

Guidelines for test of use of ISO 639-3 language codes in MARC records

ISO 639-3 is a comprehensive list of nearly 8,000 three-letter codes for individual languages that expands on ISO 639-2. ISO 639-2 includes the same languages as the MARC language codes.¹ Using ISO 639-3 to encode language information in MARC bibliographic and authority records will allow us to provide users with more accurate and granular language information.

Please submit feedback and questions through this [Google form](#)

Differences between ISO 639-3 and MARC language codes (ISO 639-2/B)

Differences in code without differences in meaning

The strings used for some language codes are different in ISO 639-3. For example, German is “deu” not “ger” and Maori is “mri” not “mao.”

Additional historical languages

ISO 639-3 distinguishes more historical languages, such as Old Georgian (“oge”) and Middle Korean (“okm”).

Additional codes for other types of languages

ISO 639-3 includes many specific codes for sign languages, such as Brazilian Sign Language “bzs”, artificial languages, such as Quenya (“qya”), ancient languages, such as Church Slavic (“chu”) and extinct languages, such as Powhatan (“pim”).

Macrolanguages

ISO 639-3 uses macrolanguage codes to identify “clusters of closely-related language varieties that ... can be considered distinct individual languages, yet in certain usage contexts a single language identity for all is needed.”² Examples of macrolanguage codes include “ara” (Arabic), “zho” (Chinese), “nor” (Norwegian) and “hbs” (Serbo-Croatian). The macrolanguage for Chinese (“zho”) is related to 16 individual languages, including Mandarin Chinese (“cmn”) and Yue Chinese/Cantonese (“yue”).

[Recommendations for coding macrolanguages](#) are given below.

Collective codes

There is no ISO 639-3 equivalent for the MARC language codes for language groups, such as “cel” for Celtic (Other) and “nub” for Nubian languages.

Special codes

The special codes “mis,” “mul,” “und” and “zxx” are unchanged from the MARC language codes.

¹ ISO 639-2 is composed of two standards: ISO 639-2/B (bibliographic) and ISO 639-2/T (terminology). These both cover the same languages, but in some cases the codes are different. ISO 639-2/B is equivalent to the MARC language codes. ISO 639-3 uses the ISO 639-2/T codes.

² <https://iso639-3.sil.org/about/scope#Macrolanguages>

Backwards compatibility

Some systems can only handle MARC language codes or require a code in the language fixed field (008/35-37). Non-MARC language codes cannot be input in 008/35-37. In order to ensure that records created or updated during this test are usable in all systems, in shared cataloging environments, code all languages using both MARC language codes and ISO 639-3. If ISO 639-3 is implemented in PCC cataloging, we anticipate that a macro can be developed to generate MARC language codes from ISO 639-3 codes.

How to input

Bibliographic record

Code languages using MARC language codes (in the 008/35-37, and if needed in an 041) as you normally would. If your regular cataloging practice includes entering a 377 field under certain circumstances, enter it using the MARC code as you normally would.

Additionally, code languages using the ISO 639-3 codes in a separate 041 field (and a separate 377 field, if a 377 with a MARC code has been entered). If possible, use the specific language. Otherwise, use “und” for undetermined.

041 (Language Code)

Use second indicator 7 and \$2 iso639-3.

```
041 17 $a fra $h grc $2 iso639-3
```

[OPTIONAL] 377 (Associated Language)

Use second indicator 7 and \$2 iso639-3.

```
377 #7 $a ben $2 iso639-3
```

Authority record

Code languages using MARC language codes (in 377) as you normally would.

Additionally, code languages using the ISO 639-3 codes in a separate 377 field. If possible, use the specific language. Otherwise, use “und” for undetermined.

377 (Associated Language)

Use second indicator 7 and \$2 iso639-3.

```
377 #7 $a leh $2 iso639-3
```

Where to find the codes

SIL is the official maintainer of ISO 639-3. ISO 639-3 language codes may be searched at the SIL site at: https://iso639-3.sil.org/code_tables/639/data. To download, navigate to the link for 'Code Tables'.

