

# Oh, The Places Your Metadata Will Go!

A State of the State of LMT  
Smart Content Summit, March 16, 2021



# Agenda

- LMT Background
- Working Group
- Where LMT is Today
- SMPTE Partnership
- Use Case: When ViacomCBS Met LMT
- Q & A



# LANGUAGE METADATA TABLE

## No Longer the Tower of Babel

Yonah Levenson, Co- Chair  
Meg Morrissey, Co-Chair  
Smart Content Summit | March 16, 2021

# LMT Timeline

2017

- Began at HBO (now WarnerMedia)
- Started with 128 Languages

2018

- LMT presented at MESA's Smart Content Summit
- Formed Working Group
- LMT V1.0

2019-2020

- Added languages
- Updated Versions
- Nominated for Best Practice: Taxonomy Bootcamp London
- Partnership with SMPTE initiated

2021

- LMT 3.03
- 239 Languages

...and more languages coming!

# 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.*

- To create a standardized table of language codes for implementation by entertainment and other industries using IETF BCP 47 (a.k.a., RFC 5646).
- To facilitate efficient and consistent LMT usage through best practices.
- To extend 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.

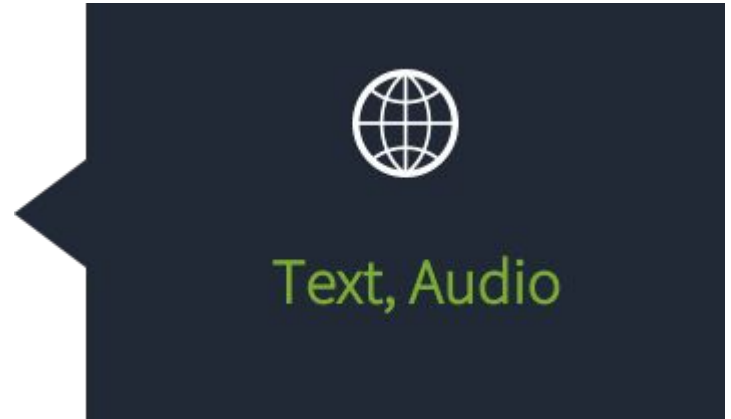
MESA SMPTE **LMT** Committee Members & Contributors



# LMT Scope

Populate asset language elements (text, audio)

- Languages only
- Audio
- Visual: Notation of script/writing system
- Endonyms: Language name in the country's language (Français)



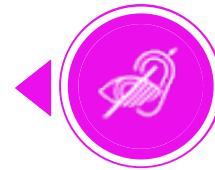
# Advantages of Adopting LMT

The working group: checks & balances across the industry

Standardized distinctions between spoken and written languages



Consistent codes between service providers, clients, and content owners





# LMT Use Cases Include:



- Audio
- Visual (Written/Text)
  - Subtitles
  - Closed Captions
  - Burned in/Forced Narratives
  - Accessibility (SDHH, Sign Language)
- Content licensing
- Rights
- Acquisition
- Distributing content
  - One or more languages
  - Electronic Sell Through (EST)
- Localization preferences

# IETF BCP 47

Full code syntax:

language-script-region-variant-extension-privateuse

Standard	Pattern
ISO 639	2 and 3 character Language Codes
ISO 15924	4 character Script Codes
ISO 3166	2 character Country Codes
UN M.49	3 digit numeric Territory Codes

# IETF BCP 47

Full code syntax: **language**-script-region-variant-extension-privateuse

Kirghiz

Visual: **ky**-Cyril

Audio: **ky**

Spanish as spoken in Latin  
America

Visual: **es-419**

Audio: **es-419**

Standard	Pattern
<b>ISO 639</b>	2 and 3 character Language Codes
ISO 15924	4 character Script Codes
ISO 3166	2 character Country Codes
UN M.49	3 digit numeric Territory Codes

English as spoken in the US

Visual: **en-US**

Audio: **en-US**

# IETF BCP 47

Full code syntax: language-script-region-variant-extension-privateuse

Kirghiz

Visual: ky-Cyrl

Audio: ky

Spanish as spoken in Latin  
America

Visual: es-419

Audio: es-419

Standard	Pattern
ISO 639	2 and 3 character Language Codes
ISO 15924	4 character Script Codes
ISO 3166	2 character Country Codes
UN M.49	3 digit numeric Territory Codes

