



Rosetta Stone 2.0

Expanding the Language Metadata Table

Ben Kass, MLIS

Sponsored by MESA



What is the Language Metadata Table?

The Language Metadata Table (LMT) is a language standard created to provide a unified source of reference for language codes for use in media and entertainment.

What is the LMT used for?

The LMT is used for audio tracks; visual/written languages including subtitles, closed captioning, burned-in captions and UI; rights, licensing, and distribution information; and accessibility features such as audio description and sign language interpretation.

Formula for an LMT Code:
 lang + script + country.
 For Example:
 Spanish as Spoken in Latin America = es-419
 French as Spoken in Canada = fr-CA
 Chinese in traditional script = zh-Hant

My Project:

Improving Coverage of the LMT

Viacom, Warner Media, Sony, Disney, and the LMT Working Group requested new languages for addition into the LMT as well as updates to existing codes. I cross-mapped these language requests to identify overlaps between them. and compared them to the LMT's existing coverage as well as other language standards. I then researched Autonyms, Names, Script(s), Language Family, Usage, and more to prepare these languages for further review and ingest into the LMT.

Language Requests, by number:

Viacom: 111
 Warner Media: 92
 Sony: 33
 Disney: 6
 LMT Working Group: 3

Results:

95 languages are ready to go and have been sent to the LMT Working Group for review, for an increase of almost 40% in languages covered! Furthermore, an additional 23 Languages of the Indian subcontinent and 19 Chinese Languages were sent for further discussion and review by LMT partners. 18 Arabic and 3 Quechua languages were sent to subcommittee to determine which codes to use, while an additional 33 miscellaneous languages require further research and troubleshooting. Finally 1 fictional language, Bulbovian, has been set aside while we create fictional language guidelines.

