

Lmt-contro.pdf

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This document is an unpublished, confidential work under development and shall not be referred to as a SMPTE Standard, Recommended Practice, or Engineering Guideline.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

Proposed SMPTE Standard

This document is subject to change
Copyright (C) 202x SMPTE 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

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 a 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 standardised 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 implementers.

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>

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 the JSON Schema standard and IETF BCP 47 apply. Terms defined in the LMT schema are described in section §5.

4 JSON Schema Definitions

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

The normative schema for the published data is provided as element a of this document. The semantics and normative provisions of the schema are described in the text of this document. Portions of the full schema are included in-line to improve the readability of constraints.

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 JSON Document complying with the schema defined in Annex A. Several views of this data may be presented on the SMPTE registers website along with the canonical representation of the document and schema.

5.2 Organisation of the register (informative)

The register is divided into 3 sections that are present in every published version of the register.

1. lmt - metadata describing a published version of the register
2. terms - an array of language metadata terms

3. groups - an array identifying groups of terms e.g. terms for Serbian languages

5.3 terms

An array of term objects. The order of term objects within the a terms array in an LMT JSON document has no meaning.

5.3.1 term

A collection of key value pairs defining the properties of a unique Language Metadata term element. Note that the object key term does not appear in the JSON because the parent terms object is an array of individual term elements. In JSON the elements of an array do not have keys.

The remaining definitions in this section are children of a term object.

5.3.2 AudioLanguageTag

Required. Shall be a valid IETF BCP-47 langtag value. Every term object in the terms array shall have a unique value of AudioLanguageTag.

5.3.3 Name

Required. Should correspond to the Description field of the IANA SubTag registry entry whose subtag matches the object's AudioLanguageTag property

5.3.4 Code

Required. Shall be a url encoded form of the language identified by AudioLanguageTag that uniquely identifies the term in the register. Used to disambiguate term and group objects as well as to disambiguate GroupTag from other non-SMPTE applications.

The url encoding is described in clause §5.5

NOTE: SMPTE might provide an endpoint for the url to help users of this register

5.3.5 AudioLanguageDisplayName1

Optional. Shall be an endonym for the spoken form of AudioLanguageTag. Used when there are alternate spellings or scripts for writing Name. Typically, the same as VisualLanguageDisplayName1.

5.3.6 AudioLanguageDisplayName2

Optional. Shall be an alternative endonym of AudioLanguageDisplayName1. Used when there alternate spellings or scripts for writing Name.

5.3.7 LongDescription1

Required. Shall be a Description of AudioLanguageDisplayName1 in English following IETF BCP 47 standard

This seems to always be the same as Name - what's it for?

5.3.8 LongDescription2

Optional. Alternate description of language name in Latin script following IETF BCP 47 standard

5.3.9 VisualLanguageTag1

Conditional. Shall be a valid IETF BCP-47 langtag value for a written form of the language identified by AudioLanguageTag. Multiple term objects in the terms array may have the same value of AudioLanguageTag.

A VisualLanguageTag1 value shall be present if a term has a value for VisualLanguageDisplayName1.

5.3.10 VisualLanguageTag2

Optional. Shall be a valid IETF BCP-47 langtag value for an alternate written form of the language identified by AudioLanguageTag. Multiple term objects in the terms array may have the same value of AudioLanguageTag.

5.3.11 VisualLanguageDisplayName1

Conditional. Shall be an endonym of the written form of the language identified by VisualLanguageTag1. Typically the same as AudioLanguageDisplayName1.

A VisualLanguageDisplayName1 value shall be present if a term has a value for VisualLanguageTag1.

5.3.12 VisualLanguageDisplayName2

Optional. Shall be an endonym of the written form of the language identified by VisualLanguageTag1.

5.4 groups

An array of group objects. The order of group objects within the a groups array in an LMT JSON document has no meaning.