English as spoken in the US

Visual: en-US

Audio: en-US

# IETF BCP 47

Full code syntax: language-script-region-variant-extension-privateuse

Kirghiz

Visual: ky-Cyrl

Audio: ky

Spanish as spoken in Latin  
America

Visual: es-419

Audio: es-419

Standard	Pattern
ISO 639	2 and 3 character Language Codes
ISO 15924	4 character Script Codes
ISO 3166	2 character Country Codes
UN M.49	3 digit numeric Territory Codes

English as spoken in the US

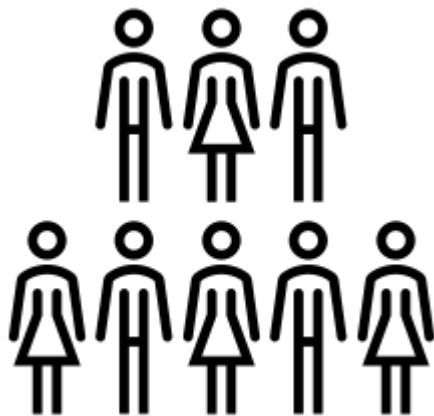
Visual: en-US

Audio: en-US

# 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	հայերէն	
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	հայերէն	American Sign Language
Visual Language Display Name 2			српски				14

# LMT Working Group



- Requests for 150+ additional languages!
  - Sources include: WarnerMedia, ViacomCBS, Disney
  - MESA resource is working through the requests and suggesting codes
    - Language Groups
    - Individual Languages
- Chinese
  - Additional Chinese dialects (Audio) have been submitted for review and approval
  - Request to separate Visual and Audio languages/dialects is under consideration
- Languages of the Indian Subcontinent
  - Qube Cinema is working on the Language Group Codes and Language Codes
  - Rolling submissions for approval

# LMT Sponsorship/Support & Technical Home

MESA

- ADOPTION GUIDELINES  
TEMPLATES
- RESOURCES



- TOOLS
- INFRASTRUCTURE

## Standards Partners Include:

EIDR (Entertainment ID Registry)

ISDCF (InterSociety Digital Cinema  
Forum)

MovieLabs





*The home of media professionals, technologists, and engineers*

# LMT – SMPTE progress Feb 2021



Bruce Devlin  
Standards Vice President  
[svp@smpte.org](mailto:svp@smpte.org)

LMT will be in  
Tech Committee 30  
Metadata Registry



## • 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://smp-te-ra.org>
- UMIDs

# TC 30MR

## Metadata & Registers

doc	slug (unofficial)	type	scope	summary	age	update	work done	0% project review	10% project approved	30% pre-FCO review	40% CD	50% FCO ballot	60% FCO comment review	70% pre-DP review	80% DP vote	90% ST audit	100% Published
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.

# Any questions?



## LMT

Meg Morrissey, Co-Chair

Yonah Levenson, Co-Chair

[LMT@mesaonline.org](mailto:LMT@mesaonline.org)

## ViacomCBS

Eric Emeric, [eric.emic@showtime.net](mailto:eric.emic@showtime.net)

Sarah Nix, [sarah.nix@viacom.com](mailto:sarah.nix@viacom.com)

## SMPTE

Bruce Devlin

[svp@smpte.org](mailto:svp@smpte.org)

# Appendix

# Resources and Links

- General Inquiries: [LMT@mesaonline.org](mailto:LMT@mesaonline.org)
  - Yonah Levenson (WarnerMedia) and Meg Morrissey (Netflix): co-chairs
- LMT Codes in Workbook and XML formats, plus documentation and templates:  
<https://www.mesaonline.org/language-metadata-table>
- SMPTE Validation Tool Demo  
<https://drive.google.com/file/d/164EiH9DwvR7pDusKp7hVv5PLW4d7aYk7/view?usp=drivesdk>

## Primary LMT Code Validation Sources

- 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>
- And others

# 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.