
Archived Reference Label Generation Rules

Version 24 January 2024

This version of Second-Level Reference Label Generation Rules (LGRs) were developed in consultation with the respective script communities following detailed analysis of the [Root Zone Label Generation Rules](#) (RZ-LGR) and after a [Public Comment proceeding](#).

This version includes:

- The script-based Reference LGRs including seven new scripts: Armenian, Cyrillic, Greek, Latin, Japanese, Korean, and Myanmar.
- In some cases, LGRs may be used in concert with other LGRs under a TLD. In such case, a set of "full-variant" LGR has been defined that collectively contain the suggested cross-script variants identified to mitigate whole-script homograph labels.
- A Common LGR has been created by merging the data from the script based LGRs. This file is intended for collision checking, particularly in the case where multiple LGRs are used in the same zone.
- Language LGRs which were derived a repertoire that is based on the Root Zone LGR but restricted to specific languages.

See the [Overview and Summary document](#) for further details about these LGRs. The package of all LGRs is available [here](#) [ZIP, 7.3MB]. The changes from previous are document in this [change log document](#).

Script-based LGRs

Name	Language Tag ^[1]	LGR Document
Arabic	und-Arab	HTML , XML
Armenian	und-Armn	HTML , XML
Bangla (Bengali)	und-Beng	HTML , XML
Cyrillic	und-Cyrl	HTML , XML
Devanagari	und-Deva	HTML , XML
Ethiopic	und-Ethi	HTML , XML
Georgian	und-Geor	HTML , XML
Greek	und-Grek	HTML , XML
Gujarati	und-Gujr	HTML , XML
Gurmukhi	und-Guru	HTML , XML
Hebrew	und-Hebr	HTML , XML
Japanese	und-Jpan	HTML , XML
Kannada	und-Knda	HTML , XML
Khmer	und-Khmr	HTML , XML
Lao	und-Laoo	HTML , XML
Latin	und-Latn	HTML , XML
Malayalam	und-Mlym	HTML , XML

Name	Language Tag ^[1]	LGR Document
Myanmar	und-Mymr	HTML , XML
Oriya	und-Orya	HTML , XML
Sinhala	und-Sinh	HTML , XML
Tamil	und-Taml	HTML , XML
Telugu	und-Telu	HTML , XML
Thai	und-Thai	HTML , XML

[1]: The prefix 'und' (Undetermined) identifies linguistic content whose language is not determined. Please see [RFC5646](#) for details of the language tag syntax and [IANA language sub tag registry](#) for the available language tags.

Full Variant Set LGRs and Common LGR

Name	Language Tag	Script Collection	LGR Document
Chinese (Full Variant Set)	und-Hani	Han used in Chinese, Korean, Japanese scripts	HTML , XML
Devanagari (Full Variant Set)	und-Deva	Devanagari, Bengali, and Gurmukhi	HTML , XML
Korean (Full Variant Set)	und-Kore	Hangul and Han used in Chinese and Korean script	HTML , XML
Latin (Full Variant Set)	und-Latn	Armenian, Cyrillic, Greek, Hebrew, and Latin	HTML , XML
Myanmar (Full Variant Set)	und-Mymr	Georgian, Latin, Malayalam, Myanmar, and Oriya	HTML , XML
Tamil (Full Variant Set)	und-Taml	Tamil and Malayalam	HTML , XML
Telugu (Full Variant Set)	und-Telu	Kannada and Telugu	HTML , XML
Common LGR	Multiple Tags	All scripts	HTML , XML

Language-based LGRs

Name	Language Tag ^[2]	LGR Document
Arabic	ar	HTML , XML
Belarusian	be	HTML , XML
Bosnian (Cyrillic)	bs-Cyrl	HTML , XML
Bosnian (Latin)	bs	HTML , XML
Bulgarian	bg	HTML , XML
Chinese	zh	HTML , XML
Danish	da	HTML , XML
English	en	HTML , XML
Finnish	fi	HTML , XML
French	fr	HTML , XML
German	de	HTML , XML
Hebrew	he	HTML , XML

Name	Language Tag ^[2]	LGR Document
Hindi	hi	HTML , XML
Hungarian	hu	HTML , XML
Icelandic	is	HTML , XML
Italian	it	HTML , XML
Japanese (Standalone)	ja	HTML , XML
Korean (Hangul)	ko	HTML , XML
Latvian	lv	HTML , XML
Lithuanian	lt	HTML , XML
Macedonian	mk	HTML , XML
Montenegrin	cnr-Cyrl	HTML , XML
Norwegian	no	HTML , XML
Polish	pl	HTML , XML
Portuguese	pt	HTML , XML
Russian	ru	HTML , XML
Serbian	sr-Cyrl	HTML , XML
Spanish	es	HTML , XML
Swedish	sv	HTML , XML
Thai	th	HTML , XML
Ukrainian	uk	HTML , XML

[2]: Where the default script is not identified, the script information is included to avoid ambiguity.