NOTE: There is no requirement that a term must 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)

5.4.1 group

A collection of key value pairs defining the properties of a unique Language Metadata group element. Note that the object key group does not appear in the JSON because the parent groups object is an array of individual group elements

The remaining definitions in this section are children of a group object.

5.4.2 GroupTag

Required. Shall be a valid IETF BCP-47 langtag value. Every group object in the groups array shall have a unique value of GroupTag.

NOTE: Some applications need to differentiate a term object that has an AudioLanguageTag value equal to the GroupTag value of a group object. This can be done by inspecting the Code property of the objects. See §5.5.

5.4.3 Name

Required. Should correspond to the Description field of the IANA SubTag registry entry whose subtag matches the object's GroupTag property

5.4.4 Code

Required. Shall be a url encoded form of the language identified by GroupTag that uniquely identifies the group in the register. Used to disambiguate term and group objects as well as to disambiguate GroupTag from other non-SMPTE applications.

The url encoding is described in clause §5.5

NOTE: SMPTE might provide an endpoint for the url to help users of this register

5.4.5 members

Required. Array of member objects that describe term object which are members of the language group described by GroupTag. Note that the array shall be present. The array may be empty.

5.4.5.1 member

A collection of key value pairs defining the properties of a member object. Note that the object key member does not appear in the JSON because the parent members object is an array of individual member elements

The remaining definitions in this section are children of a member object.

5.4.5.2 relationType

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)

Value	Meaning	Notes
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.

5.4.5.3 relationWeight

Required. An integer between 0 and 100 indicating the strength of relationType

Note that all values are always 100 - is this always the case???

5.4.5.4 AudioLanguageTag

Required. Shall be either:

- the AudioLanguageTag of an entry in the terms array e.g. ar-q

or

- a dialect of an AudioLanguageTag of an entry in the terms array e.g. ar-TN

Note that ar-TN does not appear in the terms array

Note that Norwegian is missing AudioLanguageTag values

5.5 url encoding of Code values

Each term object and each group object has a Code property whos value shall be encoded as shown below in the modified ABNF syntax.

In the example below <AudioLanguageTag> is a placeholder for a term's AudioLanguageTag and <GroupTag> is a placeholder for a term's GroupTag.

```

ROOT    = "https://smpte-ra.org/register/lmt/"
TTAG    = "term/code/" <AudioLanguageTag>
GTAG    = "group/code/" <GroupTag>
TAG     = TTAG / GTAG
CODE    = ROOT TAG

```

EXAMPLE: The Code property:

```
"AudioLanguageTag": "es-419"
```

Will be encoded as

```
"Code": "https://smpte-ra.org/register/lmt/term/code/es-419"
```

6 Register updates {#register-updates} (informative procedure)

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> after a project has been initiated.

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.json` An JSON document of the current LMT version
2. `lmt-ref.json` An JSON document of the current published `lmt.JSON`
 - a. Line endings and indent style shall match the new document
3. `lmt.diff` A human readable difference between the `lmt.JSON` and `lmt-ref.JSON`. 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-narrative.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 a SMPTE standards participant.
2. The `lmt` metadata object is correct for a submission
3. Line endings are style (`\n`) and indenting is performed with spaces
4. `lmt.JSON` validates against the schema defined in the latest version of this document
5. The narratives and different files for the `smpte-ra` website render correctly
6. 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 (containing one or more submissions) 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 Home Office (HO) publishing

The resulting table shall be made available on the SMPTE-RA website with a persistent URL SMPTE may optionally make available other views of the standardised JSON available.

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

7 Register Maintenance

7.1 Official Maintainers

The Automated 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 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.

