

IDN Implementation Guidelines (IDNG) Working Group (WG)

Notes from Meeting on 2 Dec. 2015

Meeting Attendees

WG members (in alphabetical order):

1. Dennis Tan Tanka
2. Chris Dillon
3. Jian Chang
4. Edmon Chung
5. Mats Dufberg
6. Ram Mohan
7. Wael Nasr

Staff:

8. Alireza Saleh
9. Sarmad Hussain

Regrets

None

Meeting Notes

1. **String Contention.** Need for analysis for string similarity and related contention for different scripts was raised. This may be applicable both at top-level and second level. The group discussed the issue and determined that this may not be specifically IDN related. If IDN specific aspect of this issue can be identified, the WG would discuss this further.
2. **Issue list for IDN Guidelines.** The issue list in Appendix A, circulated on the email, was discussed, as per the details below.
3. **Transition to IDNA2008.** It was suggested that the guidelines should specify that IDNA2008 would be followed. The members shared that many tools are still based on IDNA2003. It was concluded to look at this issue, first to gauge how much progress has been made, and then to decide how to make the recommendation in the current version of the Guidelines.
4. **Transition across Unicode Versions.** The members noted that the Unicode version is also getting updated and that is an important consideration in the discussion around IDNs. It was noted that though IDNA2008 is Unicode version agnostic, it has not been upgraded due to issues with some characters as identified in the IAB statement¹. The group concluded to have a section on IDNA and Unicode versions.

¹ <https://www.iab.org/documents/correspondence-reports-documents/2015-2/iab-statement-on-identifiers-and-unicode-7-0-0/>

5. **Glossary of Relevant Terms.** Glossary is important but should be produced as the discussions proceed. The WG agreed that staff should start with an initial list of terms on the wiki and the WG members can also contribute. Initially only the terms should be identified and then, towards the end of this work, they be defined and included in the Guidelines.

Staff to create an initial glossary of terms on the wiki. All members to contribute to this list of terms as the work progresses.

6. **Language Table Format.** The members noted that the new Label Generation Rules (LGR) format is well defined and contains all the relevant information, like variant information and contextual rules. It was recommended that the registries be required to publish their IDN tables in this format, if the RFC for the LGR format is finalized while these Guidelines are being formulated. The members noted that this would be one of the main items in the current revision, and the Guidelines must require this format. In this context, the WG should also elaborate and clarify differences between language vs. script based IDN tables.
7. **Variant Management.** It was explained that currently there is no general reference on what the possible states of a variant label are, how these states are defined, and how a variant may transition from one state to another. For example, it is not clear what is a “blocked” variant, and if such a variant can be unblocked? The WG agreed that the types of variants should be defined either as part of the glossary or as a recommendation – to be determined later. The WG also agreed to discuss the variant state management. For example, the WG could suggest that changing a blocked state may require change in LGR. Such cases should be identified and WG should decide which ones to comment on. Some level of variant management should be done by the registries offering IDNs with variants.

The group also considered that allocation of variants to the same registrant was a gray area as this may not be directly related to IDNs and the group may come back to it, if needed. It was pointed out that in some cases, e.g. through a dispute, some variants had to be dis-allocated from a registrant.

It was noted that allocatable variants may be determined by the LGR directly so the issue in 3.c.iii may be redundant. However, it was further explained that this issue was raised for the contexts in which the number of allocatable variants generated by the LGR may be very large and additional management mechanism may be needed.

Due to limitation of time, the group decided to continue the discussion in two weeks on Wednesday, 16 Dec.

Action Items

S.No.	Action	Owner
1	<i>Staff to create an initial glossary of terms on the wiki. All members to contribute to this list of terms as the work progresses.</i>	<i>Staff</i>

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)