# The Language Metadata Table (LMT)

**MESAlliance Smart Content Summit, March 4, 2020**

Yonah Levenson, LMT Chair
LMT Committee Members & Contributors include:

WARNER MEDIA
NBCUniversal
ISDCF
HBO
LIONSGATE
SHOWTIME
Paramount
TURNER
SMPTE
movie labs
SONY
Gracenote
FOX
Discovery Channel
Entertainment Identifier Registry
Hasbro
Disney
Haymillian
Lots of Languages! Everywhere!

No single unified standard of language codes
LMT History

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

2018
- LMT was presented at a MESAlliance Summit in NYC
- Working group was formed
- MESAlliance published LMT v1.0

2020
- MESAlliance published LMT LMT 3.0: 200+ languages!
Ex: Consumer Facing Language Display
LMT Scope

- Populate asset language elements, e.g. text, audio
- Languages only
- Notation of script/writing system included
- Includes:
  - Endonyms: Language name in the country’s language. Ex: Français
  - Exonyms: Language name as spoken in other countries. Ex: Französisch
- Codes for:
  - Audio and timed text for content
  - Visual or written languages: Closed Captions, Subtitles
  - Rights and Licensing localization
  - Distribution territories including Electronic sell through
  - Accessibility for the visually and hearing impaired (SDH)
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. LMT’s mission is:

● 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.
Advantages to Adopting LMT

- Standardized distinctions between spoken and written languages
- The working group: checks & balances across the industry
- Consistent codes between service providers, clients, and content owners
LMT Use Cases

- Licensing international content
- Distributing non-English content
- Accessibility requirements
- End-user localization preferences
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
<table>
<thead>
<tr>
<th>Column Header Name</th>
<th>Example 1: English</th>
<th>Example 2: Spanish</th>
<th>Example 3: Serbian</th>
<th>Example 4: Mandarin</th>
<th>Example 5: Armenian (Eastern)</th>
<th>Example 6: Armenian (Western)</th>
<th>Example 7: American Sign Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language Group Name</td>
<td>English</td>
<td>Spanish</td>
<td>Serbo-Croatian</td>
<td>Chinese</td>
<td>Armenian Family</td>
<td>Armenian Family</td>
<td></td>
</tr>
<tr>
<td>Language Group Tag</td>
<td>en</td>
<td>es</td>
<td>sh</td>
<td>zh</td>
<td>hyx</td>
<td>hyx</td>
<td></td>
</tr>
<tr>
<td>Audio Language Tag</td>
<td>en</td>
<td>es-419</td>
<td>sr</td>
<td>cmn</td>
<td>hy</td>
<td>hyw</td>
<td></td>
</tr>
<tr>
<td>Long Description 1</td>
<td>English</td>
<td>Spanish as Spoken in Latin America</td>
<td>Serbian</td>
<td>Mandarin</td>
<td>Armenian</td>
<td>Armenian as spoken by the Armenian Diaspora</td>
<td>American Sign Language</td>
</tr>
<tr>
<td>Long Description 2</td>
<td>English</td>
<td>Español como se habla en América Latina</td>
<td>srpski</td>
<td>zh-Hans</td>
<td>hy</td>
<td>hyw</td>
<td>ase</td>
</tr>
<tr>
<td>Audio Language Display Name 1</td>
<td>English</td>
<td>Español como se habla en América Latina</td>
<td>srpski</td>
<td>zh-Hans</td>
<td>hy</td>
<td>hyw</td>
<td>ase</td>
</tr>
<tr>
<td>Audio Language Display Name 2</td>
<td>srpski</td>
<td>sr-Latn-RS</td>
<td>zh-Hans</td>
<td>hy</td>
<td>hyw</td>
<td>ase</td>
<td></td>
</tr>
<tr>
<td>Visual Language Tag 1</td>
<td>en</td>
<td>es-419</td>
<td>sr-Latn-RS</td>
<td>zh-Hans</td>
<td>hy</td>
<td>hyw</td>
<td>ase</td>
</tr>
<tr>
<td>Visual Language Tag 2</td>
<td>sr</td>
<td>sr-Cyrl-RS</td>
<td>zh-Hans</td>
<td>hy</td>
<td>hyw</td>
<td>ase</td>
<td></td>
</tr>
<tr>
<td>Visual Language Display Name 1</td>
<td>Español como se habla en América Latina</td>
<td>srpski</td>
<td>zh-Hans</td>
<td>hy</td>
<td>hyw</td>
<td>ase</td>
<td></td>
</tr>
<tr>
<td>Visual Language Display Name 2</td>
<td>srpski</td>
<td>srpski</td>
<td>zh-Hans</td>
<td>hy</td>
<td>hyw</td>
<td>ase</td>
<td></td>
</tr>
</tbody>
</table>
Supportive Use Cases & Tools

