

Staff Report of Public Comment Proceeding

Reference Label Generation Rulesets (LGRs) for the Second Level

Publication Date: 15 September 2016

Prepared By: Sarmad Hussain

Public Comment Proceeding

Open Date:	7 Jun 2016
Close Date:	1 August 2016
Staff Report Due Date:	15 September 2016

Important Information Links

Announcement
Public Comment Proceeding
View Comments Submitted

Staff Contact: Sarmad Hussain

Email: sarmad.hussain@icann.org

Section I: General Overview and Next Steps

During the New gTLD Program's Pre-Delegation Testing (PDT), ICANN has noted a large volume of IDN table submissions for review. ICANN also continues to receive requests to add additional language support in the delegated TLDs through the Registry Services Evaluation Policy (RSEP) process. The registries are generally encouraged to collaborate in defining common language-based or script-based IDN tables to allow for consistency for end users. For this reason, the IDN tables used by each gTLD and some ccTLDs are posted at [the IANA Repository for IDN Practices](#). To further promote consistency of these submissions, and to simplify the related PDT and the RSEP, ICANN has undertaken to develop and publish reference second level LGRs. The process to develop these reference LGRs, as detailed in the [Guidelines](#) [PDF, 207 KB], ensures both linguistic and technical expert input with finalization based on community review through a formal public comment process. This work builds on the [baseline IDN tables](#) already being used in PDT.

The proposed process for [Evaluation of Deviation from the Reference Second Level LGRs](#) [PDF, 255 KB] will be updated based on the feedback received, as discussed below. Then the reference second level LGRs will be integrated into the operations, replacing the [baseline IDN tables](#) already being used.

Section II: Contributors

At the time this report was prepared, a total of three (3) community submissions had been posted to the forum. The contributors, both individuals and organizations/groups, are listed below in chronological order by posting date with initials noted. To the extent that quotations are used in the foregoing narrative (Section III), such citations will reference the contributor's initials.

Organizations and Groups:

Name	Submitted by	Initials
Registries Stakeholder Group	Paul Diaz, Chairman RySG	RySG
Internet Architecture Board	Andrew Sullivan, Chair IAB	IAB

Individuals:

Name	Affiliation (if provided)	Initials
Yoshiro Yoneya		YY

Section III: Summary of Comments

General Disclaimer: This section intends to summarize broadly and comprehensively the comments submitted to this public comment proceeding but does not address every specific position stated by each contributor. The preparer recommends that readers interested in specific aspects of any of the summarized comments, or the full context of others, refer directly to the specific contributions at the link referenced above (View Comments Submitted).

ICANN thanks RySG, IAB and YY for their comments.

RySG notes that the proposed process remains “overly restrictive and makes the reference LGRs de facto requirements for future IDN implementations.” RySG suggests that this work “should not limit the range of implementation options for gTLD registries unless clear issues of security and stability are identified” and that RySG does not believe that there is contractual basis for ICANN to approve or deny changes to IDN tables. Thus, they suggest that:

- [RYSG1] Justification of deviation from ref. LGRs should not be required from registries. ICANN should consult with registry operator to determine rationale and ICANN should point out any security and stability issues.
- [RYSG2] The only basis for ICANN to consider a submission incomplete is if IDN table is not well formed.
- [RYSG3] RySG is unaware how having a smaller number of code points pose a security or stability risk; so it is unnecessary for ICANN to evaluate tables that are purely a subset of the reference LGRs.
- [RYSG4] In cases where registry operator has been approved to offer IDNs in a specific language/script, ICANN does not have contractual basis to reject changes to relevant IDN tables (reference to Section 1.4 of Spec. 6). So the relevant portions of the process should be advisory only.

RySG suggest that these changes bring ICANN in an advisory role which allows it to share its language expertise with registries, which is “both advantageous for internet users as well as consistent with ICANN’s Registry agreements.” The changes will foster healthy competition while safeguarding security and stability of the internet for its users.