Mouseover without clicking at first, and draw your mouse down and to the right, to a link to 'Download Tables', then click there.

Unfortunately, the SIL site does not include cross-references for language names. We list below some other sources that do include cross-references and may assist in identifying language codes for particular situations, but please confirm your choice of language codes on the official SIL site.

Ethnologue can be searched directly (if you have an institutional subscription) at www.ethnologue.com.

LCSH often has cross-references to language names using Ethnologue as a source. The way that language names are authorized in LCSH often corresponds, but not always completely, to the way they have been established in the Ethnologue and in the ISO 639-3 data files.

Lexvo.org allows lookup of three-letter ISO 639-3 codes with a search box for "Lookup Code" at <http://www.lexvo.org/>.

Wikidata supplies links to three-letter ISO 639-3 codes and is searchable by cross-references of language names at www.wikidata.org.

Glottolog also references ISO 639-3 codes, and is searchable by parts of language names and by a "glottocode" at <https://glottolog.org/>.

Macrolanguages

https://iso639-3.sil.org/code_tables/macrolanguage_mappings/read

For languages that are considered to be part of macrolanguages in ISO 639-3, use the most specific code possible.

For standard literary languages, use the specific code for the literary language, if one exists. Otherwise, use the code corresponding to the spoken language on which the standard literary language is based.

In any situation, if you are not able to identify the appropriate specific language code, use the macrolanguage code.

Guidelines for languages commonly found in library collections follow below. These are all intended to be consistent with the principles given above.

Albanian - For standard literary Albanian, use als (Tosk Albanian).

Arabic - For standard modern Arabic, use arb (Standard Arabic).

Azerbaijani - For standard Azerbaijani of Azerbaijan (usually written in Latin or Cyrillic script), use azj (North Azerbaijani). For Azerbaijani of Iran (usually written in Arabic script), use azb (South Azerbaijani).

Bosnian, Croatian, Montenegrin, Serbian - Continue current MARC code practice and distinguish among these using separate codes: bos (Bosnian), hrv (Croatian), cnr (Montenegrin), srp (Serbian).

Chinese - For standard written Chinese, use cmn (Mandarin Chinese).

Estonian - For standard Estonian, use ekk (Standard Estonian).

Kurdish - For standard Kurmanji Kurdish (usually written in Latin script), use kmr (Northern Kurdish). For standard Sorani Kurdish (usually written in Arabic script), use ckb (Central Kurdish).

Latvian - For standard Latvian, use lvs (Standard Latvian).

Malagasy - For standard Malagasy, use plt (Plateau Malagasy).

Malay - For standard Malay, use zsm (Standard Malay).

Mongolian - For standard Mongolian from Mongolia (usually written in Cyrillic script), use khk (Halh Mongolian).

Nepali - For standard Nepali, use npj (Nepali (individual language))

Norwegian - For the two standard forms of Norwegian, use nob (Norwegian Bokmål) or nno (Norwegian Nynorsk), as appropriate. Note: this is a change from current practice (LC-PCC PS for 6.11.1.3), where we are told to use nor in all cases, although these two more specific codes are also available in MARC.

Oriya - For standard Oriya, use ory (Odia).

Persian - For standard Persian, use pes (Iranian Persian).

Swahili - For standard Swahili, use swl (Swahili (individual language)).

Uzbek - For standard Uzbek of Uzbekistan (usually written in Latin or Cyrillic script), use uzn (Northern Uzbek). For Uzbek of Afghanistan (usually written in Arabic script), use uzs (Southern Uzbek).

Examples

All ISO 639-3 language codes are the same as corresponding MARC language codes

```
008/35-37 (Lang): spa
041 07 $a spa $2 iso639-3
245 10 $a Cien años de soledad / $c Gabriel García Márquez.
```

```
008/35-37 (Lang): eng
041 0# $a eng $a haw
041 07 $a eng $a haw $2 iso639-3
245 00 $a Hawaiian word book just for kids.
546 ## $a Text in English and Hawaiian.
```

Some ISO 639-3 language codes are the same as corresponding MARC language codes

008/35-37 (Lang): bnt
041 1# \$a bnt \$h heb \$h grc
041 17 \$a mhw \$h heb \$h grc \$2 iso639-3
245 10 \$a Testamende ghompya naPisalemi.
500 ## \$a "The New Testament and Psalms in Mbukushu"--Title page verso.

