



# USING EIDR LANGUAGE CODES

---

*Technical Note*

## **Table of Contents**

Introduction.....	2
Recommended Data Entry Practice .....	2
Original Language.....	2
Version Language .....	3
Title, Alternate Title, Description .....	3
Constructing an EIDR Language Code.....	3
Language Tags .....	4
Extended Language Tags .....	4
Script Tags .....	5
Region Tags .....	5
Variant Tags.....	5
Language Code Tables.....	6
Common Languages .....	6
Sign Languages .....	10
Common Scripts.....	12
Regions .....	13
Special Dialects and Language Variants.....	19

## Introduction

Language codes appear in several places within the EIDR registry, generally either indicating the language expressed in a data field (for example, the language of a title string) or the language used in a piece of media (for example, the language spoken in a particular version of a work). To ensure universal comprehension and facilitate automated record de-duplication, EIDR uses a particular standard means of expressing language information.

EIDR language codes are quite flexible, and this can sometimes cause confusion. They allow for the designation of basic languages (“es” for Spanish) and regional dialects (“es-MX”, the dialect of Spanish most often spoken in Mexico, as differentiated from “es-419”, a form of Spanish that is intelligible throughout Latin America). It is also possible to designate the script used with a written language when a language might be expressed in different forms (“sr-Latn” for the Serbian language written using a Latin script as opposed to “sr-Cyrl” for Serbian written in Cyrillic). All of this is defined by RFC 5646, *Tags for Identifying Languages*, published September 2009.<sup>1</sup> This is the specification used by the XML lang tag, so EIDR language encodings can be validated using a standards-compliant XML parser. Such tags are case-insensitive, but to aid in human recognition, EIDR conforms to the casing conventions recommended in RFC 5646.

Using RFC 5646, it is possible to construct the same language code in different ways, though the shortest option is generally preferred. The EIDR Web UI provides the most common language codes as a drop-down selection list to aid data entry. If necessary, the user can select “Other” and manually enter a less common or more complex language code.

**NOTE:** If you know the language you wish to designate, but are not certain what code to use, you can skip ahead to the Common Languages table below.<sup>2</sup>

## Recommended Data Entry Practice

Where possible, use the shortest valid language code for a given situation. If only the primary language family (or macro-language) is known for certain, do not guess at the possible extended language sub-tags or region codes. For example, if you know a work is in Chinese, but are not sure if it is Mandarin or Cantonese, do not guess. Code that as “zh”.

**NOTE:** There is an implied precedence within the languages, so list the most important or common first, followed by any others in decreasing order of importance. The order has no impact on search or de-duplication; it only exists as a convenience for human review.

## Original Language

### Audio

Identify the most common or contextually important audible language(s) in the work as presented during the work’s original release. If the region, dialect, or language extension is not a critical identifying characteristic of the work, then only include the primary language code.

---

<sup>1</sup> See <https://tools.ietf.org/html/rfc5646>.

<sup>2</sup> For a full list of all valid language codes, see <http://www.loc.gov/standards/iso639-2>.

**NOTE:** If it is important to further clarify the particular spoken language (as is often the case with the macro-language Chinese, “zh”), then add a suitable sub-tag to create a compound language tag (“zh-cmn” for Mandarin Chinese).

### Visual

If visual languages are important to a work’s narrative, such as when there is no significant audible language in the work, then code the most common or contextually important visual (generally, written) language(s) in the work as presented during the work’s original release.

**NOTE:** In many cases, such works qualify as silent films and should be coded as described in *EIDR: Interim Best Practice – Silent Films*.

**NOTE:** In rare cases, the script used to represent a language may be an important distinguishing characteristic, for example “sr-Latn” vs. “sr-Cyrl” (Serbian written in Latin script vs. Serbian written in Cyrillic). These script codes are only valid for written languages, and even then should be used only when necessary.

### Version Language

A single work may have more than one language variant. In the EIDR system, these are recorded at the Edit and Manifestation levels. (In some cases, a Clip may be so specific that it warrants separate version language coding.) These language variants may be identified using Audio and/or Visual Version Language codes that differ from the Original Language codes. These are coded following the same practices as for Original Language.

### Title, Alternate Title, Description

Identify the language used in the text field, irrespective of the audio or visual language(s) that may appear within the referenced work. For fanciful words that are not part of a specific language (such as *Jumanji*) or foreign words that have been borrowed into a language (such as *Ronin* into English), code the language based on the primary language of the work’s original release (in both of these cases, “en” for English).

