The LMT (Language Metadata Table) Helps The World Manage Media Assets

YONAH LEVENSON, LMT CHAIR

MANAGER, METADATA STRATEGY

WARNERMEDIA

WARNER MEDIA

MESA Media & Entertainment Services Alliance


WarnerMedia

LMT Committee Members & Contributors

WARNER MEDIA

NBCUniversal

HBO

ISDCF

SHOWTIME

FOX

turner

SONY

amazon

Haymillian

Lots of Languages!

Everywhere!

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 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.

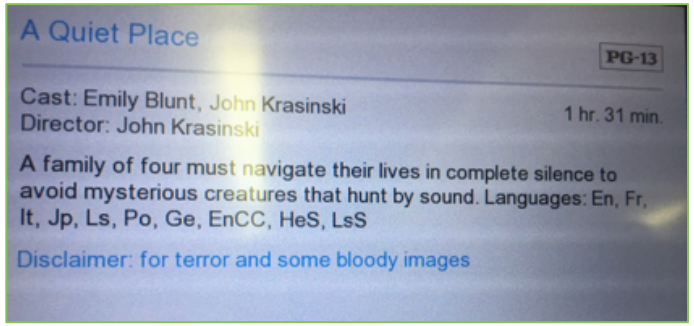
LMT Scope

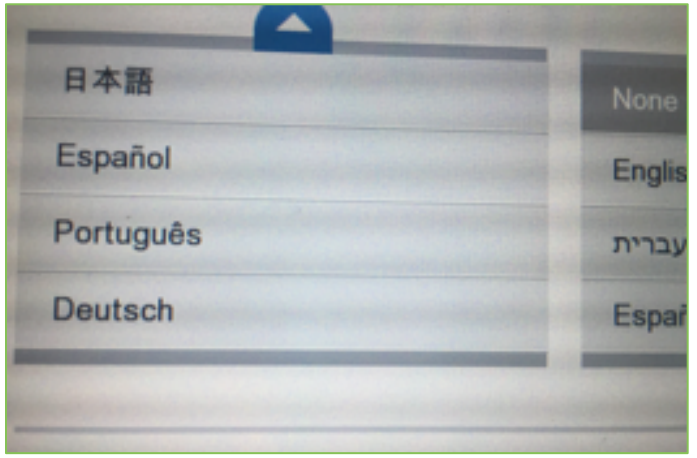
Populate asset language elements (text, audio)

- ✓ **Languages only**
- ✓ **Notation of script/writing system included**
- ✓ **Endonyms: Language name in the country's language (Français)**
- ✓ **Exonymns: Language name as spoken in other countries (Französisch)**

Text, Audio

Consumer Facing Language Display

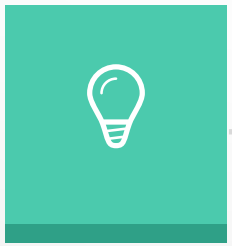




LMT History

2017

Began at HBO
(now WarnerMedia)



2017

Started with 128
languages

2018

LMT presented at MESAlliance Summit in NYC

Working Group was formed

MESAlliance published LMT v1.0

2020

MESAlliance published LMT 3.0

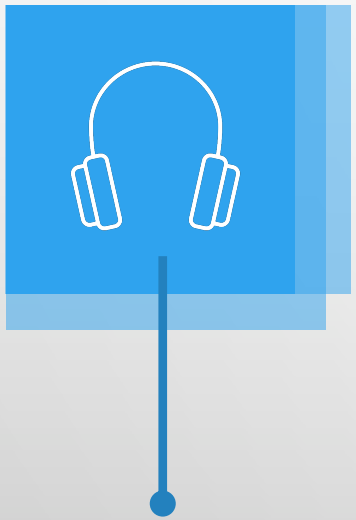
200+ languages!

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

LMT Use Cases & Examples

1. Licensing international content
2. Distributing non-English content
3. Accessibility requirements
4. End-user localization preferences

- ✓ Audio
- ✓ Closed captions
- ✓ Burned in or forced narratives
- ✓ Accessibility (sign language, SDH)
- ✓ Acquisition/rights
- ✓ Electronic sell-through partners

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.

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				

LMT for Publishing and More

METADATA

- LANGUAGE OF THE AUDIO CONTENT
  - AUDIO BOOK
  - CLIPS, SUCH AS PRONUNCIATION GUIDE
  - ACCESSIBILITY
- LMT ALSO WORKS FOR:
  - MANUFACTURING/PACKAGING
  - TECH DOCS
  - ETC.
- LMT CONTAINS THE NEEDED CODES AND DESCRIPTIVE TEXT FOR **200+ LANGUAGES**

New Home and Partners

MESA continues to sponsor and promote LMT

SMPTE to become LMT technical home

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

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

Standards Partners

EIDR: Entertainment ID Registry

ISDCF: InterSociety Digital Cinema Forum

MovieLabs

MESAlliance & SMPTE: Partnership Status

Create a standardized Register of the LMT codes

- Adhere to SMPTE's Public Committee Draft (PCD) mechanism – field test the process prior to full standardization
- LMT: first standard for the new Register process
- July 2020
  - start work on the Register
  - PCD expected a couple of months later

- Prior to standardized Register approval:
  - URN form for testing: *urn:example:smpte:ra:lmt:xxx*
  - Register will be available to test for humans & machines at the SMPTE-RA website
  - Test how multiple maintainers add terms without conflict
  - PCD process: 6 months – 3 years

Resources & Links

Email

General inquiries - LMT@mesalliance.org

Updates & edition requests - LMT@mesalliance.org

Direct contact with the Co-chairs - LMTChairs@mesalliance.org

Documentation

https://www.mesalliance.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.mesalliance.org/2018/08/07/mesa-publishes-hbo-developed-me-industry-language-metadata-table/
https://www.mesalliance.org/2018/08/08/hbo-looks-to-demystify-language-metadata/
https://www.mesalliance.org/2019/02/20/me-journal-the-language-metadata-table-lmt-an-industrywide-effort-to-collaborate/

Validating Sources

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

LMT code 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