lmt-control-2022-04-12.docx

Language Metadata Table Register

Warning

This document is an unpublished, confidential work under development and shall not be referred to as a SMPTE Standard, Recommended Practice, or Engineering Guideline. It is distributed for review and comment; distribution does not constitute publication. Recipients of this document are strongly encouraged to submit, with their comments, notification of any relevant patent rights of which they are aware and to provide supporting documentation.

Information for document editors

- This template is designed to focus on content and not on style. This means that editors should:
  - Refrain from adding styles inline unless absolutely necessary
  - Refrain from modifying styles unless absolutely necessary
  - Use the sample styles & layouts for all headings, figures, tables and math
  - Use the supplied auto numbering styles for headings and annexes
- In the event that the content cannot be adequately expressed using this template, contact SMPTE HQ.
  - The blue text (like this sentence) is editing information or boilerplate. This is information used by SMPTE HQ for publication and should not be changed by the author.
  - The document editor must fill in the fields in black before submitting it to the Drafting Group.
  - The red text (like this sentence) is helpful information that is not part of the final document. This information **shall be deleted** prior to FCD Ballot.
  - The last page of this document has editing hints and a list of useful reference documents. :::

Project Group: *DG-LMT*

Project Technology Committee: TC-30MR

Document type: *ST*

Document state: *WD*

Project chair(s): *TBD*

Document editor(s): *Bruce Devlin*

Document number: *xxxx-*

Document title: *Language Metadata Table Register*

Title Page

This page will be provided by SMPTE HQ Staff.

See AG-16 clause 3.1 (Title Page), and ISO Directive Part 2 clause 11 (Title).

Copyright © 2022 by THE SOCIETY OF MOTION PICTURE AND TELEVISION ENGINEERS
445 Hamilton Avenue., White Plains, NY 10601
+1 (914) 761-1100

Proposed SMPTE Standard

This document is subject to change

Copyright (C) 202x SMPTE All rights reserved

 © SMPTE 2022 – All Rights Reserved

Foreword

See AG-16 3.2 (Foreword), and ISO Directive Part 2 clause 12 (Foreword).

SMPTE (the Society of Motion Picture and Television Engineers) is an internationally-recognized standards developing organization. Headquartered and incorporated in the United States of America, SMPTE has members in over 80 countries on six continents. SMPTE"s Engineering Documents, including Standards, Recommended Practices, and Engineering Guidelines, are prepared by SMPTE"s Technology Committees. Participation in these Committees is open to all with a bona fide interest in their work. SMPTE cooperates closely with other standards-developing organizations, including ISO, IEC and ITU. SMPTE Engineering Documents are drafted in accordance with the rules given in its Standards Operations Manual. This SMPTE Engineering Document was prepared by Technology Committee TC-30MR.

Normative text is text that describes elements of the design that are indispensable or contains the conformance language keywords: "shall", "should", or "may". Informative text is text that is potentially helpful to the user, but not indispensable, and can be removed, changed, or added editorially without affecting interoperability. Informative text does not contain any conformance keywords.

All text in this document is, by default, normative, except: the Introduction, any section explicitly labeled as "Informative" or individual paragraphs that start with "Note:"

The keywords "shall" and "shall not" indicate requirements strictly to be followed in order to conform to the document and from which no deviation is permitted. The keywords "should" and "should not" indicate that, among several possibilities, one is recommended as particularly suitable, without mentioning or excluding others; or that a certain course of action is preferred but not necessarily required; or that (in the negative form) a certain possibility or course of action is deprecated but not prohibited.

The keywords "may" and "need not" indicate courses of action permissible within the limits of the document.

The keyword "reserved" indicates a provision that is not defined at this time, shall not be used, and may be defined in the future. The keyword "forbidden" indicates "reserved" and in addition indicates that the provision will never be defined in the future.

A conformant implementation according to this document is one that includes all mandatory provisions ("shall") and, if implemented, all recommended provisions ("should") as described. A conformant implementation need not implement optional provisions ("may") and need not implement them as described. Unless otherwise specified, the order of precedence of the types of normative information in this document shall be as follows: Normative prose shall be the authoritative definition; Tables shall be next; then formal languages; then figures; and then any other language forms.

