

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

Approved Nov. 22, 2022; last revised: Jan. 12, 2023

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.

Many current 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 PCC records are usable in all systems, the guidelines that follow instruct PCC catalogers working in shared cataloging environments to encode all languages using MARC language codes and, optionally, also using ISO 639-3 language codes. However, it is possible that in the future macros and other automated means will be developed to generate MARC language codes if catalogers have manually entered ISO 639-3 codes, eliminating the need for double coding.

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

How to input	2
Bibliographic record	2
041 (Language Code)	2
Authority record	3
377 (Associated Language)	3
Where to find the codes	3
Macrolanguages	3
Historical languages	5
Examples	6
1. ISO 639-3 adds new information for <i>all</i> languages related to a resource	6
2. ISO 639-3 adds new information for <i>some</i> languages related to a resource	7
3. ISO 639-3 and MARC language codes for languages related to a resource are equivalent	8
4. The ISO 639-3 language code for at least one language related to a resource cannot be identified	8

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.

5. Languages within macrolanguages	9
Appendix. Differences between ISO 639-3 and MARC language codes (ISO 639-2/B)	11
Differences in code without differences in meaning	11
Additional historical languages	12
Additional codes for other types of languages	12
Macrolanguages	12
Collective codes	12
Special codes	12

How to input

Bibliographic record

Always code the language(s) of the resource being cataloged by using MARC language codes (in the 008/35-37, and if needed in an 041).

Optionally, also code the language(s) of the resource in an additional 041 field using ISO 639-3 codes. It is especially encouraged to enter an 041 field using ISO 639-3 codes when those codes will identify any language more specifically than the MARC codes do:

- when the MARC codes include only a collective code applicable to the language of the resource, but the ISO 639-3 codes include a specific code for that language (see [Example 1.1](#) and [Example 1.2](#)).
- when the MARC code for a different language is also used as a collective code for the language of the resource, but the ISO 639-3 codes include a specific code for the language of the resource (see [Example 1.3](#)).

If using ISO 639-3 codes does not add any additional information about the languages of the resource, it is not necessary to create an 041 field using ISO 639-3 codes. An 041 using ISO 639-3 language codes should be able to be created programmatically later.

If multiple MARC codes have been recorded in the record, generally enter all the corresponding ISO 639-3 codes, including those that do not add new information, whenever an 041 with ISO 639-3 codes is entered. (See [Example 2.1](#)).

If the cataloger is creating an 041 field using ISO 639-3 codes and the appropriate ISO 639-3 code cannot be identified for one or more languages of the resource, use the code **und** (undetermined) as a placeholder. (See [Example 4.2](#)).

041 (Language Code)

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

041 07 \$a fra \$a gsl \$2 iso639-3

Authority record

When appropriate and when information is available, code languages in a 377 field using MARC language codes.

Optionally, also code languages using the ISO 639-3 codes in a separate 377 field. It is especially encouraged to enter a 377 field using ISO 639-3 codes when those codes will identify any language more specifically than the MARC codes do. If the cataloger is creating a 377 field using ISO 639-3 codes and the appropriate ISO 639-3 code cannot be identified for one or more languages of the resource, use the code **und** (undetermined) as a placeholder.

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 or downloaded at the SIL site at: https://iso639-3.sil.org/code_tables/639/data.

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. It covers languages in modern use, but not historical ones.

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.

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

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 <https://www.wikidata.org/>.

Macrolanguages

ISO 639-3 defines [macrolanguages](https://iso639-3.sil.org/code_tables/macrolanguage_mappings/read) as "clusters of closely-related language varieties" where "in certain usage contexts a single language identity for all is needed." A list of ISO 639-3 Macrolanguage Mappings can be found on the SIL site: 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 follow below for standard (literary) languages within macrolanguages that are commonly found in library collections. 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). (See [Example 5.1](#)).

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 **npi** (Nepali (individual language))

Norwegian - For the two standard forms of written Norwegian, use **nob** (Norwegian Bokmål) or **nno** (Norwegian Nynorsk), as appropriate. Note: However, when recording the MARC language codes, continue to follow current LC/PCC 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 the MARC list. (See [Example 5.2](#)).

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

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

Swahili - For standard Swahili, use **swh** (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).

Historical languages

ISO 639-3 includes more codes for earlier forms of languages than the MARC codes do. For example, while both vocabularies include the code **fro** (Old French (842-ca. 1400)), only ISO 639-3 includes the code **osp** (Old Spanish). The MARC code list instructs catalogers to use the group code **roa** (Romance (Other)) for Old Spanish. Catalogers are encouraged to add the ISO 639-3 code for a historical language whenever it is appropriate for the resource being described and would provide more detailed information than the MARC code does.