- **EIDR: Entertainment ID Registry**
  - Captures Language Now: Original Language, Version Language
  - Already adheres to IETF BCP 47
  - IETF RFC5646 compliant
  - Excellent data source for the analysis of languages applied to content

- **ISDCF: InterSociety Digital Cinema Forum**
  - Language codes for the subtitles, captions, closed captions for content delivered to theatres
  - Gap analysis underway for a number of languages

- **Code validators include:**
  - Validator for checking codes: [https://r12a.github.io/app-subtags/](https://r12a.github.io/app-subtags/)
  - ISO 639-3 Registration Authority: [https://iso639-3.sil.org/](https://iso639-3.sil.org/)
New Home and Partners!

Home Sweet Home!

- MESA will continue to sponsor and promote LMT
- Partnership with SMPTE as LMT Home
  - Infrastructure
  - Resources for hosting, posting, and notifying participants and users
  - Tools
  - adoption guidelines/templates

Standards Partners!

- [EIDR](#): Entertainment ID Registry
- [ISDCF](#): InterSociety Digital Cinema Forum
- [MovieLabs](#)
Moving Forward

- **Goal:**
  - Common Language Code Register
  - Harmonize differences
  - Document

- **Partners:** MESAAlliance, SMPTE, EIDR, ISDCF, & MovieLabs
  - MESAAlliance continued to lead the way
  - Working Group continues

- **2020 Timeline**
  - mid-March: Release LMT 3.1 (audio codes and minor corrections)
  - Ongoing: Meetings with Partners to work out processes and details
  - Working group meeting to be scheduled
  - mid-April: Target for release of new processes and updates at NAB
LMT Contact Information + Links

Email Addresses:
- LMT@mesalliance.org  For General inquiries
- LMTWG@mesalliance.org  For update and edition requests
- LMTChairs@mesalliance.org  For direct contact with the Co-chairs
- Demo of the SMPTE site & source code (desktop/tablet only for now): https://mrmxf.com/project/proto-lmt

LMT Documentation Links
- https://www.mesalliance.org/language-metadata-table  (scroll down for current docs)
- https://www.mesalliance.org/2018/08/07/mesa-publishes-hbo-developed-me-industry-language-metadata-table/
Appendix
Deliverables

- Excel, PDF: Currently being generated out of a taxonomy tool; update pending
- Schema, XML: XML is generated from a taxonomy tool; update info to be automated
- Policy doc: Follow SMPTE’s RDD -- (Standard) Registration Definitions Document
- Validation process
- SMPTE landing page
- Distribution lists: Project Group, Working Group, General Distribution/Subscription
Project Plan Details

- Project Plan framework defined
- Business Case defined
- Project Plan details underway
- MESA & SMPTE agreements underway
- Documentation
  - RDD in development
  - Create Operations manual
  - URN patterns and updates are in development
- Execution:
  - Update LMT workbook and schema, then test
  - Deploy final assets
- New submissions
  - Template development underway
  - Includes documentation and form
- Distribution
  - List creation underway
  - Validation tool development underway
  - Document library to be created