SMPTE ST xxx:202x

DRAFT SMPTE STANDARD

Language Medata Table

# Foreword

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*This document is an unpublished, confidential work under development and shall not be referred to as a SMPTE Standard, Recommended Practice, or Engineering Guideline.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.*

# Introduction

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

The LMT is a controlled vocabulary of codes from the larger 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.

The contributions to the register are delivered to SMPTE by Maintainer organizations. The SMPTE process is used to scrutinize the submissions prior to publication.

This document has one additional element.

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

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

# 2 Normative References

SMPTE ST 2029:202x Uniform Resource Names for SMPTE Resources

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

IETF RFC 5119 SMPTE URN Definition

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

W3C XML Schema Part 1: Structures

W3C XML Schema Part 2: Datatypes

# 3 Terms and Definitions

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

# 4 XML Schema Definitions

This section shall apply whenever a data structure is specified using XML schema definitions as specified in W3C XML Schema Part 1: Structures and W3C XML Schema Part 2: Datatypes.

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, 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.

## 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
   * 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
    * 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 urn encoding of termNote values

IETF RFC 2141 defines the general syntax of URNs as:

<URN> ::= "urn:" <NID> ":" <NSS>

SMPTE URNs use the NID smpte, which was defined by IETF RFC 5119 for registration in the IANA registry of URN NIDs. The first part of the Namespace Specific String for SMPTE Registers is defined in SMPTE ST 2029:

<NSS> ::= "ra" ":" <REGISTER\_IDENTIFIER> “:” <REGISTER\_VALUE>

The register specific identifier for this document shall be

<REGISTER\_IDENTIFIER> ::= "lmt"

The register specific value for this document shall be the value of the termNote XML element with the attribute Audio Language Tag.

<REGISTER\_VALUE> ::= "lmt"

EXAMPLE: The termNote XML element:

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

Will be urn encoded as urn:x-smpte:ra:lmt:es-419

### 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.

#### 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 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
3. Line endings and indent style shall match the new document
4. 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.
5. lmt-release-summary.md A markdown narrative to be published with this version
6. lmt-control.pdf A PDF of the published version of this document for reference by maintainers.
7. 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.

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.

# 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>  
 </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>

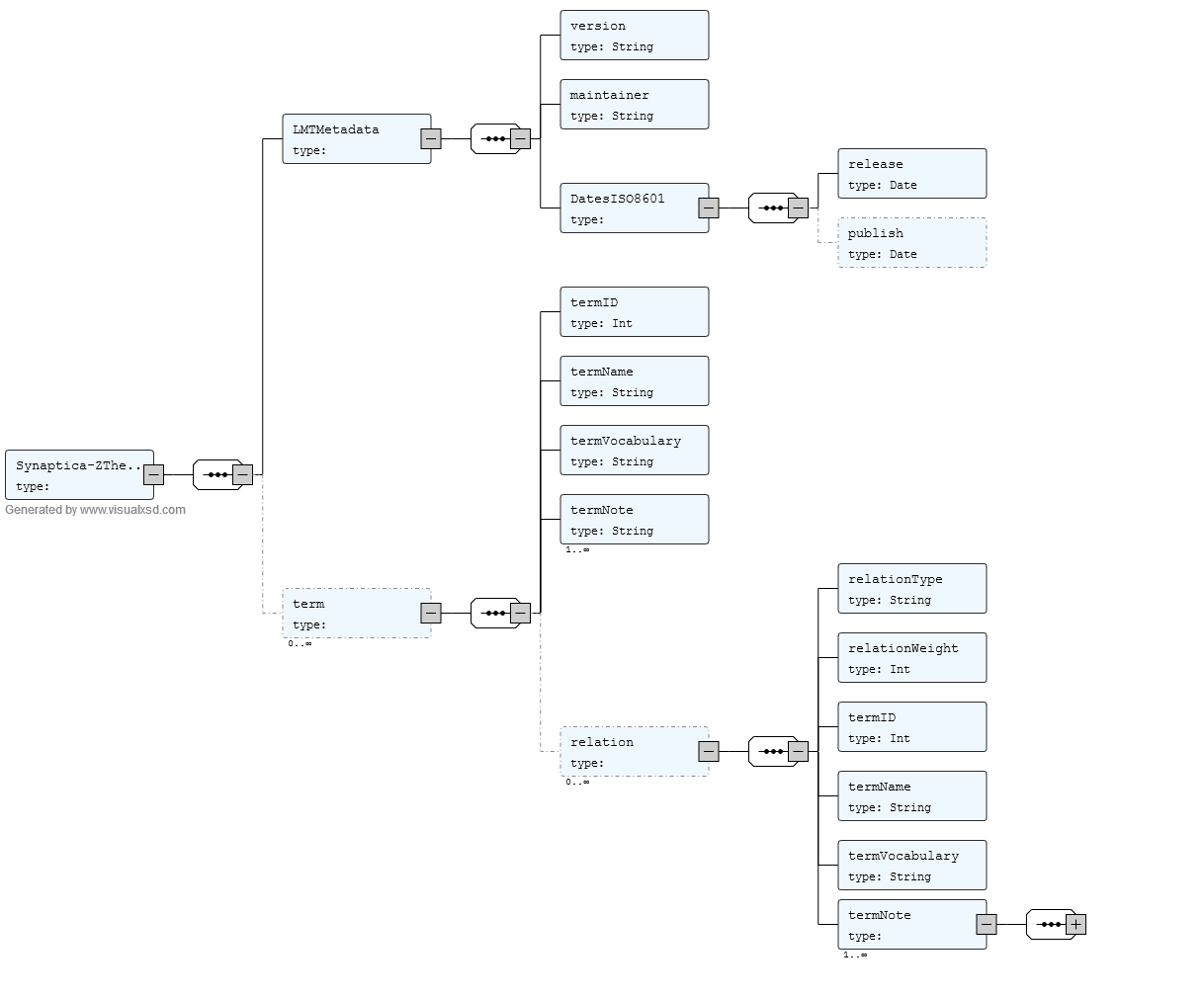


Figure A-1 Pictorial View of LMT Schma (Informative)