- If the ISO language name does not contain dates indicating the time period covered by the code, research or community guidance may be required.
- If there is no separate code for an earlier form of a language, use the code for the modern language for all periods.
- Whenever needed, additional information about the language can be recorded in a 546 note.

041 17 \$a **ang** \$2 iso639-3 [*code used for Old English (ca. 450-1100)*]

041 17 \$a **osp** \$2 iso639-3 [*code used for Old Spanish; name does not include dates*]

041 17 \$a **por** \$2 iso639-3 [*code used for Portuguese: all periods*]

For resources in historical languages, particularly when the codes are not associated with date ranges, catalogers may wish to add information about the date of composition using 046 and/or 388 following the ALA Subject Analysis Committee best practices at <https://alair.ala.org/handle/11213/16710>.

Cantar de mio Cid was composed in Old Spanish sometime between 1140 and 1207.

008/35-37 (Lang): **fre**

041 1# \$a **fre** \$a **roa** \$h **roa**

041 17 \$a **fra** \$a **osp** \$h **osp** \$2 iso639-3

046 ## \$k [1140..1207] \$2 edtf

245 00 \$a Cantar de mio Cid = \$b Chanson de mon Cid.

546 ## \$a Old Spanish and French on opposite pages.

Examples

1. ISO 639-3 adds new information for *all* languages related to a resource

In this situation, we strongly recommend adding ISO 639-3 codes if the language(s) can be accurately determined.

1.1. *ISO 639-3 has a code specifically for Fuliiru (also spelled Fuliru), while in the MARC code set the collective code for Bantu (Other) must be used for this language.*

```
008/35-37 (Lang): bnt  
041 07 $a flr $2 iso639-3  
245 00 $a Akabwa, ne'mbeba, no'mundu.  
546 ## $a In Fuliru.
```

1.2. *ISO 639-3 has a code specifically for Old Spanish, while in the MARC code set the collective code **roa** for Romance (Other) must be used for this language.*

```
008/35-37 (Lang): roa  
041 07 $a osp $2 iso639-3  
100 0# $a Alfonso $b X, $c King of Castile and Leon, $d 1221-1284.  
245 10 $a Privilegio rodado : $b [Segovia, Spain] : $k ms., $f [1256 Oct. 24].
```

1.3. *In the MARC language codes, Alemannic dialects like Basel German are assigned the collective code **ger**, but ISO 639-3 has a separate code for Alemannic.*

```
008/35-37 (Lang): ger  
041 07 $a gsw $2 iso639-3  
100 1# Wolf, David, $d 1918-  
245 10 $a Was blybt? : $b Gidicht und Gidanggen uff Baaseldytsch / $c David Wolf ; mit Fotene vom Beat Trachsler
```

2. ISO 639-3 adds new information for *some* languages related to a resource

In this situation, we strongly recommend adding ISO 639-3 codes if the language(s) can be accurately determined. Recode all the languages that were coded using MARC language codes in ISO 639-3 whether or not they add new information. This will improve consistency and make the fields easier to evaluate for completeness.

2.1. *The codes for Ancient Greek are the same in both MARC and ISO 639-3. However, ISO 639-3 has a code specifically for Mbukushu, while in the MARC code set the collective code for Bantu (Other) must be used for this language.*

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

2.2. *The code for English is the same in both MARC and ISO 639-3. However, ISO 639-3 has a code specifically for Middle Welsh, while in the MARC code set the collective code for Celtic (Other) must be used for this language.*

008/35-37 (Lang): **eng**
 041 1# \$a **eng** \$h **cel**
 041 17 \$a **eng** \$h **wlm** \$2 iso639-3
 245 10 \$a The Mabinogion / \$c translated with an introduction by Jeffrey Gantz.
 546 ## \$a English, translated from Middle Welsh.

2.3. *The codes for French in MARC and ISO 639-3 are different but equivalent; ISO 639-3 has specific codes for Pije, Fwâi, Nemi, and Jawe, while in the MARC code set the collective code for Austronesian (Other) must be used for all these languages.*

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.

3. ISO 639-3 and MARC language codes for languages related to a resource are equivalent

Although it is fine to enter ISO 639-3 codes for the language(s) of these resources, they do not add any new information to the record. Since these ISO 639-3 codes could be generated programmatically from the corresponding MARC codes at any point in the future, it is not necessary to add them manually in current cataloging.

3.1. *Although the string for the ISO 639-3 code for French (fra) is different from the MARC language code (fre), they are equivalent in meaning, so it is not necessary to add the ISO 639-3 code.*

008/35-37 (Lang): **fre**

041 07 \$a **fra** \$2 *iso639-3* [**may be entered but adds no new information**]²

245 10 \$a Dictionnaires de langue française du Canada : \$b lexicographie et société au Québec / \$c Annick Farina ; préface de Claude Poirier.