[IAB1] IAB reiterates from RFC 5894 and 6912 that registries should restrict registrations to code points which are well-understood and the desirability of more-restrictive rules closer to the root. Therefore, if a registry selects a smaller subset of the repertoire for these reasons, it should not be considered deviant and be allowed.

[IAB2] IAB further adds that for public zones the decision to include a code point should be conservative (as given in RFC 6912). It notes that for Belarusian, U+02BC is permitted, which has been discussed as problematic under the principle of conservatism.

[IAB3] IAB says that associating a language with LGRs is also problematic because DNS labels are not in a language. IAB states that it “understands the reason why these evaluations proceed by language: writing systems are often used to write multiple languages, and a character in a script that is appropriate for one language may be inappropriate or even dangerous in the context of another.”

[IAB4] IAB adds that “Yet there does not appear to be the kind of guidance in these LGRs that would be necessary in order to put them together successfully”, and as registry operator will know correct code points in the context of user population, ICANN should not permit more code points than they would require.

IAB concludes by encouraging ICANN to continue to provide evaluation of code points, but not impose inclusivity conditions on subordinate zones.

[YY1] YY suggests some specific editorial comments. In addition, YY suggests that core set of code points, variant and WLE rules “must be referred to” but not required to be supported.

[YY2] Finally, YY gives specific suggestions about the LGR for Japanese. This includes removing the context rule for U+3006.

[YY3] And also changing the context rule for U+30FC or have a consistent definition with current PDT specification.

Section IV: Analysis of Comments

General Disclaimer: This section intends to provide an analysis and evaluation of the comments submitted along with explanations regarding the basis for any recommendations provided within the analysis.

[BACKGROUND] Need to review language-based or script-based IDN tables during the Pre-Delegation Testing (PDT) originates from the Applicant Guidebook (AGB). It is stated in section 1.3.2 IDN Tables of the AGB that “IDN tables must also be submitted for each language or script in which the applicant intends to offer IDN registrations at the second or lower levels ... including specification of any variant characters ... Developing tables and registration policies in collaboration with relevant stakeholders to address common issues ... An applicant’s IDN tables should help guard against user confusion in the deployment of IDN gTLDs.”

AGB further states that “Applicants will be asked to describe the process used in developing the IDN tables submitted. ICANN may compare an applicant’s IDN table with IDN tables for the same languages or scripts that already exist in the IANA repository or have been otherwise submitted to ICANN. If there are inconsistencies that have not been explained in the application, ICANN may ask the applicant to detail the rationale for differences,” concluding that “ICANN will accept the applicant’s IDN tables based on the factors above.”

[RYSG1 & 2] For the PDT, as per the AGB, ICANN may compare the submitted IDN tables for a language/script and “ask the applicant to detail the rationale for differences.” Where there is a rationale, with no security and stability implications, the changes will be accepted. See [RYSG3] and [IAB1] for further discussion regarding smaller repertoire.

[RYSG3] It should be noted that the reference second level LGRs define a set of code points along with relevant variants and label rules. Creating a subset of the code point repertoire may make it incoherent with the corresponding variant sets and label rules. Thus, though conservatism encourages smaller repertoire of code points, the process should be undertaken in a way which holistically addresses the corresponding implications on variant sets and label rules. See also [IAB1] for further discussion and resolution.

Further, it should also be noted that due to stability concerns, it may not be straight forward to add a code point back to the repertoire in the future if, for example, the addition creates a variant set which makes existing non-variant labels, registered in the interim, to become variants of each other.

[RYSG4] Section 2.1 of the 2014 Registry Agreement asks for a material change to registry services, such as domain name registration, to go through the RSEP to evaluate its security, stability and competition effects. The RSEP and approval process aim to create an environment that encourages gTLD registry operators to discuss any changes that may impact third parties with ICANN before they are made. As discussed, changes in content of IDN tables may have an adverse security and stability impact. Under the RSEP, ICANN will evaluate security and stability effects, and, we agree that, where there is no significant impact, the changes will be able to proceed. As has been pointed out in Section 1.4 of Spec. 6, any such revisions to IDN tables should be re-published at IANA Repository.