008/35-37 (Lang): eng
041 1# \$a eng \$h wel
041 17 \$a eng \$h cym \$2 iso639-3
245 10 \$a One moonlit night / \$c Caradog Prichard ; translated by Philip Mitchell ; [introduction by Harri Pritchard Jones].
546 ## \$a English translated from Welsh.

No ISO 639-3 language codes are the same as corresponding MARC language codes

008/35-37 (Lang): fre
041 07 \$a fra \$2 iso639-3
245 10 \$a Dictionnaires de langue française du Canada : \$b lexicographie et société au Québec / \$c Annick Farina ; préface de Claude Poirier.

008/35-37 (Lang): ssa
041 07 \$a mls \$2 iso639-3
245 00 \$a Âsurti, âsurtu geya = \$b Le champ, la culture.
546 ## \$a In Masalit.

008/35-37 (Lang): fre
041 0# \$a fre \$a map
041 07 \$a fra \$a piz \$a fwa \$a nem \$a jaz \$2 iso639-3
245 10 \$a Dictionnaire thématique des langues de la région de Hienghène (Nouvelle-Calédonie) : \$b Pije, Fwâi, Nemi, Jawe / \$c André-Georges Haudricourt, Françoise Ozanne-Rivierre. Précédé d'une Phonologie comparée des langues de Hienghène et du proto-océanien / par Françoise Ozanne-Rivierre.
546 \$a Contains text in Pije, Fwâi, Nemi, and Jawe.

Languages within macrolanguages

008/35-37 (Lang): cnr
041 0# \$a cnr \$a ger
041 07 \$a cnr \$a deu \$2 iso639-3
245 00 \$a Crnogorska poljska bolnica = \$b Montenegrinisches Feldlazaret.
300 ## \$a 1 postcard : \$b black and white ; \$c 9 x 14 cm
Montenegrin is part of macrolanguage Serbo-Croatian (hbs)

008/35-37 (Lang): nor
041 07 \$a nob \$2 iso639-3
245 10 \$a Inne i fjordene / \$c Peter Egge
Norwegian (Bokmål) is part of macrolanguage Norwegian (nor).

008/35-37 (Lang): eng
041 0# \$a eng \$a twi \$a gaa \$a fat \$a ewe
041 07 \$a eng \$a twi \$a gaa \$a fat \$a ewe \$2 iso639-3
245 10 \$a Christian messenger.
546 ## \$a Chiefly in English, some Twi, Ga, Fanti and Ewe.
Fanti and Twi are part of macrolanguage Akan (aka).

008/35-37 (Lang): ara
041 0# \$a ara \$b eng
041 07 \$a arz \$b eng \$2 iso639-3
245 00 \$a CALLHOME Egyptian Arabic Speech.
520 ## \$a The CALLHOME Egyptian Arabic corpus of telephone speech consists of 120 unscripted telephone conversations between native speakers of Egyptian Colloquial Arabic (ECA), the spoken variety of Arabic found in Egypt.
Egyptian Arabic is part of macrolanguage Arabic (ara).

008/35-37 (Lang): chi
041 1# \$a chi \$j chi \$j eng \$h chi
041 17 \$a cmn \$j cmn \$j eng \$h cmn \$2 iso639-3
245 10 \$a 危险关系 / \$c 制片人, 陈伟明; 导演, 许秦豪; 影片摄制, 北京中博世纪影视传媒有限公司 = Dangerous liaisons / producer, Chen Weiming; director, Hur Jin Ho; produced by Zonbo Media.
245 10 \$a Wei xian guan xi / \$c zhi pian ren, Chen Weiming; dao yan, Hǒ Chin-ho; ying pian she zhi, Beijing zhong bo shi ji ying shi chuan mei you xian gong si = Dangerous liaisons / producer, Chen Weiming; director, Hur Jin Ho; produced by Zonbo Media.
546 ## \$a In Mandarin with optional subtitles in traditional Chinese or English.
Mandarin Chinese is part of macrolanguage Chinese (zho).