3.2. *The codes for English and Hawaiian are the same in both MARC and ISO 639-3.*

008/35-37 (Lang): **eng**

041 0# \$a **eng** \$a **haw**

041 07 \$a **eng** \$a **haw** \$2 *iso639-3* [**may be entered but adds no new information**]

245 00 \$a Hawaiian word book just for kids.

546 ## \$a Text in English and Hawaiian.

4. The ISO 639-3 language code for at least one language related to a resource cannot be identified

If one or more languages associated with the resource cannot be accurately identified with the granularity required by ISO 639-3, but the cataloger is recording an ISO 639-3 041 field to better identify another language in the resource, represent the unknown language or languages with the code **und** where appropriate.

4.1 *No information can be determined about the language of the resource*

008/35-37 (Lang): **und**

041 07 \$a **und** \$2 *iso639-3* [**may be entered but adds no new information**]

245 00 \$a Our point of view / \$c Blue Note All-Stars.

264 1# \$a Hollywood, California : \$b Blue Note Records, \$c [2017]

546 ## \$a Unknown language sung in vocal selections.

² Strictly speaking, the MARC language code *fre* is also used as the collective code for additional languages, such as Allemand French and Judeo-French. However, those languages are very rarely used in bibliographic records. Therefore, for pragmatic reasons, we do not currently recommend routinely adding a 041 field using ISO 639-3 codes just to mark the majority of cases that are standard modern French. Instead, we hope that heuristics and algorithmic approaches will be developed that can automatically map this coding in the majority of cases.

4.2 Both Border Kuna and Choco are covered by the MARC language code for indigenous languages of Central America (**cai**). Border Kuna has a specific ISO 639-3 code **kvn**. Choco refers to a language family, so has no ISO 639-3 code. Code an 041 field with ISO 639-3 codes, and enter **kvn** for Border Kuna and **und** for Choco, because the specific language within the Choco language family is not known.

008/35-37 (Lang): **cai**
 041 0# \$d **cai** \$h **cai**
 041 07 \$d **kvn** \$d **und** \$h **kvn** \$h **und** \$2 iso639-3
 245 00 Music of the Indians of Panama.
 260 ## [Washington, D.C.] : \$b Smithsonian Folkways Recordings, \$c [©2001?]
 546 ## Sung in Border Kuna or Choco.

5. Languages within macrolanguages

When it is possible to identify a language within a macrolanguage and locate a matching ISO 639-3 code, prefer the specific code over the macrolanguage one.

5.1. *Montenegrin (cnr) is part of macrolanguage Serbo-Croatian (hbs)*

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

5.2. *Norwegian (Bokmål) is part of macrolanguage Norwegian (nor).*

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

5.3. *Egyptian Arabic is part of macrolanguage Arabic (ara).*

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.

5.4. *Yue Chinese (Cantonese) and Mandarin Chinese are 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.

Appendix. 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.” These are listed below. Note that some of these languages are treated as macrolanguages or may have additional codes for related historical languages or other variants in ISO 639-3.

639-3 Label	MARC	639-3
Albanian	alb‡	<i>sqj*</i>
Armenian	arm‡	hye
Basque	baq	eus
Burmese	bur	mya
Chinese	chi	<i>zho*</i>
Czech	cze	ces
Dutch	dut	nld
French	fre‡	fra
Georgian	geo‡	kat
German	ger‡	deu

639-3 Label	MARC	639-3
Modern Greek (1453-)	gre‡	ell
Icelandic	ice	isl
Macedonian	mac	mkd
Maori	mao‡	mri
Malay (macrolanguage)	may‡	<i>msa*</i>
Persian	per‡	<i>fas*</i>
Romanian	rum‡	ron
Slovak	slo	slk
Tibetan	tib‡	bod
Welsh	wel	cym

‡These MARC language codes are also used as collective codes for other languages.

*Macrolanguage code; use **only** if a more specific ISO 639-3 code cannot be identified (see [Macrolanguages section](#) above).

Additional historical languages

ISO 639-3 distinguishes more historical languages, such as Old Georgian (“oge”) and Middle Korean (“okm”). To see the complete list of historical language codes, go to the [ISO 639 Code Tables Entry View](#).

Additional codes for other types of languages

ISO 639-3 includes many specific codes for sign languages, such as Brazilian Sign Language “bzs”; artificial or constructed 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.”³ 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 above.

Collective codes

There is no ISO 639-3 equivalent for the MARC language codes for language groups, such as “cel” for Celtic (Other), a remainder group, and “nub” for Nubian languages, an inclusive group. Unlike in the MARC language codes, ISO 639-3 never uses the code for one language as a collective code for other languages.

Special codes

The [special codes](#) “mis,” “mul,” “und” and “zxx” are unchanged from the MARC language codes.

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