LMT Committee Members & Contributors include:

WARNER MEDIA  NBCUniversal  LIONSGATE
SHOWTIME  HBO  Paramount
movie labs  WARNER BROS. PICTURES  FOX
gracenote  Hasbro  Turner
SONY  Amazon  Discovery Channel
EIDR  Entertainment Identifier Registry  SMPTE
Disney  Haymillian
The broadcast and media industry, as well as others, does not have a single unified standard of language terminology.

**RESULT:**

Each point of distribution/information exchange creates their own language code table, thus creating a kind of Tower of Babel.
Ex: Consumer Facing Language Display

**A Quiet Place**

**Cast:** Emily Blunt, John Krasinski  
**Director:** John Krasinski

1 hr. 31 min.

A family of four must navigate their lives in complete silence to avoid mysterious creatures that hunt by sound. Languages: En, Fr, It, Jp, Ls, Po, Ge, EnCC, HeS, LsS

Disclaimer: for terror and some bloody images
Why Create LMT?

- Language metadata is used by every aspect of media & entertainment, but is rarely standardized between systems or across supply chains.
- There was a need for a more flexible, granular, and modular language standard, applicable in both broad and narrow contexts.
- People are often confused or overwhelmed by all the options available.
- LMT includes values for several language applications in the media and entertainment industries, including codes for:
  - Audio and timed text for content
  - Visual or written languages for display on storefronts and packaging
  - Rights and Licensing localization
  - Distribution territories
  - Accessibility for the visually and hearing impaired
LMT Scope

- Used to populate the language elements of an asset, e.g. text, audio
- Languages only; regions and territories are a separate discussion
- Notation of script/writing system where necessary
- Includes
  - Endonyms: Language name in the country’s language. Ex: Français
  - Exonyms: Language name as spoken in other countries. Ex: Französisch
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.
LMT History

- The Language Metadata Table (LMT) initiative began at HBO in 2017 to normalize language codes and provide a unified source of reference throughout the enterprise.
- The initial table had 128 languages, each with a production use case.
- In July 2018, MESAlliance invited HBO to share their language tables as the basis of an industry standard. A working group was formed with representatives from studios, post houses, and other media affiliates and companies.
- In August 2018, LMT v1.0 was published with 128 languages.
- In September 2019, LMT v2.1 was published with over 200 languages, including best practices.
Advantages to Adopting LMT

- Allows standard distinctions between spoken and written languages
- Provides flexibility for capturing language metadata for various departments
- Having a working group to manage the LMT can better account and coordinate the changing nature of languages
- Allows for better communication and the same labeling between service providers, clients, and content owners
LMT Use Cases

- Licensing international content
  - As the industry seeks to scale new content development, more organizations are looking at international productions to expand inventory

- Distributing non-English content
  - As that international content gets distributed, platforms need metadata describing the language elements: audio tracks, subtitles, UI, etc.
  - Geographic information needs to be considered as part of the distribution metadata

- Accessibility requirements
  - How are closed-captions and other accessible elements described to the end-user?

- End-user localization preferences
  - Consumers want to view both content and settings in their desired languages wherever they are viewing, watching, and/or listening
  - UI/UX preferences
LMT Implementation Examples

- Audio: Allows for standard description of the audio languages corresponding to content to a provider or client
- Closed Captions: LMT can distinguish between the audio language of the content and the caption or written language when sending materials to a vendor
- Burned In or Forced Narratives: Physical signs in the content often appear in different written language than subtitles for audio language
- Accessibility: Visual description or American Sign Language can be distinguished from other languages for the content
- Acquisition/Rights: The overall language can be used to show territorial or distribution for content that rolls up any audio or other languages
- Electronic Sell-Through Partners: Languages for viewing can be displayed on the screen in the correct dialect vs the subtitle or audio languages
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 provides commonly used codes supported by use cases in actual use, pre-constructed for easy reference.**
- **Within LMT, 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>Audio Language Display Name 1</td>
<td>English</td>
<td>Español como se habla en América Latina</td>
<td>Srpski</td>
<td>普通话</td>
<td>հայերեն</td>
<td>Ῥուսականորեն</td>
<td></td>
</tr>
<tr>
<td>Audio Language Display Name 2</td>
<td></td>
<td></td>
<td></td>
<td>zh-Hans</td>
<td>hy</td>
<td>hyw</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></td>
<td></td>
<td>sr-Cyr-RS</td>
<td>zh-Hans</td>
<td>hy</td>
<td>hyw</td>
<td></td>
</tr>
<tr>
<td>Visual Language Display Name 1</td>
<td></td>
<td>Español como se habla en América Latina</td>
<td>Srpski</td>
<td>简体中文</td>
<td>հայերեն</td>
<td>Ῥուսականորեն</td>
<td></td>
</tr>
<tr>
<td>Visual Language Display Name 2</td>
<td></td>
<td></td>
<td></td>
<td>zh-Hans</td>
<td>hy</td>
<td>hyw</td>
<td></td>
</tr>
</tbody>
</table>

**MESA**
ISDCF and LMT: Differences & Discussions

- Source of truth? LMT or existing ISDCF
  - How hard is it for existing users to switch? Timeframe needed?
- LMT has more languages than ISDCF
  - Should all LMT languages be included in ISDCF?
- How should code differences/conflicts be resolved?
Next Steps

● Expand coverage for Latin American countries.
  ○ Add Spanish for each Spanish-speaking country
  ○ Add Portuguese for each country as supported by use cases

● Explore language use cases for Asian countries.
  ○ Subject matter experts: please apply!

● Explore more languages and dialects use cases for India and Pakistan.

● Bear in mind what is needed from Unicode for the support of diacritics and non-Latin characters

● Next LMT Working Committee meeting: November 14 @1pm
  ○ Contact MESAlliance if interested in attending

● Please share your Language needs if not yet covered by LMT
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

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/