

Guidelines for Developing Reference Label Generation Rules (LGRs) for The Second Level Version 2

Publication Date: 26 May 2020

Prepared By: Sarmad Hussain

Public Comment Proceeding

Open Date:	30 March 2020
Close Date:	12 May 2020
Staff Report Due Date:	25 May 2020 (extended from 18 May 2020)

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

ICANN org had published [reference Label Generation Rules \(LGRs\)](#) for multiple languages to facilitate the security and stability of Internationalized Domain Name (IDN) operations of generic top-level domain (gTLD) registries and improve transparency and consistency of testing IDN tables. The [existing guidelines](#) used for these reference IDN tables focus on developing language-based references LGRs.

Since then, the detailed analyses of multiple scripts have been completed for the [Root Zone Label Generation Rules \(RZ-LGR\)](#) by the relevant communities. These analyses can now be used for creating reference second-level LGRs for these scripts. Therefore, the [Guidelines for Developing Reference LGRs for the Second Level](#) are being updated to also include details of how to design script-based reference LGRs, in addition to language-based reference LGRs.

These updated guidelines will be finalized after incorporating the input from the community. Using these guidelines, ICANN org will continue to develop the IDN tables for additional languages and scripts and post these for Public Comment.

Section II: Contributors

At the time this report was prepared, a total of four (4) 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
gTLD Registries Stakeholder Group	Samantha Demetriou	RySG

Individuals:

Name	Affiliation (if provided)	Initials
Bill Jouris		BJ1
Bill Jouris		BJ2
Bill Jouris		BJ3

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).

BJ1:

The reference for top-level domain (TLD) (<https://www.icann.org/resources/pages/variant-tlds-2012-05-08-en>) returns an error indicating that the page does not exist.

BJ2:

Section 5.1 paragraph 1 (page 9) notes that second level domains (SLDs) can include a variety of code points not allowed in TLDs, for example digits. These were not considered when identifying variants for TLDs. How will variants involving these code points be identified?

BJ3:

Section 6, paragraph 1 (page 11) states: "[TLD] provides a definition of variant as well as suggests which variants are appropriate, and [TLD] adds additional information. The results of these projects can be seen as authoritative for script-based LGRs." But for TLDs an extremely narrow definition of "variants" is used -- one which includes only those cases which not even a tiny fraction of the user population will notice. A rather larger group is classified as "confusables". For them, a Similarity Review Panel is envisioned to conduct a manual review process. Is it planned that something similar will be done with SLDs? Or will the "confusables" from the TLD project be included as "variants" here?

RySG:

RySG welcomes the further development of reference LGRs for the second level, by ICANN and the community, acknowledging that the combined knowledge of how to use scripts of the

world in the DNS is a valuable asset. The RySG notes that script based LGRs for the root zone are designed to be conservative; however, LGRs for lower levels could relax some of those rules that better meet the needs of its users.

RySG further states that, although a registry operator, as a contracted party with ICANN, does want consistency and predictability in its operations and obligations which reference LGRs may bring, the RySG observes that there is not a contractual basis to require registry operators to adopt these reference LGRs during [Registry Services Evaluation Policy \(RSEP\)](#) and [Registry System Testing \(RST\)](#) procedures. Registry operators have the prerogative to determine the registration policies that best suit the market realities, within applicable law and contractual obligations.

Therefore, RySG welcomes these good practices and looks forward to working with ICANN and the community to develop a process for adopting the anticipated reference LGRs within the multistakeholder, consensus process.

In addition, RySG suggests errata for the document.

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.

ICANN org thanks BJ and RySG for their valuable input and feedback.

The errata in BJ1 and RySG comments will be addressed in the updated version of the Guidelines document which is to be published.

BJ2. Yes, adding code points can impact existing variant sets and label-level rules. Therefore, any code points added in the reference LGRs in addition to the repertoire already analyzed for RZ-LGR, including digits, will be carefully considered in this context. The resulting draft reference LGRs will be published for Public Comment for input from the community for further confirmation before these are finalized.

BJ3. Variant code points are those which are considered “same” by the script community. The variant code points can be motivated by a variety of reasons based on the script, as have been documented by the community-based panels in their work on the Root-Zone Label Generation Rules (RZ-LGR). Visual “sameness” or indistinguishability is one of these reasons. However, not all code points which are considered visually “similar” may qualify to be variants of each other, so cannot be addressed through IDN tables. As noted by BJ, visually “similar” cases are handled by the string similarity review process for TLDs. For the second level, other policies such as dispute resolution policies, may be used to mitigate against abusive registrations exploiting such visually similar characters.

RySG. [RFC 6912](#) provides a consistent set of principles for inclusion of code points for forming labels in the DNS, including second-level and top-level domains, with one salient difference that the top-level labels will be alphabetic. As the same set of principles are used

for developing RZ-LGR by the community, the work for root zone forms a linguistically acceptable, secure and stable baseline for the reference script-based IDN tables for the second level after including digits and hyphen.

It is not ICANN org's intention to require registry operators (ROs) to adopt these reference IDN tables. These reference IDN tables are intended to be used to bring consistency and transparency in the IDN table review during RSEP and RST processes. ROs may utilize reference IDN tables and may decide to change code points, variants or rules included in these baseline reference IDN tables, while developing the IDN tables which ROs would like to implement to suit the market needs. Such differences will be reviewed and if any of these changes are found to cause security or stability issues, these will be addressed in consultation with the relevant RO.

ICANN org also looks forward to working with RySG and the community for adopting the additional reference LGRs in order to help make the IDN table review processes more consistent and transparent for ROs.