IDN Implementation Guidelines (IDNG) Working Group (WG)

Notes from Meeting on 16 Dec. 2015

<u>Meeting Attendees</u>

WG members (in alphabetical order):

1. Chris Dillon
2. Christian Dawson
3. Dennis Tan Tanaka
4. Jian Chang
5. Kal Feher
6. Mats Dufberg

Staff:
7. Sarmad Hussain

<u>Regrets</u>

None

<u>Meeting Notes</u>

1. **Variants.** Variant discussion started with reference to primary vs. other variants. Primary variant could be interpreted as the label applied-for by use or alternatively if a script considers a particular version as canonical. There should be clarity on such definitions. In another case, primary variant is automatically allocated to the registrant and secondary variants are not. Thus, there is a behavior where some variant(s) are activated automatically whereas others are not automatically activated. There is another kind of variants which are never activated. In the case of Chinese, some make such distinctions but some gTLDs do not make this distinction. Chinese, Japanese and Korean communities are working together to define a collective set for the top-level. The group should look at this work for providing guidance and possibly suggest similar consideration at the second level, keeping in mind that there may be script vs. language level differences. Implementers, for interoperability, follow the community based work e.g. LGR. LGR may also have a smaller set for Root Zone than what is needed for the second level. But second level does not need to follow same set but should follow similar principles to deal with variants.

   For Armenian variants are for preventing cross-script homoglyphs. In Arabic, there are multiple kinds of variants, including homoglyphs, different writing styles, "simplified" spellings, etc.

   When developing tables, we should recommend that the implementers should consider the variants being developed by the relevant communities through LGR and other work.

2. **Automatically Activating Variants.** The group discussed that this may not be relevant for the guidelines and leave it to the individual registry. Historical practice for China and Taiwan has been that if simplified Chinese label is applied, traditional is automatically activated, and vice

versa.   In the future the practice may be that a particular label is registered and its variants are reserved, with the primary concern to avoid phishing issues. So such decisions should be left to registry which may not require automatic activation.   Registries would generally follow consistent approach but we should not restrict them to a single mechanism to allow them to meet their individual requirements.

3. **Variants at Second Level Should be Allocated to Same Registrant.** There is an expectation or understanding that this may be the case, but this has not been clearly specified anywhere. In the ccTLD space, there is an example, where the variants are initially offered to the registrant but if the original registrant does not reserve them, they are made openly available to other registrants. The group generally agreed that this is a good requirement to consider, but should be worded in a way that it allows registries to update their language tables (allowing some leeway in such exceptional cases). This may also be possible through the dispute process.

4. **Implication of Variants on TMCH.** SAC 60 reports discusses this issue, on whether registries should provide all variants to TMCH or TMCH should generate it. There has been some discussion that there are no obligations on registries and TMCH should use the IDN tables provided by the registries. Best if TMCH makes the call on how marks and their variants should be interpreted. However, if the IDN tables are under-defined for variants, this would cause a challenge for TMCH. This could be resolved once the LGR format is adopted. Thus, TMCH is not relevant for the IDN Guidelines.

5. **Similarity/Confusability and Upper/Lower Case.** This includes within-script confusability and cross-script confusability. The group should say that the implementers should keep up to date from guidelines from IAB and other relevant script and language communities in the context of labels. Georgian has three cases, upper, lower and neutral case, where lower case and neutral case are PVALID. The discussion was not completed and would be continued next time.

The group decided to continue the discussion and meet on 13 January 2016.

Appendix A: Issues list discussed

1. Transition and Terminology
    a. Any (residual) issues/items from IDNA2003 to IDNA2008 transition?
    b. Glossary of terms around LGRs and variants from the IDN TLD project and other work
2. Language table format and managing consistency of end-user experience
    a. IDN Table format in XML, based on new LGR specifications being developed
    b. Role of reference second level tables in managing consistency and differences across TLDs for a predictable end-user experience
    c. Relationship between language tables and script tables? Other categories (some languages but not entire script)?
    d. Managing consistence across levels; relationship of Root Zone LGR and the second level IDN tables?
3. Variants
    a. Variant states (number and nomenclature) and state-change mechanisms (e.g. blocked, allocatable, allocated, reserved, etc.)
    b. Primary vs. secondary variants in Chinese language (other scripts or languages?)
    c. Policy for activating variants
        i. Automatic or registrant requested allocation
            1. Automatic activation of variant labels at second level for Chinese language domain names
        ii. Minimum and Maximum number of allocated variants/Ceiling value?
        iii. Choosing which variants may be activated
    d. Variants at second level must(?) be allocated to the same registrant?
    e. Implication of second level variants on TMCH
4. Similarity/Confusability of labels
    a. Homographic issues within script
        i. General homographic issues within a script
        ii. Scope of confusability: upper to lower case mapping
    b. Cross-script homoglyphs management to prevent phishing possibilities– cox.com, where "cox" can be in Latin or Greek or Cyrillic
        i. Script mixing within a second level label
        ii. Script mixing across levels
        iii. Scope of confusability: upper to lower case mapping
        iv. Is ASCII a special case for mixing?
5. Registration data
    a. Registration data for IDNs
    b. Registration data of variants – information regarding variant sets and variant label disposition(s)