Language tagging using the new RFC 5646
Language tags

Before

- ISO 639 language codes
- ISO 3166 country codes

RFC 1766 / RFC 3066

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!
- codes available only from new IANA Language Subtag Registry
- incorporates ISO and UN codes
- over 7,000 codes

az-Cyrl
zh-Hant-HK
es-419
sl-IT-nedis
Language tags

Extended language subtags (extlangs)

Language – Script – Region – Variant – Extension – Private Use

BCP 47 (includes RFC 5646)

ar-aao
ar-afb
ar-arq
ar-arz
etc...

Better as...

aao
afb
arq
arz
etc...
Language tags
Macrolanguages, collections, etc.
Language tags

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://rishida.net/utils/subtags/
- Choosing a Language Tag
  http://www.w3.org/International/questions/qa-choosing-language-tags