

IDN Implementation Guidelines (IDNG) Working Group (WG)

Notes from Meeting on 15 December, 2016

Meeting Attendees (in alphabetical order)

WG members:

1. Dennis Tanaka
2. Jian Zhang
3. Kal Feher
4. Mats Dufberg

Staff:

5. Sarmad Hussain

Meeting Notes

The WG members continued the discussion on the document and the recommendations to be proposed for public comment. Updated document *IDN Guidelines 4.0 20161215* was discussed.

- 1. Authoritative sources for homoglyphs.** The WG group looked at the *confusables.txt* and *intentional.txt* files from the Unicode consortium's TR39 which had been filtered to exclude code points disallowed by the IDNA 2008 protocol. It was noted that *confusables.txt* includes similar code points comprehensively across a diverse set of scripts but not limited to homoglyphs. The WG also noted that *intentional.txt* is limited to homoglyphs but only contains code points from Cyrillic, Greek and Latin scripts and does not include other scripts, e.g. Arabic script. The WG discussed the possibility of recommending registries to look at *confusables.txt* and sifting through it to find the relevant subset themselves, however, it was concluded that such a recommendation would not be actionable for the registries, and even if they did consider it they would likely end up having different sets, which will not address the user-confusion but contribute further to it. An alternative discussed was to say that *intentional.txt* should be used as a minimum and then additional conditions, as needed, be added to make up for the limitations of *intentional.txt*.

As another option, the WG also considered developing a relevant subset of *confusables.txt* as a reference. In this context, the WG discussed who would undertake this work. The members agreed that ICANN may not be the best organization and that this should be taken up either at IETF or Unicode consortium. They noted that IETF had looked at similar problems in RFC 5646 and RFC 5992. However, this may not be in the control of ICANN. The members noted that recommendation 5 from version 3.0 is relevant to this discussion and should be added back to the current draft, and that term "homoglyph" should be defined. One option is to use "visually confusable" term instead of homoglyph because it is already used in previous version of the guidelines and the community is familiar with it, but that may make the set of cases much wider. The members discussed what would be the best way forward. Asking ICANN to undertake another study to define this set of homoglyphs was considered, but not recommended as it is not a very large problem today (due to limited cases of phishing in IDNs) and that there is already a lot relevant work going on at ICANN for the root zone, which is being done by script communities, is transparent and reasonably comprehensive across the script code point sets,

has a very rigorous review process, and therefore should be leveraged (even if in some cases it may be slightly less comprehensive than needed for the second level).

The group noted that the root zone work is still on-going. So the guidelines should include how this evolving root zone work should be incorporated by the registries. The WG group was informed that root zone LGR-1 is already out, containing Arabic script. Also, Georgian, Lao, Khmer and Thai proposals have been finalized by the community and will be considered for the second version of LGR, being targeted for release around ICANN 58. And then LGR-3 containing many more script (e.g. Chinese, Japanese and Korean) may be available in early 2018. Further, due to stability constraints on LGR for the root zone, it provides a firm reference, even as it grows in its coverage over time.

The WG agreed that data from intentional.txt and root zone LGR work must be used. And further, the data from confusables.txt should be considered (but not as a “must”). The recommendation should be drafted in this context.

- 2. Next meeting.** The WG agreed to meet on 22 December. Action items from previous week, except AI1 and AI3 which has been addressed, will be carried to next week for discussion.

Action Items

S. No.	Action Items	Owner
1	<i>Update the recommendation on harmonization for the registries adhering to data from intentional.txt and root zone LGR as a “must” and also encourage registries to look at confusables.txt (but latter not as a “must”)</i>	MD
2	<i>Re-write the recommendation on automatic activation based on the current input for further discussion</i>	EC
3	<i>Add recommendation 5 from version 3.0 of the guidelines for further discussion and inclusion in the current version</i>	SH