MESA **LMT**

SMPTE
The home of media professionals, technologists, and engineers

Language Metadata Table
No Longer the Tower of Babel

Yonah Levenson, Co-Chair
Meg Morrissey, Co-Chair
Bruce Devlin, SMPTE
MESA: February 25, 2021

LMT Committee Members & Contributors

Lots of Languages! EVERYWHERE!

*Yet … no single **unified** standard of language codes*
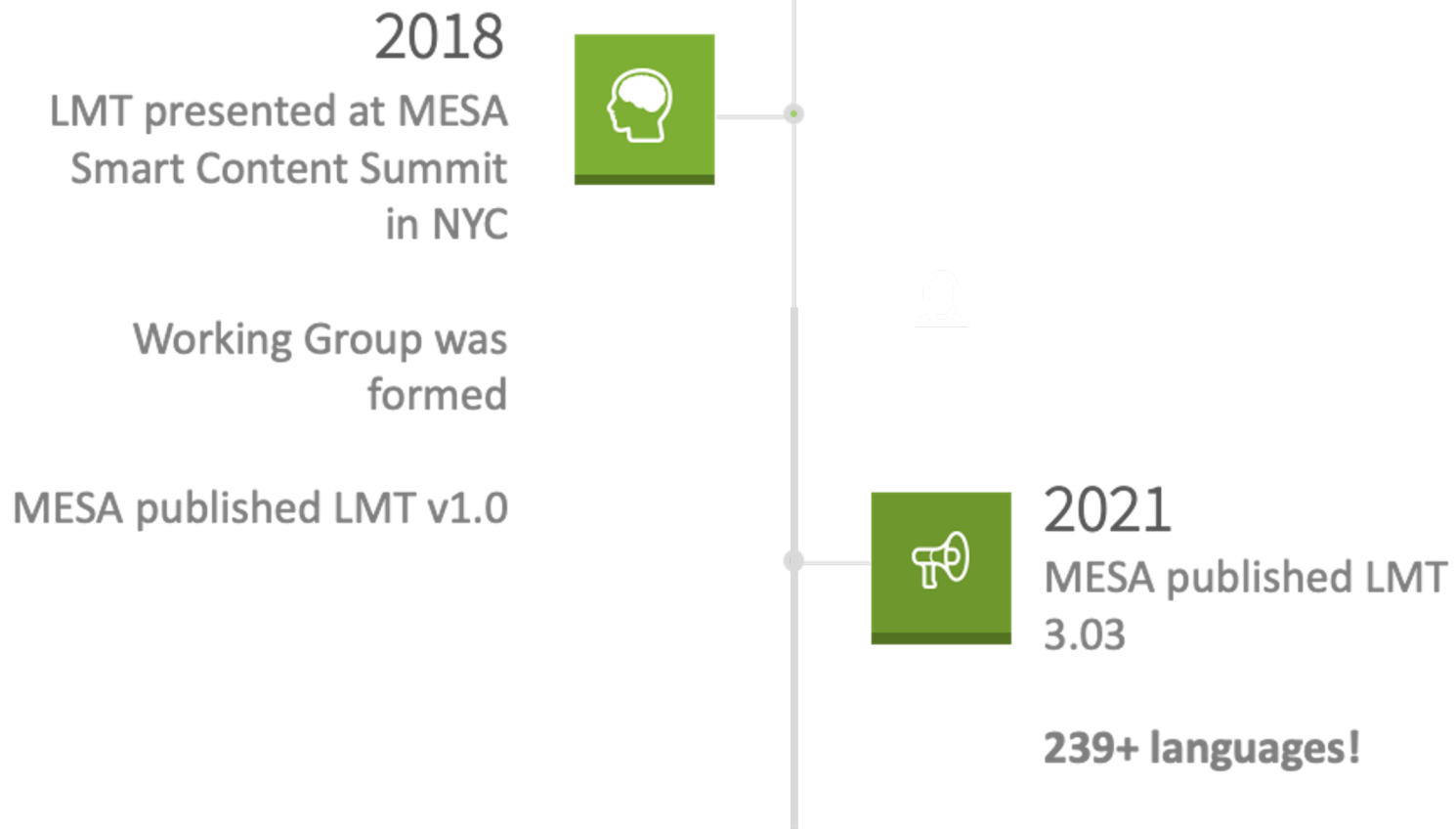
LMT Mission Statement

The Language Metadata Table (LMT) was created to provide a unified source of reference for language codes for use throughout the media and entertainment industries.

