Decision Tree Discussion on Data Provisioning for Script Tags

Action Items:
- engage expert for advice on:
    - Unicode
    - Language and script subtags

Issue: Do different versions of Unicode need to be specified?

JG: policy could state a minimum version of Unicode to use
- but is Unicode backwards compatible?
- code versions will change over time
- do code versions delete code points rather than add?

SH: Unicode stable and code points are only added (but double check)
- a minimum version of Unicode should support later versions
- see RFC 7940

Issue: Is a subtag an override or complement to automated script detection?

JG: we should either support a standard or not
FA: don't need to specify this if one overrides the other
SH: Han script can be "overridden" to be called Japanese
JG: re: override: can't invent things that don't exist
- can only go with what's in the Unicode table
SH: how do you choose language with many scripts? Han = Japanese, Chinese, Korean
FA: RFC 5646 subtag shouldn't contradict automated Unicode script detection
JG: script tag must be valid for code points in use, but don't have to specify script property
- **script tag, if specified, must be valid for all code points**

Issue: A Registry cannot alter what script information is sent to them by a registrar

- This is generally accepted practice/status quo
- FA: A registry must validate data given to them and if data not valid, only then will an issue arise
    - JG: **Registry can perform a kind of syntactic validation: check that script tags match with Unicode points** (check script tag against data that's presented)

- if script tag does not match unicode points, registration on hold until resolved
    - RC: this is more than syntactic validation

Issue: Private subtags should not be allowed

Recall: language tags are a sequence of subtags

FA: [reading RFC 5646] private use subtags used in private agreements, and not intended for general use

JG: these could be useful for providing additional information

- problem with use of word "subtag" here: the script tag should not be private use (nor should language), but allow for the private use fields available in RFC 5646 guidelines
- allow for potentially valuable private subtags
- **but private subtags should only complement language and script tags provided under the policy we create**
- don't want everything being private use, but we should allow for other information to be entered once language and script entered into registration data system

[from follow-up email discussion]

JG:

Adding to this, the ABNF looks like this:
Language-Tag = langtag ; normal language tags
/ privateuse ; private use tag
/ grandfathered ; grandfathered tags
langtag = language
["-" script]
["-" region]
*("-" variant)
*("-" extension)
["-" privateuse]
Note that "privateuse" appears in two places: at the "root" of the ABNF and in the "langtag" rule.
What I'm saying **is that "privateuse" in the expansion of "Language-Tag" should be invalid. However, "privateuse" in the expansion of "langtag" is valid.**

RC:

As far as the discussion on privateuse, I wonder if we are confusing two topics: privateuse as an alternate language-tag; and the privateuse sub-tag.

As I read RFC 5646 these are two distinct concepts. I was unable to locate the specifics of the alternates (langtag, privateuse, grandfathered) so I am inferring what is meant by this identifiers.

I think that we all were agreeing that the idea of a privateuse language-tag was inappropriate for our use.

I think that Jim and I were trying to state that the use of the privateuse sub-tag within the langtag alternative was acceptable.

I know that we have not discussed language yet but for illustrative purposes: a Language-tag "x-priv" would not be ok but Language-tag "en-Latn-x-priv" could be acceptable.