**NOTE:** If the characters used in the text field are not part of the standard script for that language, it may be helpful to identify the actual script used by appending a script tag to the language code. For example, use “ja-Hani” for Japanese expressed in Kanji characters. While it is valid to do so, it is not necessary to identify the script when a language has been transliterated into the Latin script (known as “Romanized”), since they can be identified by inspection. For example, Japanese transliterated into Latin script would just be “ja” rather than “ja-Latn”. This has the benefits of being both easier and shorter. However, if it is important to identify the transliteration (for example, to distinguish otherwise similar subtitle tracks), then do so.

### Constructing an EIDR Language Code

EIDR language codes follow this general pattern:

language[-extended\_language][-script][-region][-variant]

Where each sub-tag in square brackets is optional, but if present, separated from the preceding tag(s) by a hyphen. Each sub-tag may consist of a mix of letters or number, but no whitespace or punctuation (other than the separating hyphens) is allowed. RFC 5646 allows a more complex language code structure, but EIDR best practice is to limit the codes to the format provided.

Possible language codes include:

- en – English
- fr-BE – The dialect of French common to Belgium
- fr-015 – The dialect of French commonly spoken across North Africa
- ja-Kana – Japanese written with Katakana script
- ja-Latn-hepburn – Japanese transliterated into the Latin alphabet using the Hepburn Romanization standard
- sgn-qmm – Mongolian Sign Language
- zh – Chinese
- zh-cmn – Mandarin Chinese
- sr-Latn – Serbian written in Latin script
- sr-Cyrl – Serbian written in Cyrillic script

The full list of sub-tags allowed by 5646 is maintained by IANA (Internet Assigned Numbers Authority) and can be found at [www.iana.org/assignments/language-subtag-registry](http://www.iana.org/assignments/language-subtag-registry). It consists of selected elements drawn from a mix of code lists maintained by ISO (International Standards Organization), the UN (United Nations), and IANA itself.

## Language Tags

The shortest and most preferred language code consists solely of a language tag, such as en, de, and fr. All other sub-tags are optional.

Generally, a language tag is a 2-letter ISO 639-1 code, but it may also be a 3-letter 639-2, -3, or -5 code if a suitable 639-1 code is not available. See Common Languages, below, for a list of common EIDR language tags. All language tags are taken from the 639 family of ISO standards, “Codes for the representation of names of languages.” By convention, language tags are represented in all lower-case letters.

## Extended Language Tags

Certain language families (or macro-languages) are further clarified using extended language tags. These are 3-letter ISO 639-3 codes. The RFC 5646 standard practice is to use the shortest possible representation for any given language, which encourages the use of the 3-letter extended language tags on their own. However, Section 4.1.2 of RFC 5646 provides an exception that allows for 2-letter primary language prefixes in front of extended language tags. This is the format used in EIDR to assist in language family grouping and de-duplication. See Common Languages and Sign Languages, below, for a list of common EIDR extended language tags. By convention, extended language tags are represented in all lower-case letters.

For example, Mandarin Chinese can be represented using the 3-letter ISO 639-3 code `cmn` by itself. However, if this is coded with the ISO 639-1 2-letter Chinese macro-language prefix (“zh”), the result is `zh-cmn`. The advantage for EIDR of this coding is that all of the Chinese variants can be identified easily (they all begin with “zh”) and if one party has coded a work `zh` (generically Chinese), the EIDR system can readily compare that to works coded `zh-cmn` (Mandarin) or `zh-yue` (Cantonese) when looking for a duplicate match.

## Script Tags

Script tags are 4-letter codes drawn from ISO 15924, "Information and documentation – Codes for the representation of names of scripts." See Common Scripts, below, for a list of common EIDR script codes. By convention, they are represented with an initial capital letter, such as `Cyrl` (Cyrillic). Most languages are associated with a single script for their native written form, so it is not necessary to identify the script with which a visual language component may have been written. For example, with English, there is nothing to be gained by specifying `en-Latn`. In fact, doing so will only complicate de-duplication (and is a violation of international standard practice).

**NOTE:** In certain cases, a written language may be transliterated into a different language’s native script without any attempt at translation. In these cases, the transliterated script is not in a new language, but is simply a phonetic representation of the original language. For example, the Greek word “δὲ” can be transliterated into Latin script as “de” (`el-Latn`) or translated into English as “moreover” (`en`). If the transliterated material stands alone, it can be coded as the source language in an alternate script. However, if it is mixed in as a lesser element within material in a different language, then it should not be considered when determining the language code. The surrounding language should be identified instead.

