ns=<http://www.example.org/ns/sample>>

<header>

<title>Sample module</title>

</header>

<datCatSet>

<descripSpec name=”subjectField” datcatId=”<http://www.datcatinfo.net/datcat/DC-489>”>

<levels>

<level>conceptEntry</level>

</levels>

<content datatype=”PCDATA”/>

</descripSpec>

<descripSpec name=”definition” datcatId=”<http://www.datcatinfo.net/datcat/DC-168>”>

<levels>

<level>conceptEntry</level>

<level>langSec</level>

<levels>

<content datatype=”noteText”/>

</descripSpec>

<termNoteSpec name=”partOfSpeech” datcatId=”<http://www.datcatinfo.net/datcat/DC-1345>”>

<content datatype=”picklist”>

<value>adjective</value>

<value>noun</value>

<value>other</value>

<value>verb</value>

<value>adverb</value>

</content>

</termNoteSpec>

</datCatSet>

</tbxmd>