If this is a revision, a topical list of changes [should/shall be included here]

Introduction

An Introduction section is Optional / Conditional

The introduction provides specific information or commentary about the technical content of the document, and about the reasons prompting its preparation. See AG-16

© SMPTE 2022 – All Rights Reserved

clause 3.3 (Introduction), AG-16 clause 4.2 (Conformance Terms), and ISO Directive Part 2 clause 13 (Introduction).

This section is entirely informative and does not form an integral part of this Engineering Document.

The Language Metadata Table (LMT) is an controlled vocabulary that is used to organize language metadata via locations and dialects. It was created to provide a unified, standards based source of reference for language codes for use throughout the media and entertainment industries.

LMT values are selected from the standardsed values in the IETF BCP 47 dictionary. The goal of LMT is to encourage interoperability in code usage by restricting the options to those in practical use within the industry.

Users of the register might maintain versions of the SMPTE register for their own administrative purposes. The format of the SMPTE register is intended for direct use by implmenters.

This document has one element in addition to the printed document.

[Editors notes: The following paragraph will be replaced with the appropriate patent information during the SMPTE Headquarters publication process.]

At the time of publication, no notice had been received by SMPTE claiming patent rights essential to the implementation of this Engineering Document. However, attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. SMPTE shall not be held responsible for identifying any or all such patent rights.

1 Scope

The scope clearly defines the subject of the document and the aspects covered, thereby indicating the limits of applicability of the document. See AG-16 clause 3.4 (Scope), and ISO Directive Part 2 clause 14 (Scope).

This SMPTE Standard defines a method for maintaining and validating a published register of Language Metadata Identifier Records.

2 Normative References

The normative references clause lists, for information, those documents which are cited normatively in the document. See AG-16 clause 3.5 (Normative References), AG-16 clause 4.3 (Normative References to Standards and Recommended Practices), and the ISO Directives Clause 15 (Normative References).

The following SMPTE STANDARD contains provisions that, through reference in this text, constitute provisions of this standard. [Dated references require that the specific edition cited shall be used as the reference]. Undated citations refer to the edition of the referenced document (including any amendments) current at the date of publication of this document. All SMPTE STANDARDs are subject to revision, and users of this engineering document are encouraged to investigate the possibility of applying the most recent edition of any undated reference.

IETF BCP 47 Tags for Identifying Languages, https://tools.ietf.org/html/bcp47

SMPTE Registration Authority, https://smpte-ra.org.

JSON Schema Language draft-json-schema-language-02
https://tools.ietf.org/html/draft-json-schema-language-02

 © SMPTE 2022 – All Rights Reserved

3 Terms and Definitions

The terms and definitions clause provide definitions necessary for the understanding of certain terms used in the document. See AG-16 clause 3.6 (Terms and Definitions), AG-16 clause 4.4 (Terms and Definitions), and ISO Directive Part 2 clause 16 (Terms and Definitions).
Select one of the following sentences and delete the others:

a) For the purposes of this document, the following terms and definitions apply:

b) For the purposes of this document, the terms and definitions given in [external reference(s)] apply.

c) For the purposes of this document, the terms and definitions given in [external reference(s)] and the following apply:

d) No terms and definitions are listed in this document.

If sentence b) or d) is selected, nothing else appears in the Clause.
If sentence a) or c) is selected, one or more terms is defined in the clauses immediately following.
Your terms and definitions follow.

For the purposes of this document, the terms and definitions given in SMPTE ST 2029 and IETF BCP 47 apply.

4 JSON Schema Definitions

Public CD Note

JSON Schema is still in draft form. It is expected that this document will stay in public CD until the IETF standardisation is complete

In order to avoid duplication between text and schema, the cardinality and default values of elements are specified in the schema definitions only.

In the event of a conflict between schema definitions and the prose in this document, the prose shall take precedence

5 The LMT Controlled Vocabulary Register and its Terms

5.1 General (Informative)

The register is represented as an XML Document complying with the XML Schema defined in Annex A. Several views of this data will be presented on the SMPTE registers website along with the XML document. One or more maintainers are responsible for providing updates to SMPTE.