## Region Tags

Region tags allow for the specification of regional variants (generally dialects) of different languages, such as `de-AT` (the dialect of German spoken in Austria) and `es-419` (the form of Spanish that is generally understood throughout Latin America). In general, region tags are only used when it is important to distinguish particular language variations. For example, the `fr-CA` (French as spoken in Canada) version of a work may be distinct from the `fr-FR` (French as spoken in France) version. It is unlikely that there would ever be a reason to distinguish between `en-GB` and `en-US` (United Kingdom and United States versions of English) for audio language, but due to their spelling differences, it may be helpful to make the distinction for visual languages.

Region tags are 2-letter country codes drawn from ISO 3166-1, “Codes for the representation of names of countries and their subdivisions – Part 1: Country codes” or 3-digit region codes drawn from UN M.49, “Standard Country or Area Codes for Statistical Use.” See Common Languages, below, for a list of commonly occurring EIDR region tags and Special Dialects and Language Variants for exceptional cases. By convention, the ISO country codes are represented in all caps, such as `US`, `GB`, or `TW`, while the UN region codes are presented with leading zeros, from `001` (the world) to `061` (Polynesia).

## Variant Tags

Variant tags identify language or dialect variations that are not possible to code using extended language or region tags and are generally tied to a specific primary language. For example, in 1901 and 1996 there were major spelling revisions in the German language, so if it were important to distinguish between

versions of a work subtitled with pre-1996 spellings vs. a version subtitled with post-1996 spellings, those visual languages could be coded as de-1901 and de-1996, respectively. However, Spanish did not undergo a spelling revision in 1996, so es-1996 is not valid. Such language variants are highly unusual and should be used with caution. See Special Dialects and Language Variants, below, for a list of the variant tags used within EIDR.

Variant tags may be up to 8-characters long and may contain a mix of letters and numbers. If they begin with a letter, then they must be at least 5-characters long. If they begin with a number (including a leading zero), they must be at least 4-characters long. By convention, any letters in a variant tag are presented in lower case, such as sl-nedis (the Natisone or Nadiza dialect of Slovenian).

## Language Code Tables

### Common Languages

Code	Language Name	Notes
ab	Abkhazian	
aa	Afar	
af	Afrikaans	
sq	Albanian	
am	Amharic	
ar	Arabic	
arc	Aramaic	
hy	Armenian	
as	Assamese	
aus	Australian Languages	
ay	Aymara	
az	Azerbaijani	
bat	Baltic Languages	
ba	Bashkir	
eu	Basque	
be	Belarusian	
bn	Bengali	
dz	Bhutani	
bh	Bihari	
bi	Bislama	
bs	Bosnian	
br	Breton	
bg	Bulgarian	
my	Burmese	
km	Cambodian	
ca	Catalan; Valencian	
zh	Chinese	Use when the dialect is unknown.
zh-gan	Chinese, Gan	Special format as per RFC 5646 §4.1.2.
zh-hak	Chinese, Hakka	Special format as per RFC 5646 §4.1.2.
zh-cmn	Chinese, Mandarin	Special format as per RFC 5646 §4.1.2.
zh-hsn	Chinese, Xiang (Hunanese)	Special format as per RFC 5646 §4.1.2.

Code	Language Name	Notes
zh-min	Chinese, Min (Taiwanese)	Special format as per RFC 5646 §4.1.2.
zh-wuu	Chinese, Wu (Shanghaiese)	Special format as per RFC 5646 §4.1.2.
zh-yue	Chinese, Yue (Cantonese)	Special format as per RFC 5646 §4.1.2.
co	Corsican	
crp	Creoles; Pidgins	There are specific codes for many creoles and pidgins, but it is rarely necessary to distinguish them for the purposes of deduplication.
cpe	Creoles; Pidgins – English-based	
cpf	Creoles; Pidgins – French-based	
cpp	Creoles; Pidgins – Portuguese-based	
hr	Croatian	
cs	Czech	
da	Danish	
nl	Dutch; Flemish	Modern Dutch, since 1350.
egy	Egyptian, Ancient	
en	English	Modern English, since 1500.
eo	Esperanto	
et	Estonian	
fo	Faroese	
fj	Fiji	
fil	Filipino; Pilipino	
fi	Finnish	
fr	French	Modern French, since 1600.
fy	Frisian	
gd	Gaelic; Scottish Gaelic	
gl	Galician	
ka	Georgian	
de	German	Modern German, since 1500.
el	Greek	Modern Greek, since 1453.
gn	Guarani	
gu	Gujarati	
ht	Haitian; Haitian Creole	Could be confused with fr-HT or cpf.
ha	Hausa	
haw	Hawaiian	
he	Hebrew	Was “iw”.
hi	Hindi	
hmo	Hmong; Mong	
hu	Hungarian	
is	Icelandic	
id	Indonesian	Was “in”.
ia	Interlingua	
ie	Interlingue; Occidental	
iu	Inuktitut	
ik	Inupiaq	
ga	Irish	
it	Italian	
ja	Japanese	
jw	Javanese	

