

# Language tagging using the new RFC 5646



A word cloud of language tags in various colors and sizes. The tags include: en-scouse, en-gb, mas, sl-it-nedis, es-419, de, zh-Hant-HK, zh-Hans, zh-yue, de-CH-1901, az-Latn, and fr.

Richard Ishida

W3C Internationalization Activity Lead

Language tags

Before

RFC 1766 / RFC 3066

language region

- ◆ ISO 639 language codes
- ◆ ISO 3166 country codes

en

en-GB

en-scouse

Language tags

Now

**BCP 47** (includes RFC 5646)

language script region variant extension private\_use

- ◆ only language subtag required – use other tags only when needed!

az-Cyrl

- ◆ codes available only from new IANA Language Subtag Registry

zh-Hant-HK

- ◆ incorporates ISO and UN codes

es-419

- ◆ over 7,000 codes

sl-IT-nedis

Language tags

# Extended language subtags (extlangs)

**BCP 47** (includes RFC 5646)

language script region variant extension private\_use



extended-language

ar

kok

ms

sw

uz

zh

sgn



ar-aa0

ar-afb

ar-arq

ar-arz

etc...



Better as...

aa0

afb

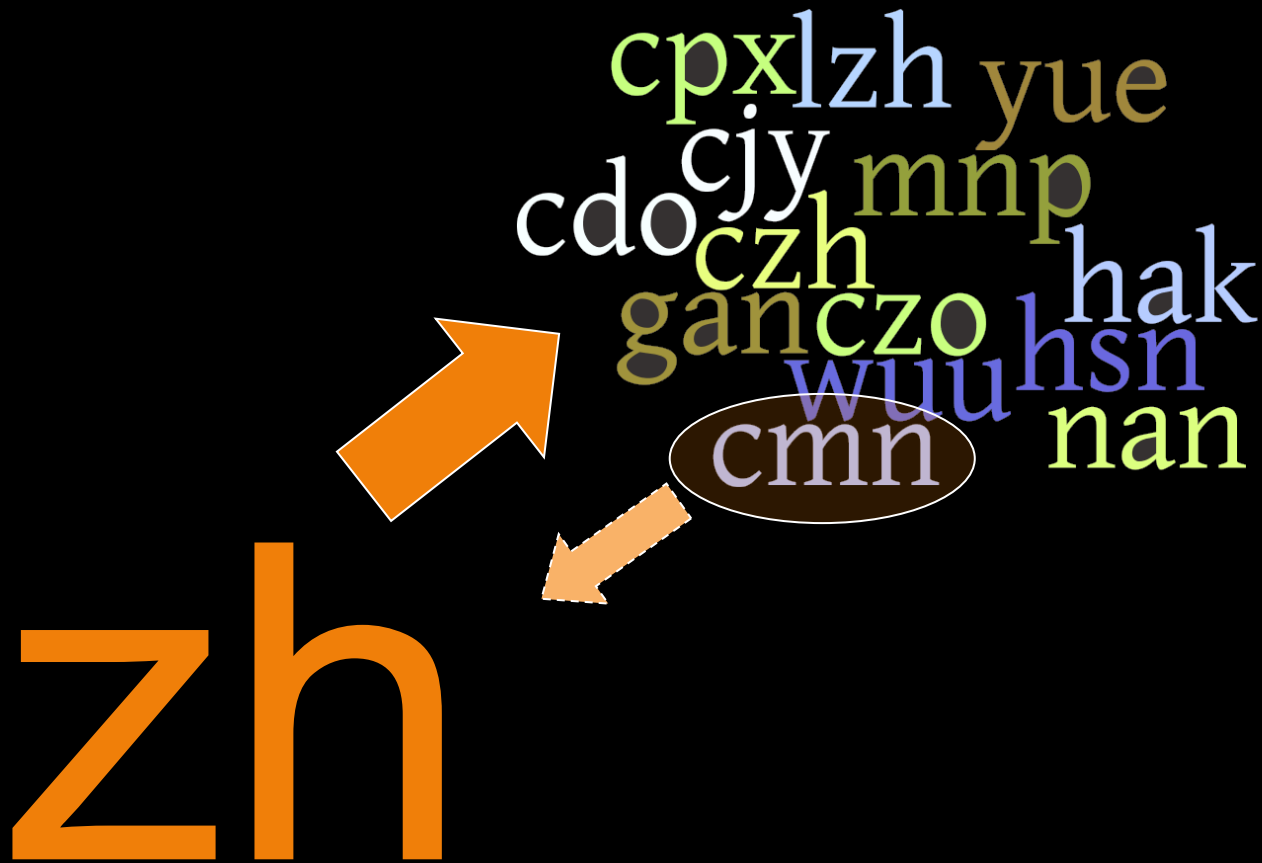
arq

arz

etc...

Language tags

Macrolanguages, collections, etc.



### Use

- ◆ Supported by a community of experts at the IETF
- ◆ Registry is regularly updated through a published process allowing submissions
- ◆ Language tags are mostly simple, but can be flexed to meet more complex needs
- ◆ Extension mechanism can provide mappings to other systems
- ◆ Accompanied by a language tag matching specification

Language tags

Further reading



- ◆ BCP 47  
<http://www.rfc-editor.org/rfc/bcp/bcp47.txt>
- ◆ IANA Language Subtag Registry  
<http://www.iana.org/assignments/language-subtag-registry>
- ◆ Subtag Lookup Tool  
<http://rshida.net/utls/subtags/>
- ◆ Choosing a Language Tag  
<http://www.w3.org/International/questions/qa-choosing-language-tags>