Bibliography

IANA list of BCP 47 subtags <https://www.iana.org/assignments/language-subtag-registry/language-subtag-registry>

Annex A LMT JSON Schema (informative)

Element-a of this document is a JSON schema document. It is presented with line numbers in the table below

```
{
  "$schema": "https://json-schema.org/draft/2020-12/schema",
  "$id": "https://mrmxf.com/schema/lmt-schema-r2.json",
  "title": "LMT Schema",
  "description": "The Schmea for the SMPTE interchange JSON of the Language Metadata Table",
  "required": ["terms", "groups"],
  "type": "object",
  "properties": {

    "terms": {
      "type": "array",
      "minItems": 1,
      "items": {
        "type": "object",
        "$ref": "#/$defs/term"
      }
    },

    "groups": {
      "type": "array",
      "minItems": 1,
      "items": {
        "type": "object",
        "$ref": "#/$defs/group"
      }
    },

    "additionalProperties": true
  },
  "$defs": {
    "term": {
      "type": "object",
      "properties": {
        "Name": {
          "type": "string",
          "minLength": 2,
          "maxLength": 511
        },
        "AudioLanguageTag": {
          "$ref": "#/$defs/AudioLanguageTag"
        },
        "Code": {
          "type": "string",
          "pattern": "https:\\\\smpte-ra\\.org\\/register\\/lmt\\/code\\\[a-zA-Z0-9-\\]{2,5}"
        },
        "audio_language_display_name_1": {
          "type": "string",
          "minLength": 2,
          "maxLength": 511
        },
        "audio_language_display_name_2": {
          "type": "string",
          "minLength": 2,
          "maxLength": 511
        },
        "AudioLanguageDisplayName2": {
          "type": "string",
          "minLength": 2,
          "maxLength": 511
        },
        "LongDescription1": {
          "type": "string",
          "minLength": 2,
          "maxLength": 511
        },
        "long_description_2": {
          "type": "string",

```



```

    },
    "AudioLanguageTag": {
      "type": "string",
      "pattern": "[a-zA-Z0-9-]{2,5}"
    }
  }
}

```

Figure A-1 Highlighted & Numbered View of LMT Schema

Annex B LMT Fragment (informative)

The fragment below shows some fragments of the register's format.

```

{
  "Metadata": {
    "controlDocument": "SMPTE STxxx:2021",
    "registerDate": "2021-05-27",
    "registerStatus": "Experimental",
    "createdBy": "Bruce Devlin"
  },

  "terms": [
    {
      "Name": "Abkhazian",
      "audio_language_display_name_1": "аҧсҭа бызшәа",
      "AudioLanguageTag": "ab",
      "Code": "https://smpte-ra.org/register/lmt/code/ab",
      "LongDescription1": "Abkhazian",
      "VisualLanguageDisplayName1": "аҧсҭа бызшәа",
      "VisualLanguageTag1": "ab"
    },

    ...

    {
      "Name": "Chinese",
      "audio_language_display_name_1": "繁体中文",
      "audio_language_display_name_2": "简体中文",
      "AudioLanguageTag": "zh",
      "Code": "https://smpte-ra.org/register/lmt/code/zh",
      "LongDescription1": "Chinese",
      "VisualLanguageDisplayName1": "简体中文",
      "VisualLanguageDisplayName2": "繁体中文",
      "VisualLanguageTag1": "zh-Hans",
      "VisualLanguageTag2": "zh-Hant"
    },

    ...

  ],
  "groups": [
    ...

    {
      "Name": "Portuguese",
      "Code": "https://smpte-ra.org/register/lmt/code/pt",
      "GroupTag": "pt",
      "members": [
        {
          "relationType": "NT",
          "relationWeight": 100,
          "AudioLanguageTag": "pt-BR"
        },
        {
          "relationType": "NT",
          "relationWeight": 100,
          "AudioLanguageTag": "pt-PT"
        }
      ]
    }
  ]
}

```

```

...
]
}

::: {custom-style="smpte-caption"}

Figure B-1 Fragment of the LMT Register
::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::

```{=openxml}
<w:p><w:r><w:br w:type="page"/></w:r></w:p><!-- insert a page break -->

```

### 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 publishing process.