Code	Language Name	Notes
kl	Kalaallisut; Greenlandic	
kn	Kannada	
ks	Kashmiri	
kk	Kazakh	
rw	Kinyarwanda	
ky	Kirghiz	
rn	Kirundi	
ko	Korean	
ku	Kurdish	
lo	Lao	
la	Latin	
lv	Latvian	
ln	Lingala	
lt	Lithuanian	
mk	Macedonian	
mg	Malagasy	
ms	Malay	
ml	Malayalam	
mt	Maltese	
mi	Maori	
mr	Marathi	
mo	Moldavian	
mos	Mossi	
mn	Mongolian	
na	Nauru	
ne	Nepali	
no	Norwegian	
oc	Occitan	
or	Oriya	
om	Oromo	
pa	Panjabi; Punjabi	
fa	Persian; Farsi	
pl	Polish	
pt	Portuguese	
ps	Pushto; Pashto	
qu	Quechua	
rm	Rhaeto-Romance	
ro	Romanian; Moldavian; Moldovan	
rom	Romany	
ru	Russian	
sm	Samoan	
sg	Sangho	
sa	Sanskrit	
sco	Scots	Scottish Gaelic is “gd”.
gd	Scottish Gaelic	Scots is “sco”.
sr	Serbian	
hbs	Serbo-Croatian	Was “sh”.
st	Sesotho	



<b>Code</b>	<b>Language Name</b>	<b>Notes</b>
tn	Setswana	
sn	Shona	
sgn	Sign Language	Use when the type of sign language is unknown.
sd	Sindhi	
si	Sinhala; Sinhalese	
ss	Siswati	
sk	Slovak	
sl	Slovenian	
so	Somali	
es	Spanish; Castilian	
su	Sundanese	
sw	Swahili	
sv	Swedish	
tl	Tagalog	
ty	Tahitian	
tg	Tajik	
ta	Tamil	
tt	Tatar	
te	Telugu	
th	Thai	
bo	Tibetan	
ti	Tigrinya	
to	Tonga	
ts	Tsonga	
tr	Turkish	
tk	Turkmen	
tw	Twi	
ug	Uighur; Uyghur	
uk	Ukrainian	
ur	Urdu	
uz	Uzbek	
vi	Vietnamese	
vo	Volapuk	
wln	Walloon	
cy	Welsh	
wo	Wolof	
xh	Xhosa	
yi	Yiddish	Was “ji”.
yo	Yoruba	
za	Zhuang; Chuang	
zu	Zulu	
art	Artificial Languages	Use for made-up languages, such as Na’vi.
mul	Multiple	Multiple, unidentified languages are present.
und	Undetermined	Use only when linguistic content is clearly present, but it is impossible to determine the language.

Code	Language Name	Notes
zxx	No Linguistic Content	Only use as a visual language code to identify works that have no linguistic elements, audio or visual. Do not use as an audio language.

