Language Tagging in the Rough
NRSI, SIL International
What are Language Tags?
**Language Tag: Language**

<table>
<thead>
<tr>
<th>Language component</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>ISO 639 2-letter or 3 letter</td>
<td>en</td>
</tr>
<tr>
<td>qaa-qtz private use</td>
<td>kyu</td>
</tr>
<tr>
<td>und - Unknown language</td>
<td>pce</td>
</tr>
</tbody>
</table>

Ing
Language Tag: Script

Ing-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

Ing-Scpt-RE

Country component
ISO 3166-1

Examples
en-GB
kyu-Kali-TH
pce-Thai
# Language Tag: Variants

**lng-Scpt-RE-variant**

<table>
<thead>
<tr>
<th>Variants</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language (dialects)</td>
<td>en-fonipa</td>
</tr>
<tr>
<td>Script</td>
<td>kyu-Kali-TH</td>
</tr>
<tr>
<td>Orthography</td>
<td>pce-Thai-rulai</td>
</tr>
</tbody>
</table>

For mature orthographies

IETF registered
Language Tag: Extensions

Ing-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
<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>langtag</td>
<td>likely_subtag</td>
<td>LangName</td>
<td>script</td>
<td>regions</td>
<td>default_region</td>
<td>is_unwrittten</td>
<td>areas</td>
<td>ethnicity_provided</td>
<td>confirmed</td>
</tr>
<tr>
<td>2</td>
<td>crj</td>
<td>crj-Cans-CA</td>
<td>Cree, Southern East</td>
<td>Cans</td>
<td>CA</td>
<td>CA</td>
<td>NAM</td>
<td>CLDR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>crj-Latin</td>
<td>crj-Latin-CA</td>
<td>Cree, Southern East</td>
<td>Latin</td>
<td>CA</td>
<td>CA</td>
<td>NAM</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>crk</td>
<td>crk-Cans-CA</td>
<td>Cree, Plains</td>
<td>Cans</td>
<td>CA US</td>
<td>CA</td>
<td>NAM</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>crk-Latin</td>
<td>crk-Latin-CA</td>
<td>Cree, Plains</td>
<td>Latin</td>
<td>CA US</td>
<td>CA</td>
<td>NAM</td>
<td>CLDR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>crl</td>
<td>crl-Cans-CA</td>
<td>Cree, Northern East</td>
<td>Cans</td>
<td>CA</td>
<td>CA</td>
<td>NAM</td>
<td>CLDR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>crl-Latin</td>
<td>crl-Latin-CA</td>
<td>Cree, Northern East</td>
<td>Latin</td>
<td>CA</td>
<td>CA</td>
<td>NAM</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>kyu-Kali</td>
<td>kyu-Kali-MM</td>
<td>Kayah, Western</td>
<td>Kali</td>
<td>MM</td>
<td>MM</td>
<td>SEA</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>kyu-Latin</td>
<td>kyu-Latin-MM</td>
<td>Kayah, Western</td>
<td>Latin</td>
<td>MM</td>
<td>MM</td>
<td>SEA</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>kyu-Mymr</td>
<td>kyu-Mymr-MM</td>
<td>Kayah, Western</td>
<td>Mymr</td>
<td>MM</td>
<td>MM</td>
<td>SEA</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>kyw-Beng</td>
<td>kyw-Beng-IN</td>
<td>Kudmali</td>
<td>Beng</td>
<td>IN</td>
<td>IN</td>
<td>SAS</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>kyw-Deva</td>
<td>kyw-Deva-IN</td>
<td>Kudmali</td>
<td>Deva</td>
<td>IN</td>
<td>IN</td>
<td>SAS</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>kyw-Orya</td>
<td>kyw-Orya-IN</td>
<td>Kudmali</td>
<td>Orya</td>
<td>IN</td>
<td>IN</td>
<td>SAS</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>lhu</td>
<td>lhu-Latin-CN</td>
<td>Lahu</td>
<td>Latin</td>
<td>CN LA MM TH VN</td>
<td>CN</td>
<td>EAS</td>
<td>SAS</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>lll</td>
<td>lll-Zzyy-PG</td>
<td>Lilau</td>
<td>PG</td>
<td>PG</td>
<td>1</td>
<td>MEL</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>img</td>
<td>img-Latin-PG</td>
<td>Lamogai</td>
<td>Latin</td>
<td>PG</td>
<td>PG</td>
<td>MEL</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>mnc-Mong-x-manchu</td>
<td>mnc-Mong-CN-x-manchu</td>
<td>Manchu</td>
<td>Mong</td>
<td>CN</td>
<td>CN</td>
<td>EAS</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>mnc-Mong-x-oldmanchu</td>
<td>mnc-Mong-CN-x-oldmanchu</td>
<td>Manchu</td>
<td>Mong</td>
<td>CN</td>
<td>CN</td>
<td>EAS</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Language Tags List

LangTag List

Want to invite someone?  

https://goo.gl/tl4waT

https://goo.gl/forms/jBP5IjTdoX5bViu2