1. To create a standardized table of language codes for implementation by entertainment and other industries using IETF BCP 47 (a.k.a., RFC 5646).
2. To facilitate efficient and consistent LMT usage through best practices.
3. To extent LMT code values through vetted field definitions and approved language code values with a community of thought leaders who focus on information and data from the business, professional associations, and academic institutions through the exchange of knowledge and collaboration.

MESA
SMPTE LMT

LMT History

2017
Began at HBO
(now WarnerMedia)

2017
Started with 128
languages

LMT

Where LMT Is Today

2018

LMT presented at MESA Smart Content Summit in NYC

Working Group was formed

MESA published LMT v1.0

2021
MESA published LMT 3.03

239+ languages!

MESA SMPTE LMT

Why Create LMT?

Language Metadata is used by every aspect of media & entertainment, publishing, and other content-centric organizations.

LMT codes for

| Audio and timed text for content | Visual or written languages | Rights and licensing localization | Distribution territories | Accessibility for the visually & hearing impaired (SUDH) |

Advantages of Adopting LMT

The working group: checks & balances across the industry

Standardized distinctions between spoken and written languages

Consistent codes between service providers, clients, and content owners

MESA SMPTE LMT

8
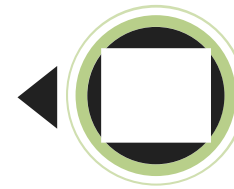
LMT Use Cases

1. Licensing international content
2. Distributing non-English content
3. Accessibility requirements (SDH, sign language)
4. End-user localization preferences
5. Audio
6. Closed Captions
7. Burned in/Forced Narratives
8. Rights
9. Acquisition
10. Distribution
11. Electronic Sell Through

IETF BCP 47

- IETF: Internet Engineering Task Force (a.k.a, the Internet people)
- BCP: Best Current Practice
- BCP 47: Tags for Identifying Language
- IETF BCP 47 defines a standard application of:
  - ISO 639: 2- and 3-character Language codes
  - ISO 3166: 2-character Country codes
  - UN M. 49: 3-digit numeric Territory codes
  - ISO 15924: 4-character Script codes
- IETF BCP 47 works because
  - Language, dialect, script, and geographic codes can be combined in more than 40K ways
    - From the general: en for English
    - To the specific: fr-FR vs. fr-CA to distinguish Parisian French from Quebecoise
  - Codes under regular review to keep the lists current:
    - "Greenlandic" updated to "Kalaallisut" to reflect contemporary cultural norms
    - A WWW standard supported by W3C (a.k.a., the Web people) for HTML, XML, etc.

MESA
SMPTE LMT

Anatomy of a Language Code

- Full code syntax: `language-script-region-variant-extension-privateuse`
  - e.g., `mn-Cyrl-MN` for Mongolian written in Cyrillic as used in Mongolia
- Selecting from 9,000 subtags to create 40,000 combinations can be overwhelming.
- LMT: pre-constructed codes supported by use cases
- Language groupings are explicitly defined – easy enough for Spanish, but hard for Chinese
- For each language, several fields are used to identify the standard:
  - Language Group Name, Tag, Code
  - Audio language tags and displays
  - Visual language tags and displays
  - Descriptions

LMT Examples

Column Header Name	English	Spanish	Serbian	Mandarin	Armenian (Eastern)	Armenian (Western)	American Sign Language
Language Group Name	English	Spanish	Serbo-Croatian	Chinese	Armenian Family	Armenian Family	
Language Group Tag	en	es	sh	zh	hyx	hyx	
Audio Language Tag	en	es-419	sr	cmn	hy	hyw	
Long Description 1	English	Spanish as Spoken in Latin America	Serbian	Mandarin	Armenian	Armenian as spoken by the Armenian Diaspora	American Sign Language
Long Description 2							
Audio Language Display Name 1	English	Español como se habla en América Latina	srpski	普通话	արեւմտահայերէն	հայերեն	
Audio Language Display Name 2			српски				
Visual Language Tag 1	en	es-419	sr-Latn-RS	zh-Hans	hy	hyw	ase
Visual Language Tag 2			sr-Cyrl-RS				
Visual Language Display Name 1		Español como se habla en América Latina	srpski	简体中文	արեւմտահայերէն	հայերեն	American Sign Language
Visual Language Display Name 2			српски				