## Sign Languages

Code	Sign Language Name	Commonly Used In
sgn-ads	Adamorobe Sign Language	Ghana
sgn-sqk	Albanian Sign Language	Albania
sgn-asp	Algerian Sign Language	Algeria
sgn-ase	American Sign Language	USA, Canada
sgn-aed	Argentine Sign Language	Argentina
sgn-aen	Armenian Sign Language	Armenia
sgn-asw	Australian Aboriginal Sign Language	Australia
sgn-asf	Australian Sign Language	Australia
sgn-asq	Austrian Sign Language	Austria
sgn-bqf	Bali Sign Language	Indonesia, Java & Bali
sgn-bvs	Belgian-Flemish Sign Language	Belgium
sgn-bvs	Belgian-French Sign Language	Belgium
sgn-bvl	Bolivian Sign Language	Bolivia
sgn-bzs	Brazilian Sign Language	Brazil
sgn-bho	British Sign Language	United Kingdom
sgn-bqn	Bulgarian Sign Language	Bulgaria
sgn-csc	Catalonian Sign Language	Spain
sgn-cds	Chadian Sign Language	Chad
sgn-csg	Chilean Sign Language	Chile
sgn-csl	Chinese Sign Language	China
sgn-csn	Colombian Sign Language	Colombia
sgn-csr	Costa Rican Sign Language	Costa Rica
sgn-cse	Czech Sign Language	Czech Republic
sgn-dsl	Danish Sign Language	Denmark
sgn-dse	Dutch Sign Language	Netherlands
sgn-ecs	Ecuadorian Sign Language	Ecuador
sgn-esn	El Salvadoran Sign Language	El Salvador
sgn-esl	Eskimo Sign Language	Canada
sgn-eth	Ethiopian Sign Language	Ethiopia
sgn-fse	Finnish Sign Language	Finland
sgn-fcs	French Canadian Sign Language	Canada
sgn-fsl	French Sign Language	France
sgn-gsg	German Sign Language	Germany
sgn-gse	Ghanaian Sign Language	Ghana
sgn-gds	Ghandruk Sign Language	Nepal
sgn-gss	Greek Sign Language	Greece
sgn-gsm	Guatemalan Sign Language	Guatemala
sgn-hps	Hawai'i Pidgin Sign Language	USA
sgn-hk	Hong Kong Sign Language	Hong Kong
sgn-icl	Icelandic Sign Language	Iceland
sgn-inl	Indonesian Sign Language	Indonesia, Java & Bali
sgn-ins	Indopakistani Sign Language	India, Pakistan

<b>Code</b>	<b>Sign Language Name</b>	<b>Commonly Used In</b>
sgn-isg	Irish Sign Language	Ireland
sgn-isl	Israeli Sign Language	Israel
sgn-ise	Italian Sign Language	Italy
sgn-jcs	Jamaican Country Sign Language	Jamaica
Sgn-jls	Jamaican Sign Language	Jamaica
sgn-jsl	Japanese Sign Language	Japan
sgn-jos	Jordanian Sign Language	Jordan
sgn-xki	Kenyan Sign Language	Kenya
sgn-kvk	Korean Sign Language	Korea, South
sgn-kgi	Kuala Lumpur Sign Language	Malaysia, Peninsular
sgn-lsl	Latvian Sign Language	Latvia
sgn-lbs	Libyan Sign Language	Libya
sgn-lls	Lithuanian Sign Language	Lithuania
sgn-lsg	Lyons Sign Language	France
sgn-xml	Malaysian Sign Language	Malaysia, Peninsular
sgn-mdl	Maltese Sign Language	Malta
sgn-mre	Martha's Vineyard Sign Language	USA
sgn-lsy	Mauritian Sign Language	Mauritius
sgn-msd	Mayan Sign Language	Mexico
sgn-mfs	Mexican Sign Language	Mexico
sgn-mzg	Monastic Sign Language	Holy See
sgn-qmm	Mongolian Sign Language	Mongolia
sgn-xms	Moroccan Sign Language	Morocco
sgn-nbs	Namibian Sign Language	Namibia
sgn-nsp	Nepalese Sign Language	Nepal
sgn-nzs	New Zealand Sign Language	New Zealand
sgn-ncs	Nicaraguan Sign Language	Nicaragua
sgn-nsi	Nigerian Sign Language	Nigeria
sgn-nsl	Norwegian Sign Language	Norway
sgn-nsr	Nova Scotian Sign Language	Canada
sgn-okl	Old Kentish Sign Language	United Kingdom
sgn-pys	Paraguayan Sign Language	Paraguay
sgn-psg	Penang Sign Language	Malaysia, Peninsular
sgn-psc	Persian Sign Language	Iran
sgn-prl	Peruvian Sign Language	Peru
sgn-ppp	Philippine Sign Language	Philippines
sgn-psd	Plains Sign Talk	USA
sgn-pso	Polish Sign Language	Poland
sgn-psr	Portuguese Sign Language	Portugal
sgn-pro	Providencia Sign Language	Colombia
sgn-psl	Puerto Rican Sign Language	Puerto Rico
sgn-rsi	Rennellese Sign Language	Solomon Islands
sgn-rms	Romanian Sign Language	Romania
sgn-rsl	Russian Sign Language	Russia, Europe
sgn-sdl	Saudi Arabian Sign Language	Saudi Arabia
sgn-spf	Scandinavian Pidgin Sign Language	Sweden
sgn-sls	Singapore Sign Language	Singapore
sgn-svk	Slovakian Sign Language	Slovakia