5.2 Schema

The schema for the published data is given in Annex A. The semantics of the schema elements and attributes are defined in section 5.3. The hierarchy and number of occurrences of each element are defined in the schema. Attributes that are required are defined in the schema.

© SMPTE 2022 – All Rights Reserved

5.3 Elements and Attributes in the Register

5.3.1 `term`

The term element defines a term in the register.

5.3.2 `termId`

A unique numerical identifier required for every term.

5.3.3 `termUpdate`

Indicates the status of the term. It shall have the value add or deprecate or delete?????.

5.3.4 `termName`

Shall be a valid IETF BCP-47 langtag value

5.3.5 `termVocabulary`

Shall take one of the enumerated values from the LMTtermVocabularyType in the Schema.

- `Language Metadata Table LMT` shall be used for individual LMT Language Codes
- `Language Groupings LMT` shall be used to define a Language Group.

5.3.6 `termNote`

Each term element shall have the required termNote child elements defined in Table 1. The label attribute contains the name of the termNote field, ex: Language Group Name, Language Group Tag, Language Group Code.

5.3.6.1 `label` attribute

The `label` attribute shall take one of the enumerated values from the `LMTtermNoteLabelType` as defined in the schema. Table 1 `LMTtermNoteLabelType` semantics

1. `Language Group Name`
   - The Group name is equivalent to the generic language name. Language dialects are subordinate to their language grouping.
   - EXAMPLE: Armenian - Western falls under Armenian Family.
   - Required when `termVocabulary` is `Language Groupings LMT`
   - Forbidden when `termVocabulary` is `Language Metadata Table LMT`
2. `Language Group Tag`
   - IETF BCP 47 tag.
   - Required when `termVocabulary` is `Language Groupings LMT`
   - Forbidden when `termVocabulary` is `Language Metadata Table LMT`
3. `Language Group Code`

 © SMPTE 2022 – All Rights Reserved

- URN or URI for the language group term. See [urn encoding of termNote values]
- Required when `termVocabulary` is `Language Groupings LMT`
- Forbidden when termVocabulary is `Language Metadata Table LMT`

4. `Audio Language Tag`
   - IETF BCP 47 language tag. Typically spoken/audio language.
   - Forbidden when `termVocabulary` is `Language Groupings LMT`
   - Required when `termVocabulary` is `Language Metadata Table LMT`
5. `Long Description 1`
   - Description of language name in Latin script following IETF BCP 47 standard
   - Forbidden when `termVocabulary` is `Language Groupings LMT`
   - Required when `termVocabulary` is `Language Metadata Table LMT`
6. `Long Description 2`
   - Alternate description of language name in Latin script following IETF BCP 47 standard
   - Forbidden when `termVocabulary` is `Language Groupings LMT`
   - Optional when `termVocabulary` is `Language Metadata Table LMT`
7. `Audio Language Display Name 1`
   - Endonym of audio language. Typically, the same as Visual Language Display Name 1 but not always.
   - Forbidden when `termVocabulary` is `Language Groupings LMT`
   - Optional when `termVocabulary` is `Language Metadata Table LMT`
8. `Audio Language Display Name 2`
   - Alternate endonym of audio language. Typically, the same as Visual Language Display Name 2 but not always.
   - Forbidden when termVocabulary is Language Groupings LMT
   - Optional when termVocabulary is Language Metadata Table LMT
9. `Visual Language Tag 1`
   - Script in which language is written following IETF BCP 47 standard (which calls for the tags to be presented in Latin Script).
   - Forbidden when termVocabulary is Language Groupings LMT
   - Optional when termVocabulary is Language Metadata Table LMT
10. `Visual Language Tag 2`
    - Alternate script in which language is written following IETF BCP 47 standard (which calls for the tags to be presented in Latin Script).
    - Forbidden when termVocabulary is Language Groupings LMT
    - Optional when termVocabulary is Language Metadata Table LMT
11. `Visual Language Display Name 1`

 © SMPTE 2022 – All Rights Reserved

- Endonym of written language. Typically the same as Audio Language Display Name 1 but not always.
- Forbidden when termVocabulary is Language Groupings LMT
- Optional when termVocabulary is Language Metadata Table LMT

