The Language Metadata Table (LMT) Standard
V2.0 Release
Smart Content Summit East
July 27, 2019
Agenda

- Language Metadata Table Committee Introductions:
  - Chairs: Yonah Levenson, Manager, & Laura Dawson, Metadata Analyst
    Metadata Management & Taxonomy @WarnerMedia/HBO
  - Working Group contributors include: Disney, Discovery, EIDR, European Union, HBO, Lionsgate,
    MESA, NBCUniversal, Paramount, Turner, Warner Bros, WWE, + vendors & many more
- Mission Statement/Background
- Why LMT?
- Use Cases with LMT Solution
- LMT Working Committee Update
- Mission Statement
- Template for adding languages
- Questions?

July 27, 2019
LMT Mission Statement

The Language Metadata Table standard 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.

● 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.
HBO and MESAlliance

● The 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.

● In addition, the LMT has the capability to provide for description of fictional languages (i.e. Klingon), to handle HBO’s Dothraki from Game of Thrones as well as other invented languages.

● In July 2018 HBO was asked by MESAlliance to share their language tables with the media and entertainment industry and the LMT working group was formed with representatives from studios, post houses, and other media affiliates and companies.
Why create a LMT?

● Language metadata is used by every aspect of entertainment and media businesses but is rarely standardized beyond the basic two or three letter BCP or ISO codes

● There was a greater need a more flexible, granular, and modular language standard, in order to describe broad and narrow contexts

● LMT includes values for several language needs in the media and entertainment industries, including codes for:
  ○ Audio 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
What is the LMT?

● The LMT is an expandable mapping flat table structure that organizes language metadata via locations and dialects

● LMT uses the Internet Engineering Task Force (IETF) Best Current Practice (BCP) 47 codes because this standard base provides 40,000 language, script, and geographic codes, which can be combined in an exponential number of ways for greater descriptive granularity

● The product is simply a listing of fields, codes, and definitions that can be incorporated into any database architecture

● Agreement for LMT 2.0 has been reached on nearly 200 language codes and display values, with more on the way

● As with any standard, the power of the LMT increases with adoption
IETF BCP 47

- IETF: Internet Engineering Task Force
- BCP: Best Common Practice
- 47: The number of this best practice
- IETF BCP 47 consists of
  - ISO 639: Language codes
  - ISO 3166: Country codes
  - UN M. 49: UN Territory standards
- IETF BCP 47 works because
  - Language and geographic codes can be combined in more than 40K ways
  - Combine codes with territories for even more precision:
    ex: “it-CH” = Italian as spoken in Switzerland
  - Updated language names reflect contemporary cultures: “Greenlandic” updated to “Kalaallisut”
  - A WWW standard supported by W3C
How the LMT is structured

● LMT uses language groupings to keep subgroups together so, for example, Latin American Spanish and Spanish from other countries appear in the list together

● 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
Use Cases

- **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
## LMT Data Dictionary

<table>
<thead>
<tr>
<th>Column Header Name</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language Group Name</td>
<td>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.</td>
</tr>
<tr>
<td>Language Group Code</td>
<td>URN or URI for each language group value in the LMT.</td>
</tr>
<tr>
<td>Audio Language Tag</td>
<td>IETF BCP 47 language tag. Typically spoken/audio language.</td>
</tr>
<tr>
<td>Long Description 1</td>
<td>Description of language name in Latin script following IETF BCP 47 standard</td>
</tr>
<tr>
<td>Long Description 2</td>
<td>Alternate description of language name in Latin script following IETF BCP 47 standard</td>
</tr>
<tr>
<td>Audio Language Display Name 1</td>
<td>Endonym of audio language. Typically the same as Visual Language Display Name 1 but not always.</td>
</tr>
<tr>
<td>Audio Language Display Name 2</td>
<td>Alternate endonym of audio language. Typically the same as Visual Language Display Name 2 but not always.</td>
</tr>
<tr>
<td>Visual Language Tag 1</td>
<td>Script in which language is written following IETF BCP 47 standard (which calls for the tags to be presented in Latin Script).</td>
</tr>
<tr>
<td>Visual Language Tag 2</td>
<td>Alternate script in which language is written following IETF BCP 47 standard (which calls for the tags to be presented in Latin Script).</td>
</tr>
<tr>
<td>Visual Language Display Name 1</td>
<td>Endonym of written language. Typically the same as Audio Language Display Name 1 but not always.</td>
</tr>
<tr>
<td>Visual Language Display Name 2</td>
<td>Alternate written endonym. Typically the same as Audio Language Display Name 1 but not always.</td>
</tr>
<tr>
<td>URN</td>
<td>URN or URI for each language value in the LMT.</td>
</tr>
</tbody>
</table>
# LMT Example Table

<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>hyx</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>ունկնսկնահայերէն</td>
<td>հայերեն</td>
<td></td>
</tr>
<tr>
<td>Audio Language Display Name 2</td>
<td>cpncka</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual Language Tag</td>
<td>en</td>
<td>es-419</td>
<td>ar-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-Cyr-RS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></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>简体中文</td>
<td>ունկնսկնահայերէն</td>
<td>հայերեն</td>
<td></td>
<td>American Sign Language</td>
</tr>
<tr>
<td>Visual Language Display Name 2</td>
<td>cpncka</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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
Next Steps

- Latin American countries
  - Add Spanish for each country
  - Add Portuguese as supported by use cases

- Explore language use cases for Asian countries
  - Subject matter experts: please apply!
  - Note: Gracenote will be participating

- Set up fall meetings
- What are your needs?
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 links
  ○ https://www.mesalliance.org/language-metadata-table
  ○ https://www.mesalliance.org/2018/08/07/mesa-publishes-hbo-developed-me-industry-language-metadata-table/
  ○ https://www.mesalliance.org/2019/02/20/me-journal-the-language-metadata-table-lmt-an-industrywide-effort-to-collaborate/