<b>Code</b>	<b>Sign Language Name</b>	<b>Commonly Used In</b>
sgn-sfs	South African Sign Language	South Africa
sgn-ssp	Spanish Sign Language	Spain
sgn-sqs	Sri Lankan Sign Language	Sri Lanka
sgn-swl	Swedish Sign Language	Sweden
sgn-ssr	Swiss-French Sign Language	Switzerland
sgn-sgg	Swiss-German Sign Language	Switzerland
sgn-slf	Swiss-Italian Sign Language	Switzerland
sgn-tss	Taiwanese Sign Language	Taiwan
sgn-tza	Tanzanian Sign Language	Tanzania
sgn-tsq	Thai Sign Language	Thailand
sgn-tse	Tunisian Sign Language	Tunisia
sgn-tsm	Turkish Sign Language	Turkey
sgn-ugn	Ugandan Sign Language	Uganda
sgn-ukl	Ukrainian Sign Language	Ukraine
sgn-uks	Urubú-Kaapor Sign Language	Brazil
sgn-ugy	Uruguayan Sign Language	Uruguay
sgn-vsl	Venezuelan Sign Language	Venezuela
sgn-yds	Yiddish Sign Language	Israel
sgn-ysl	Yugoslavian Sign Language	Yugoslavia
sgn-zsl	Zambian Sign Language	Zambia
sgn-zib	Zimbabwe Sign Language	Zimbabwe

## Common Scripts

The standard script for a written language is never identified. For example, en-Latn (English written using the Latin script) is invalid. However, some languages are commonly written using more than one script. For example, it may be important to distinguish Mandarin written in traditional script (zh-cmn-Hant) from simplified script (zh-cmn-Hanz). Finally, it may be important to identify the fact that a language has been transliterated into another script. So, while el-Grek (Greek written in Greek) would be invalid, el-Latn (Greek transliterated into Latin script) could provide important distinguishing information.

<b>Code</b>	<b>Script</b>
Arab	Arabic
Brai	Braille
Cyrl	Cyrillic
Grek	Greek
Latn	Latin
Zmth	Mathematical Notation
Sgnw	SignWriting
Zsym	Symbols
Syrn	Syriac (Eastern variant)
Syre	Syriac (Estrangelo variant)
Syrj	Syriac (Western variant)
Visp	Visible Speech
Hant	Traditional Chinese
Hanz	Simplified Chinese
Hani	Han (Kanji; Kanji; Hanja)
Hira	Hiragana

Code	Script
Hang	Hangul (Hangŭl, Hangeul)
Kana	Katakana
Zyyy	Code for undetermined script

## Regions

### Numeric Region Codes

A regional dialect of a language can be identified by qualifying the language by the region in which the dialect is commonly spoken. For example, fr-029 would be Caribbean French, while es-419 would be Latin American Spanish.

Code	Region
001	World
002	Africa
014	Eastern Africa
017	Middle Africa
015	Northern Africa
018	Southern Africa
011	Western Africa
019	Americas
419	Latin America and the Caribbean
021	North America
029	Caribbean
013	Central America
021	Northern America
005	South America
142	Asia
143	Central Asia
030	Eastern Asia
034	Southern Asia
035	South-Eastern Asia
145	Western Asia
150	Europe
151	Eastern Europe
154	Northern Europe
039	Southern Europe
155	Western Europe
009	Oceania
053	Australia and New Zealand
054	Melanesia
057	Micronesia
061	Polynesia

### Country Codes

A country-specific dialect of a language can be identified by qualifying the language by the country in which the dialect is commonly spoken. For example, fr-CA would be Canadian French, while pt-BR would be Brazilian Portuguese.

<b>Code</b>	<b>Country</b>	<b>Notes</b>
AF	Afghanistan	
AX	Åland Islands	
AL	Albania	
DZ	Algeria	
AS	American Samoa	
AD	Andorra	
AO	Angola	
AI	Anguilla	AI was French Afar and Issas.
AQ	Antarctica	
AG	Antigua and Barbuda	
AR	Argentina	
AM	Armenia	
AW	Aruba	
AU	Australia	
AT	Austria	
AZ	Azerbaijan	
BS	Bahamas	
BH	Bahrain	
BD	Bangladesh	
BB	Barbados	
BY	Belarus	
BE	Belgium	
BZ	Belize	
BJ	Benin	Was Dahomey (DY).
BM	Bermuda	
BT	Bhutan	
AN	Netherlands Antilles	
BO	Bolivia	
BQ	Bonaire, Sint Eustatius, and Saba	BQ was British Antarctic Territory.
BA	Bosnia and Herzegovina	
BW	Botswana	
BV	Bouvet Island	
BR	Brazil	
IO	British Indian Ocean Territory	
BN	Brunei	
BG	Bulgaria	
BF	Burkina Faso	Was Upper Volta (HV).
BI	Burundi	
KH	Cambodia	
CM	Cameroon	
CA	Canada	
CV	Cape Verde	
KY	Cayman Islands	
CF	Central African Republic	
TD	Chad	
CL	Chile	
CN	China	
CX	Christmas Island	