12. `Visual Language Display Name 2`
    - Alternate written endonym. Typically, the same as Audio Language Display Name 1 but not always.
    - Forbidden when termVocabulary is Language Groupings LMT
    - Optional when termVocabulary is Language Metadata Table LMT
13. `Code`
    - URN or URI for each language tag term. See 5.3.6.2
    - Forbidden when termVocabulary is Language Groupings LMT
    - Required when termVocabulary is Language Metadata Table LMT
14. `Scopes`
    - A list of comma separated values to restrict the scope of the usage of the tag. If not specified the scope is global
    - Optional
    - EXAMPLE: “ISDCF, MovieLabs”

5.3.6.2 url encoding of `termNote` element with `label` attribute of `Code`

Each `term` element shall have a `termNote` child element with a `label` attribute of the value `Code`.

The value of the `termNote` child elememt is coded as follows:

```
<CODE> ::= "https://smpte-ra.org/register/lmt/code" ":" <TERM-AUDIO-LANGUAGE-TAG>
```

The `<TERM-AUDIO-LANGUAGE-TAG>` field shall take the value of the `termNote` child element with the `label` attribute of `Audio Language Tag`

EXAMPLE: The termNote XML element:

```
<termNote label="Audio Language Tag">es-419</termNote>
```

Will be encoded as

```
<termNote label="Code">https://smpte-ra.org/register/lmt/code/es-419</termNote>
```

5.3.7 `relation`

5.3.7.1 Presence of a `relation` element

This element indicates that there are one or more related terms to this LMT code value.

If the `termVocabulary` child of a term element is set to Language Metadata Table LMT then the relation will link to a term with its `termVocabulary` child set to Language Grouping LMT.

If the `termVocabulary` child of a term element is set to Language Grouping LMT then there will be on or more relation elements that link to term elements with `termVocabulary` child set to Language Metadata Table LMT.

 © SMPTE 2022 – All Rights Reserved

5.3.7.2 relationType

The schema allows for language relationships through the use of Language Grouping.

There is no requirement a specific term to be part of a Language Group. The use of IETF BCP 47 “Macrolanguage” and “Language Family” designations allow for alphabetical sorting by grouping, keeping languages like Chinese together. If not, languages like Mandarin and Cantonese would separate. A simple hierarchy allows for the maximum flexibility. Some language grouping examples are:

- Greek: to account for ancient vs modern
- English: British, Canadian, Australian, American, etc
- Spanish: Latin American vs European, Mexican vs Argentinian
- Special: for codes such as “und” (undetermined)

The following enumerations of the relationType element are permitted.

Value	Meaning	Notes
EQT	Equivalent to	This term is equivalent to the related term
BT	Broad Term	The related term is a broader term than this term. EXAMPLE: the relation element for en-AU (Australian English) has a BT relation to en (Generic English)
NT	Narrow Term	The related term is a narrower term than this term. EXAMPLE: the relation element for en (Generic English) has a NT relation to en-au (Australian English)
TT	Top Term	The related term is the Top Most Broad Term.

5.3.7.3 termId

Required. Shall be set to the termId value of the related element in the LMT XML document

5.3.7.4 termName

Required. Shall be an identical copy of the termName value of the related term referenced by termId.

5.3.7.5 termVocabulary

Required. Shall be an identical copy of the termVocabulary value of the related term referenced by termId.

6 Register updates

6.1 SMPTE LMT Repository and Submission Package definition

SMPTE manages its copy of the register using a Git repository. The repository contains the files detailed below. An update request to start the SMPTE process is generated by a

 © SMPTE 2022 – All Rights Reserved

maintainer issuing a pull request to SMPTE's `RegLMT` repository at `https://github.com/SMPTE/RegLMT`.

6.2 SMPTE repository structure

The repository shall contain one of each files listed below. All other files in the repository will be ignored and may be deleted at the discretion of SMPTE HQ.