Based on RySG feedback, the process described in “Evaluation of Deviation from the Second Level LGR References” will be updated as per the discussion above.

[IAB1] As discussed in [RYSG3] above, the reference second level LGRs define a coherent set of code points, variants and label rules. A subset can be created by removing code points, variants and/or label level rules. Creating an arbitrary subset may make the code point repertoire and corresponding variant sets and label level rules incoherent. We agree with the overriding requirement of conservatism, as pointed out by IAB. To address the comment, the suggested evaluation process will be updated to explicitly state that a registry may include a repertoire which is a subset of a reference second level LGR, and this will not be considered a deviation. In the cases where the difference from reference LGR is limited to only the code points which are not part of the any variant set or label rule, the registry can proceed. However, in other cases, the registry should carefully review the changes to keep the repertoire, variant sets and label level rules coherent with each other. Such LGRs will be checked for security and stability issues, e.g. which may result by not including a variant rule or a context rule.

[IAB2] Recognizing the security issue concerning U+02BC, a special context rule had been used: "apostrophe-modifier-disallowed" disallowing the code point in either first or last position of a label to mitigate the security risk. But based on IAB feedback, the code point and the related context rule will be removed.

[IAB3] As given in the background information earlier in this section, the reference to a language originates from the AGB. To address that concern, it is useful to reiterate the statement made in the "Guidelines for Developing Reference LGRs for the Second Level": "The fact that something can be written in a particular language, or even looked up in its dictionary, does not imply an entitlement to have that string appear in the DNS." As such, there is no requirement in the reference LGR or a subset to fully cover a given language, but again a LGR subset must be coherent. To clarify it further, the statement from the Guidelines document will also be added to process the "Evaluation of Deviation from the Second Level LGR References."

[IAB4] There is some relevant text in the "Reference Label Generations Rules (LGR) for the Second Level - Overview and Summary" in section 2.2: "The language-based reference LGRs provided here could be used as "building blocks" in assembling local, regional or script-based LGRs. When used in that fashion, care must be taken that the resulting LGR provides for a consistent treatment of variants, for example." This text will be merged with additional text to be added in the Overview document to make clearer what is implicated in the process for subsetting, extending or merging reference LGRs.

[YY1] The reference to root zone LGR will be considered and will be added if considered relevant. Reference to LAGER specification will be updated to RFC 7940. In the sentence "*Current reference LGRs include a core set of code points, variant rules, and Whole Label Evaluation (WLE) rules that must be supported and also include an (optional) extended ruleset serving special needs based on geographical or other variations.*", the "must be supported" cannot be replaced by "must be referred to" because all these elements are normative in a LGR. Further, all references to JIS X 208-1990 will be replaced by JIS X 0208:1997.

[YY2] By using the Unicode "ScriptExtensions" property, it becomes unnecessary to maintain a context rule for U+3006, so it will be removed from the Japanese LGR.

[YY3] Concerning U+30FC HIRAGANA-KATAKANA PROLONGED SOUND MARK, the reference LGR will be updated to a context rule "preceded-by-Japanese" ("Japanese" interpreted as Kanji, Hiragana, and Katakana). This imposes a reasonable restriction on the use of the code point, among others, preventing it from occurring at the beginning of a label. Because of its resemblance to punctuation, a conservative approach motivates some restrictions on the occurrence of this code point. rules.

Additional summary: It should also be noted that, in light of the absence of any feedback concerning the set of code points marked as "excluded code points" in relevant 2nd level LGRs, all reference to these code points will be removed from the final versions of these LGRs. (For the XML files, that includes removing the context rule named "excluded-cp").