<b>Code</b>	<b>Country</b>	<b>Notes</b>
CC	Cocos (Keeling) Islands	
CO	Colombia	
KM	Comoros	
CG	Congo	
CD	Congo, The Democratic Republic of the	Was Zaire (ZR).
CK	Cook Islands	
CR	Costa Rica	
CI	Côte d'Ivoire	
HR	Croatia	
CU	Cuba	
CW	Curaçao	
CY	Cyprus	
CZ	Czech Republic	
DK	Denmark	
DJ	Djibouti	Was French Afar and Issas (AI).
DM	Dominica	
DO	Dominican Republic	
EC	Ecuador	
EG	Egypt	
SV	El Salvador	
GQ	Equatorial Guinea	
ER	Eritrea	
EE	Estonia	
ET	Ethiopia	
FK	Falkland Islands (Malvinas)	
FO	Faroe Islands	
FJ	Fiji	
FI	Finland	
FR	France	
GF	French Guiana	
PF	French Polynesia	
TF	French Southern Territories	
GA	Gabon	
GM	Gambia	
GE	Georgia	GE was Gilbert and Ellice Islands.
DE	Germany	Used for West Germany prior to 1990.
GH	Ghana	
GI	Gibraltar	
GR	Greece	
GL	Greenland	
GD	Grenada	
GP	Guadeloupe	
GU	Guam	
GT	Guatemala	
GG	Guernsey	
GN	Guinea	
GW	Guinea-Bissau	
GY	Guyana	

<b>Code</b>	<b>Country</b>	<b>Notes</b>
HT	Haiti	
HM	Heard Island and McDonald Islands	
VA	Holy See	Vatican City.
HN	Honduras	
HK	Hong Kong	
HU	Hungary	
IS	Iceland	
IN	India	
ID	Indonesia	
IR	Iran, Islamic Republic of	
IQ	Iraq	
IE	Ireland	
IM	Isle of Man	
IL	Israel	
IT	Italy	
JM	Jamaica	
JP	Japan	
JE	Jersey	
JO	Jordan	
KZ	Kazakhstan	
KE	Kenya	
KI	Kiribati	
KP	Korea, Democratic People's Republic of	North Korea.
KR	Korea, Republic of	South Korea.
KW	Kuwait	
KG	Kyrgyzstan	
LA	Lao People's Democratic Republic	Laos.
LV	Latvia	
LB	Lebanon	
LS	Lesotho	
LR	Liberia	
LY	Libya	
LI	Liechtenstein	
LT	Lithuania	
LU	Luxembourg	
MO	Macao	
MK	Macedonia, the former Yugoslav Republic of	Republic of Macedonia.
MG	Madagascar	
MW	Malawi	
MY	Malaysia	
MV	Maldives	
ML	Mali	
MT	Malta	
MH	Marshall Islands	
MQ	Martinique	
MR	Mauritania	
MU	Mauritius	
YT	Mayotte	



<b>Code</b>	<b>Country</b>	<b>Notes</b>
MX	Mexico	
FM	Micronesia, Federated States of	
MD	Moldova, Republic of	
MC	Monaco	
MN	Mongolia	
ME	Montenegro	
MS	Montserrat	
MA	Morocco	
MZ	Mozambique	
MM	Myanmar	Was Burma (BU).
NA	Namibia	
NR	Nauru	
NP	Nepal	
NL	Netherlands	
NC	New Caledonia	
NZ	New Zealand	
NI	Nicaragua	
NE	Niger	
NG	Nigeria	
NU	Niue	
NF	Norfolk Island	
MP	Northern Mariana Islands	
NO	Norway	
OM	Oman	
PK	Pakistan	
PW	Palau	
PS	Palestinian Territory, Occupied	
PA	Panama	
PG	Papua New Guinea	
PY	Paraguay	
PE	Peru	
PH	Philippines	
PN	Pitcairn	
PL	Poland	
PT	Portugal	
PR	Puerto Rico	
QA	Qatar	
RE	Réunion	
RO	Romania	
RU	Russian Federation	
RW	Rwanda	
BL	Saint Barthélemy	
SH	Saint Helena, Ascension and Tristan da Cunha	
KN	Saint Kitts and Nevis	
LC	Saint Lucia	
MF	Saint Martin (French part)	The Dutch part is SX.
PM	Saint Pierre and Miquelon	