1. `lmt.xml` An XML document of the new LMT version
2. `lmt-ref.xml` An XML document of the current published `lmt.xml`

a. Line endings and indent style shall match the new document

3. `lmt-diff.txt` A human readable difference between the `lmt.xml` and `lmt-ref.xml`. The format of this document is not mandated. It should be appropriate for a SMPTE member to find the differences between the current and previous versions of the register.
4. `lmt-release-summary.md` A markdown narrative to be published with this version
5. `lmt-control.pdf` A PDF of the published version of this document for reference by maintainers.
6. `README.md` A markdown description of the repo referencing the PDF in the repo for contact information.

6.3 QC of the submission package

When the submission package is received, SMPTE will validate:

1. The request was received from one of the approved register maintainers.
2. The `Zthes/LMTMetadata/version` field has been updated
3. The `Zthes/LMTMetadata/DatesISO8601/release` field has been updated
4. The `Zthes/LMTMetadata/DatesISO8601/release` field is absent
5. Line endings shall be unix style (`\n`) and indenting shall be performed with spaces
6. `lmt.xml` validates against the schema defined in the latest version of this document
7. The web pages for the smpte-ra website render correctly
8. The provided documents meet SMPTE process rules

6.4 SMPTE Process (informative)

The submission package shall be treated as an incoming SMPTE Standard and subject to the procedures of the SMPTE Standards Operations Manual. This section is for guidance only. If this section conflicts with the SMPTE Standards OM or any Administrative Guideline, then that other document will prevail.

NOTE: During the Public CD process, the group may choose to ask the TC to ask the SMPTE Steering Committee to put some of the provisions of this section into a new AG.

 © SMPTE 2022 – All Rights Reserved

NOTE: The SMPTE process means that individual submissions might be accepted, rejected or modified by consensus.

1. Once a submission is received and has passed validation, a meeting of the LMT working group is convened to elevate the package as a WD to the TC.
2. The TC chairs initiate a 2-week pre-FCD review to become a CD. [there may be comments]
3. The TC chair asks the SVP to publish the as a public-CD for comment.
4. When the Proponenet wish to progress the document to standard, TC Chairs commence an FCD ballot. With the question *Do the changes to the Language Metadata Table Register fulfil the register requirements in the latest version of SMPTE ST xxxx*
5. If there are comments from the Public CD process or from the Ballot, they shall be addressed.
6. A DP vote follows if there were substantive comments to be addressed in the ballot.
7. An ST Audit takes place to verify process was followed.
8. The LMT is published in the live area of SMPTE-RA and the public CD version is removed.

6.4.1 SMPTE Headquarters (HQ) publishing

The resulting table shall be made available on the SMPTE-RA website at a permanent URL SMPTE may optionally make available other views of the canonical XML available.

Tooling and processing of these alternate views is at the discretion of SMPTE HQ and outside the scope of this document.

7 Register Maintenance

7.1 Official Maintainers

The Automatied update procedure in the section above is available to SMPTE Standards Members who are identified by the controlling Technology Committee as official Maintainers. The list of maintainers shall be made available on the SMPTE Register site.

The role of a maintainer is to represent a group that is active in the maintenance of the controlled vocabulary. A maintainer has the same role as a proponent as defined in the SMPTE Standards Operations Manual.

7.2 Becoming a maintainer

A maintainer shall be a proponent of an approved project to update the register. The project shall state which of the proponents shall be maintainers and whether the proponent is proposing to stay as a a maintainer when the project is complete.

7.3 Removal of maintainers

A maintainer may revoke their status by communicating the desire to do so to the chair of the controlling SMPTE Technology Committee. The controlling SMPTE Technology Committee may remove maintainer status by consensus.

 © SMPTE 2022 – All Rights Reserved

 © SMPTE 2022 – All Rights Reserved

Annex A LMT XML Schema (normative)

This specification is accompanied element-a, which is an XML schema document as specified in XML Schema Part 1: Structures.