008/35-37 (Lang): chi
041 1# \$a chi \$j chi \$j eng \$h chi
041 17 \$a yue \$a cmn \$j cmn \$j eng \$h yue \$2 iso639-3
245 00 \$a 少林足球 = \$b Shaolin soccer / \$c 寰宇娛樂有限公司; 星輝海外有限公司; 編劇周星馳, 曾謹昌; 監製楊國輝; 周星馳導演.
245 00 \$a Shao lin zu qiu = \$b Shaolin soccer / \$c Huan yu yu le you xian gong si; Xing hui hai wai you xian gong si; bian ju Zhou Xingchi, Zeng Jinchang; jian zhi Yang Guohui; Zhou Xingchi dao yan.
546 ## \$a In Chinese (Cantonese or Mandarin) with optional subtitles in traditional Chinese, simplified Chinese or English.
Yue Chinese (Cantonese) and Mandarin Chinese are part of macrolanguage Chinese (zho).

008/35-37 (Lang): lav
041 0# \$d lat \$d lav \$e eng \$e lat \$e lav \$n eng \$n lat \$n lav \$g eng \$g fre \$g ger \$m eng
041 07 \$d lat \$d lvs \$e eng \$e lat \$e lvs \$n eng \$n lat \$n lvs \$g eng \$g fra \$g deu \$m eng \$2 iso639-3
245 10 \$a Baltic exchange : \$b choral works / \$c by Prauliņš [and others].
546 ## \$a Sung in Latin and Latvian.
500 ## \$a Program and biographical notes in English, with German and French translations, and texts with English translations (19 p. : ports.) inserted in container.
Standard Latvian is part of macrolanguage Latvian (lav).

More specific not known

008/35-37 (Lang): spa

041 0# \$d spa \$d eng \$d myn \$d zap \$e spa \$e eng \$e myn \$g eng \$g spa

041 07 \$d spa \$d eng \$d und \$d zap \$e spa \$e eng \$e und \$g eng \$g spa \$2 iso639-3

245 10 \$a Border : \$b la linea / \$c Lila Downs.

546 ## \$a Sung primarily in Spanish; some selections in Maya, Zapotec, or English.

500 ## \$a Lyrics in Spanish, English and Maya, most with English and, where necessary, Spanish translations ([20] p. : ill. (some col.)) inserted in container.

The MARC language code myn is used for all Maya languages as a group. In ISO 639-3, Maya is not considered a macrolanguage, and only codes for specific Maya languages are available, so if you do not know which Maya language is used in the resource, your only option is to use undetermined (und).

Historical

008/35-37 (Lang): rus

041 1# \$a pol \$a rus \$h rus

041 17 \$a pol \$a rus \$a orv \$h orv \$2 iso639-3

245 00 \$a Słowo o pułku Igorowym / \$c przekład i opracowanie Jana Dalibora Wagilewicz ; do druku przygotował i artykułem wstępnym poprzedził Oleg Kupczyński.

546 ## \$a Polish and Russian translation with Old Russian original.

The MARC language codes do not distinguish between modern Russian and Old Russian, but ISO 639-3 has separate codes for them.

Artificial

008/35-37 (Lang): art

041 0# \$a art \$a eng

041 07 \$a sjn \$a eng \$2 iso639-3

245 00 \$a Sindarin-English & English-Sindarin dictionary / \$c edited and arranged by J.M. Carpenter.

650 #0 \$a Sindarin (Artificial language) \$v Dictionaries.

Some artificial languages have their own MARC language codes, but this one does not, so a collective code is used. In ISO 639-3, it has its own code.