<b>Code</b>	<b>Country</b>	<b>Notes</b>
VC	Saint Vincent and the Grenadines	
WS	Samoa	
SM	San Marino	
ST	Sao Tome and Principe	
SA	Saudi Arabia	
SN	Senegal	
RS	Serbia	Russia.
SC	Seychelles	
SL	Sierra Leone	
SG	Singapore	
SX	Sint Maarten (Dutch part)	The French part is MF.
SK	Slovakia	SK was Sikkim.
SI	Slovenia	
SB	Solomon Islands	
SO	Somalia	
ZA	South Africa	
GS	South Georgia and the South Sandwich Islands	
SS	South Sudan	
ES	Spain	
LK	Sri Lanka	
SD	Sudan	
SR	Suriname	
SJ	Svalbard and Jan Mayen	
SZ	Swaziland	
SE	Sweden	
CH	Switzerland	
SY	Syrian Arab Republic	Syria.
TW	Taiwan	
TJ	Tajikistan	
TZ	Tanzania, United Republic of	
TH	Thailand	
TL	Timor-Leste	Was East Timor (TP).
TG	Togo	
TK	Tokelau	
TO	Tonga	
TT	Trinidad and Tobago	
TN	Tunisia	
TR	Turkey	
TM	Turkmenistan	
TC	Turks and Caicos Islands	
TV	Tuvalu	
UG	Uganda	
UA	Ukraine	
AE	United Arab Emirates	
GB	United Kingdom	Do not use UK.
US	United States	
UM	United States Minor Outlying Islands	

Code	Country	Notes
UY	Uruguay	
UZ	Uzbekistan	
VU	Vanuatu	Was New Hebrides (NH)
VE	Venezuela, Bolivarian Republic of	
VN	Viet Nam	
VG	Virgin Islands, British	
VI	Virgin Islands, U.S.	
WF	Wallis and Futuna	
EH	Western Sahara	
YE	Yemen	Used for North Yemen prior to 1990.
ZM	Zambia	
ZW	Zimbabwe	Was Southern Rhodesia (RH).

## Special Dialects and Language Variants

### Signed Spoken Languages

Established sign languages have their own ISO 639-3 codes, listed in Sign Languages, above. Codes for signed versions of spoken languages can be constructed by adding a spoken language and optional region code to the generic sign language code (sgn), such as sgn-eng for a signed version of spoken English, or sgn-eng-US for a signed version of the US dialect of spoken English.

**NOTE:** In the example above, English is represented using its ISO 639-2 three-letter language code, rather than its more common ISO 639-1 two-letter code, because when constructing the compound language tag, it appears in the extended language slot rather than the language slot. This same pattern can be seen in the use of zh-cmn to identify Mandarin Chinese.

### Language Variants

Code	Description	Limitations
1959acad	Academic (governmental) Belarusian codified in 1959.	Used with be (Belarusian) language.
1996	Standard German orthography codified in 1996.	Use with de (German) language.
alalc97	Romanization (transliteration into the Latin script) according to American Library Association and the Library of Congress standards.	Follows a Latn (Latin) script tag.
emodeng	Early Modern English – the variant of English used in the King James <i>Bible</i> .	Used with en (English) language.
fonipa	International Phonetic Alphabet	Acts as a Script tag
hepburn	Hepburn romanization of Japanese.	Use with ja-Latn (Japanese in Latin script).
heploc	Library of Congress variant of Hepburn romanization of Japanese.	Use with ja-Latn (Japanese in Latin script).
hognorsk	Norwegian in Høgnorsk (High Norwegian) orthography	Use with nn (Norwegian) language.
jyutping	Jyutping romanization of Cantonese.	Use with zh-yue-Latn (Cantonese in Latin script).

<b>Code</b>	<b>Description</b>	<b>Limitations</b>
pinyin	Pinyin Romanization of Chinese or Tibetan.	Use with zh* (Chinese or a Chinese dialect) or bo (Tibetan) and Latn (Latin script) such as zh-cmn-Latn-pinyin or bo-Latn-pinyin.
scotland	Scottish Standard English	Use with en (English) language.
ulster	Ulster dialect of scots	Use with sco (Scots) language.
valencia	The dialect of Catalan spoken in the Comunidad Valenciana region of Spain.	Use with ca (Catalan) language.

**NOTE:** Language variants are unlikely to be used except in academic situations where linguistic experts are drawing fine distinctions between similar language representations.