The schema is presented informatively in the table below

```
<xs:schema attributeFormDefault="unqualified" elementFormDefault="qualified"
  xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <!-- for restricting the termNote Attributes-->
  <xs:simpleType name="LMTtermNoteLabelType">
    <xs:restriction base="xs:string">
      <xs:enumeration value="Language Group Name" />
      <xs:enumeration value="Language Group Tag" />
      <xs:enumeration value="Language Group Code" />
      <xs:enumeration value="Audio Language Tag" />
      <xs:enumeration value="Long Description 1" />
      <xs:enumeration value="Long Description 2" />
      <xs:enumeration value="Audio Language Display Name 1" />
      <xs:enumeration value="Audio Language Display Name 2" />
      <xs:enumeration value="Visual Language Tag 1" />
      <xs:enumeration value="Visual Language Tag 2" />
      <xs:enumeration value="Visual Language Display Name 1" />
      <xs:enumeration value="Visual Language Display Name 2" />
      <xs:enumeration value="Code" />
      <xs:enumeration value="Notes" />
    </xs:restriction>
  </xs:simpleType>
  <xs:simpleType name="relationTypeType">
    <xs:restriction base="xs:string">
      <xs:enumeration value="EQT" />
      <!-- Equivalent To -->
      <xs:enumeration value="BT" />
      <!-- Broad Type -->
      <xs:enumeration value="NT" />
      <!-- Narrow Type -->
      <xs:enumeration value="TT" />
      <!-- Top Type -->
    </xs:restriction>
  </xs:simpleType>
  <xs:element name="Synaptica-ZThes">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="LMTMetadata" maxOccurs="1" minOccurs="1">
          <xs:complexType>
            <xs:sequence>
              <xs:element type="xs:string" name="version" maxOccurs="1" minOccurs="1"/>
              <xs:element type="xs:string" name="maintainer" maxOccurs="1" minOccurs="1"/>
              <xs:element name="DatesISO8601" maxOccurs="1" minOccurs="1">
                <xs:complexType>
                  <xs:sequence>
                    <xs:element name="release" type="xs:date" maxOccurs="1" minOccurs="1"/>
                    <xs:element name="publish" type="xs:date" maxOccurs="1" minOccurs="0"/>
                  </xs:sequence>
                </xs:complexType>
              </xs:element>
            </xs:sequence>
          </xs:complexType>
        </xs:element>
        <xs:element name="term" maxOccurs="unbounded" minOccurs="0">
          <xs:complexType>
            <xs:sequence>
              <xs:element name="termID" type="xs:int" maxOccurs="1" minOccurs="1"/>
              <!-- <xs:element name="termUpdate" type="xs:string" maxOccurs="1" minOccurs="1"/> -->
              <xs:element name="termName" type="xs:string" maxOccurs="1" minOccurs="1"/>
              <xs:element name="termVocabulary" type="xs:string" maxOccurs="1" minOccurs="1"/>
              <xs:element name="termNote" maxOccurs="unbounded" minOccurs="1">
                <xs:complexType>
                  <xs:simpleContent>
                    <xs:extension base="xs:string">
                      <xs:attribute name="label" type="LMTtermNoteLabelType" use="required"/>
                    </xs:extension>
                  </xs:simpleContent>
                </xs:complexType>
```


© SMPTE 2022 – All Rights Reserved

```
            </xs:element>
            <xs:element name="relation" maxOccurs="unbounded" minOccurs="0">
              <xs:complexType>
                <xs:sequence>
                  <xs:element type="relationTypeType" name="relationType"/>
                  <xs:element type="xs:int" name="relationWeight"/>
                  <xs:element type="xs:int" name="termID"/>
                  <xs:element type="xs:string" name="termName"/>
                  <xs:element type="xs:string" name="termVocabulary"/>
                  <xs:element name="termNote" maxOccurs="unbounded" minOccurs="1" />
                </xs:sequence>
                <xs:attribute type="xs:byte" name="weight" use="optional"/>
              </xs:complexType>
            </xs:element>
          </xs:sequence>
        </xs:complexType>
      </xs:element>
    </xs:sequence>
  </xs:complexType>
 </xs:element>
</xs:schema>
```

