Language Tagging in the Rough

NRSI, SIL International

What are Language Tags?

Language Tag: Language

lng

Language component

ISO 639 2-letter or 3 letter
qaa-qtz private use

und - Unknown language

Examples
en
kyu
pce

Language Tag: Script

lng-Scpt

Script component

ISO 15924

Qaaa-Qaai private use

Zyyy unknown script

Zxxx for unwritten, e.g. audio

Examples
en
kyu-Kali
pce-Thai

Language Tag: Region

lng-Scpt-RE

Country component

ISO 3166-1

Examples
en-GB
kyu-Kali-TH
pce-Thai

Language Tag: Variants

lng-Scpt-RE-variant

Variants

Language (dialects)
Script
Orthography
      For mature orthographies
IETF registered

Examples
en-fonipa
kyu-Kali-TH
pce-Thai-rulai

Language Tag: Extensions

lng-Scpt-RE-variant-e-xtension

Extension
namespace-values
-x-   PUA

Examples
en-Zxxx-x-audio
kyu-Kali-TH
pce-Thai-x-rulai

Resources

BCP 47

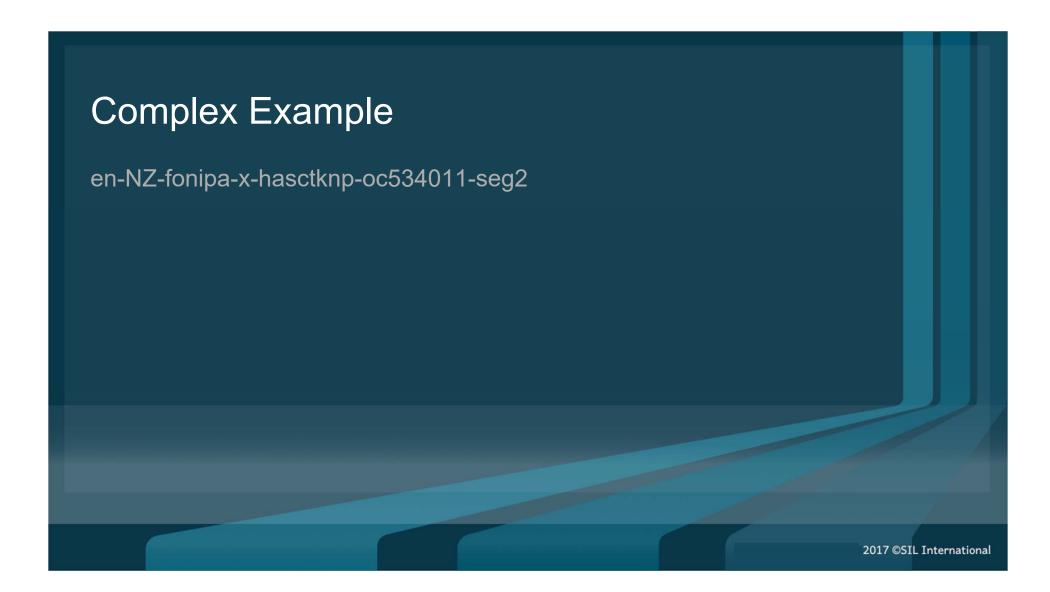ISO 639-1, 639-3, 15924, 3166-1

IANA Registry

Tag Wisely!

Only include the elements you need to distinguish

Drop region if there is no variation

Drop script if there is no variation

Agree on defaults

Many under differentiated tags and bad tags

Complex Example

en-NZ-fonipa-x-hasctknp-oc534011-seg2

Why Language Tags?

Uses of Language Tags

Controls text interaction in applications

Removes ambiguity

Locating texts on the Internet

Language needs assessments

Gives a language group a digital presence

How to LanguageTag

Tagging a Writing System

Dialects can be mutually unintelligible!

Try for iso 639-3 code first

Register any variants

Use PUA codes as interim only

Tagging can be very political

Iso 639-3 Registration Process

Two related varieties are normally considered **varieties** of the **same language** if speakers of each variety have inherent understanding of the other variety (that is, can **understand** based on knowledge of their own variety **without needing to learn the other variety**) at a functional level.

Where spoken intelligibility between varieties is marginal, the existence of a **common literature** or of a **common ethnolinguistic identity** with a central variety that both understand can be strong indicators that they should nevertheless be considered **varieties** of the same language.

Where there is enough intelligibility between varieties to enable communication, the existence of well-established **distinct ethnolinguistic identities** can be a strong indicator that they should nevertheless be considered to be **different languages**.

IETF Subtag Registration Process

Read BCP 47

Submit to: mailto://ietf-languages@iana.org

Open list anyone can join. But it's busy.

Response within 2 weeks

LANGUAGE SUBTAG REGISTRATION FORM
 1. Name of requester:
 2. E-mail address of requester:
 3. Record Requested:

   Type:
   Subtag:
   Description:
   Prefix:
   Preferred-Value:
   Deprecated:
   Suppress-Script:
   Macrolanguage:
   Comments:

 4. Intended meaning of the subtag:
 5. Reference to published description
    of the language (book or article):
 6. Any other relevant information:

Establishing Defaults

Need a default region

       Many locale categories are based on region

Default script only if very clear

Language Tag List

Language Tag

Likely Subtag

	A	B	C	D	E	F	G	H	I	J	K
1	langtag	likely_subtag	LangName	script	regions	default_region	is_unwritten	areas	ethnlg_provided	confirmed	comments
2	crj	crj-Cans-CA	Cree, Southern East	Cans	CA	CA		NAM		CLDR	
3	crj-Latn	crj-Latn-CA	Cree, Southern East	Latn	CA	CA		NAM	1		
4	crk	crk-Cans-CA	Cree, Plains	Cans	CA US	CA		NAM		CLDR	
5	crk-Latn	crk-Latn-CA	Cree, Plains	Latn	CA US	CA		NAM			
6	crl	crl-Cans-CA	Cree, Northern East	Cans	CA	CA		NAM		CLDR	
7	crl-Latn	crl-Latn-CA	Cree, Northern East	Latn	CA	CA		NAM	1		
8	kyu-Kali	kyu-Kali-MM	Kayah, Western	Kali	MM	MM		SEA	1		
9	kyu-Latn	kyu-Latn-MM	Kayah, Western	Latn	MM	MM		SEA	1		
10	kyu-Mymr	kyu-Mymr-MM	Kayah, Western	Mymr	MM	MM		SEA	1		
11	kyw-Beng	kyw-Beng-IN	Kudmali	Beng	IN	IN		SAS	1		
12	kyw-Deva	kyw-Deva-IN	Kudmali	Deva	IN	IN		SAS	1		
13	kyw-Orya	kyw-Orya-IN	Kudmali	Orya	IN	IN		SAS	1		
14	lhu	lhu-Latn-CN	Lahu	Latn	CN LA MM TH VN	CN		EAS SEA	1		
15	lll	lll-Zyyy-PG	Lilau		PG	PG	1	MEL	1		
16	lmg	lmg-Latn-PG	Lamogai	Latn	PG	PG		MEL	1		
17	mnc-Mong-x-manchu	mnc-Mong-CN-x-manchu	Manchu	Mong	CN	CN		EAS	1		
18	mnc-Mong-x-oldmanch	mnc-Mong-CN-x-oldmanchu	Manchu	Mong	CN	CN	1	EAS	1		
19											

Language Tags List

LangTag List   https://goo.gl/tI4waT

Want to invite someone?   https://goo.gl/forms/jBP5IJdTdOx5bViu2