Column Head Definitions

Column Header Name	Definition
Language Group Name	The name of the language group, if appropriate. The Group name is equivalent to the generic language name. Language dialects are subordinate to their language grouping. Ex: Armenian - Western falls under Armenian Family.
Language Group Tag	IETF BCP 47 tag.
Language Group Code	URN or URI for each language group value in the LMT
Audio Language Tag	IETF BCP 47 language tag. Typically spoken/audio language.
Long Description 1	Description of language name in Latin script following IETF BCP 47 standard
Long Description 2	Alternate description of language name in Latin script following IETF BCP 47 standard
Audio Language Display Name 1	Endonym of audio language. Typically the same as Visual Language Display Name 1 but not always.
Audio Language Display Name 2	Alternate endonym of audio language. Typically the same as Visual Language Display Name 2 but not always.
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).
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).
Visual Language Display Name 1	Endonym of written language. Typically the same as Audio Language Display Name 1 but not always.
Visual Language Display Name 2	Alternate written endonym. Typically the same as Audio Language Display Name 1 but not always.
URN	URN or URI or URL for each language value in the LMT.

MESA SMPTE LMT

LMT Operations

MESA continues to sponsor and promote LMT
SMPTE to be the LMT technical home

MESA

- **ADOPTION GUIDELINES TEMPLATES**
- **RESOURCES**

SMPTE
The home of media professionals, technologists, and engineers

- **TOOLS**
- **INFRASTRUCTURE**

Standards Partners

EIDR: Entertainment ID Registry

ISDCF: InterSociety Digital Cinema Forum

MovieLabs

MESA SMPTE LMT

The home of media professionals, technologists, and engineers

LMT – SMPTE progress Feb 2021

Bruce Devlin
Standards Vice President
svp@smpte.org

SMPTE

The home of media professionals, technologists, and engineers

10E	20F	21DC	24TB	25CSS	30MR	31FS	32NF	34CS	35PM	Misc

Metadata & Registers Technology Committee | **Dean Bullock** | **Phil Warren**

LMT will be in
TC 30 MR

- **TC Scope**
  - The application of the general scope as it applies to definition and implementation of the SMPTE Registration Authority, used to identify digital assets and associated metadata. Additionally, the common definition of metadata semantic meaning across multiple committees..

- **Topics and Standards**
  - Metadata Registers
  - Vocabularies for other TCs
  - https://smpte-ra.org
  - UMIDs

10E	20F	21DC	24TB	25CSS	30MR	31FS	32NF	34CS	35PM	Misc

TC **30MR** Metadata & Registers

doc	slug (unofficial)	type	scope summary	age	update	work done	project review 0%	project approved 10%	drafting 20%	pre-FCD review 30%	CD 40%	FCD ballot 50%	FCD comment re 60%	pre-DP review 70%	DP vote 80%	ST audit 90%	Published 100%
LMT	30MR-LMT	new	LMT Language Metadata Table	-	24 Feb 2021	project scoping & review	2	1									

Notes Project being submitted for a second review to address the fact that the scope should include more details about management of the SMPTE register and the harmonisation with other groups.

Resources & Links

 Email
General inquiries - LMT@mesaonline.org

Documentation & Articles

https://www.mesaonline.org/language-metadata-table (scroll down for current docs)

https://www.mesalliance.org/wp-content/uploads/2018/08/HBO-MESA-LMT-Press-Release-FINAL.pdf

https://www.mesaonline.org/2018/08/08/hbo-looks-to-demystify-language-metadata/

https://www.mesaonline.org/2019/02/20/me-journal-the-language-metadata-table-lmt-an-industrywide-effort-to-collaborate/

https://www.smpte.org/blog/in-collaboration-with-mesa-smpte-brings-first-of-a-kind-language-metadata-table-register-into-its-agile-new-public-cd-process

MESA
SMPTE LMT

Validating Sources

IETF BCP 47 says to use the **shortest codes possible,**
if appropriate.

Validation sources include:

- Validator for checking codes: https://r12a.github.io/app-subtags/

- ISO 639-3 Registration Authority: https://iso639-3.sil.org/

- IANA Language Subtag Registry
  https://www.iana.org/assignments/language-subtag-registry/language-subtag-registry

- Google Language Table

MESA SMPTE LMT

QUESTIONS

Thank You!