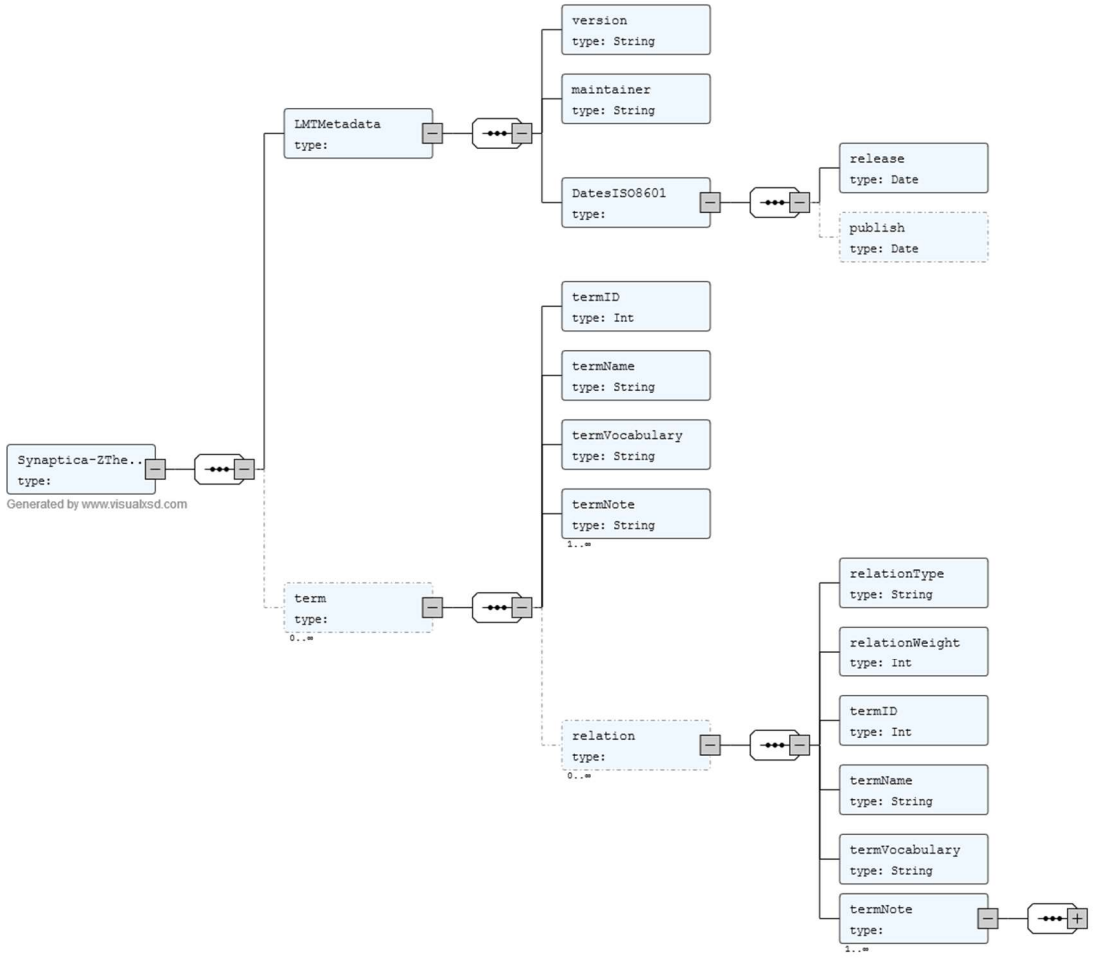

Figure A-1 Pictorial View of LMT Schma (Informative)

© SMPTE 2022 – All Rights Reserved

Information for Document Editors (this page is to be deleted prior to FCD ballot)
The following documents have useful reference material for document editors. SMPTE AG 16:2018 – SMPTE Engineering Document Style Guidelines
International Organization for Standardization (ISO) / International Electrotechnical Commission (IEC), Directives, Part 2:2016-05, Principles and rules for the structure and drafting of ISO and IEC documents, 7.0
Warning: Only copy-and-paste plain text from other documents; otherwise this style set may be corrupted.
The `refdoc` for this document type is usually `xr/refdocs/refdoc-smpte-eng.docx` unless you have used your own. There are samples of all the styles available.
If you extend the document types with your own `refdoc`, the **YOU MUST** make Home Office aware, otherwise the styles may be lost during the Home Office pubishing process.

© SMPTE 2022 – All Rights Reserved