

## Registries Stakeholder Group Statement

### Public Comment: **Reference Label Generation Rulesets (LGRs) for the Second Level**

Date statement submitted: **14 October 2020**

Reference url: <https://www.icann.org/public-comments/reference-lgrs-second-level-2020-08-24-en>

#### Background<sup>1</sup>

To improve the transparency and consistency of the Internationalized Domain Name(IDN) table review process to facilitate the registry operations of new gTLDs, ICANN has developed additional reference IDN tables in machine-readable format, called reference Label Generation Rulesets (LGRs) for the second level. The reference IDN tables are based on the [Guidelines for Developing Reference Label Generation Rules \(LGRs\)](#), which were finalized after [community review](#). These reference LGRs will be used in reviewing IDN tables submitted by the gTLD registries, e.g. through the Registry Service Evaluation Policy (RSEP) process.

Earlier RySG comments on this or related issues:

[RySG Comment on Guidelines for Developing Reference Label Generation Rules \(LGRs\) for the Second Level Version 2](#) (May 2020)

---

### Registries Stakeholder Group comment

The Registries Stakeholder Group (RySG) welcomes the opportunity to provide feedback on the *Reference Label Generation Rulesets (LGRs) for the Second Level* and wishes to submit the following comments:

1. The RySG is concerned that disproportionate burden on IDN implementation could ultimately stifle adoption by registries which ultimately limits distribution by registrars and other registration channels.
  - a. Simple and practical rules help promote and improve the implementation and adoption of IDNs by registries and other parts of the ecosystem; in contrast, complex requirements can actually do the opposite by decreasing investment and stifling growth. According to the IDN World Report Internationalized Domain

---

<sup>1</sup> *Background: intended to give a brief context for the comment and to highlight what is most relevant for RO's in the subject document – it is not a summary of the subject document.*

Names (IDNs) account for 2.5% of total domain names across gTLDs and ccTLDs<sup>2</sup>; despite the fact that about two thirds of the world population is a native speaker of a non-Latin-based language<sup>3</sup>. By making IDN implementation more complex (and more expensive) it discourages investment in such systems.

- b. The number of rules added by these reference LGRs – beyond the normative requirements of IDNA2008 – range from 3 to 14 rules per LGR (i.e. whole-label evaluation and context rules). There are more than 90 new rules combined in 17 reference LGRs. The RySG respectfully asks: what severe security and stability issues are these rules solving for? And why now? IDNs have been around for many years now without posing severe security and stability issues<sup>4</sup>. The Malayalam LGR contains 14 of these rules compared to zero in the Chinese Hani LGR. Therefore, we would like to understand the basis for developing reference LGRs that greatly deviate from the status quo.

2. ICANN by creating and enforcing policy within contractual agreements is working outside of the bottom up consensus building policy process.

- a. ICANN staff has advised the RySG that ICANN org has further developed the reference LGRs for the second-level i) to provide predictability and transparency to the IDN table review process, and ii) to address security and stability issues with IDN implementation. To this end, ICANN org will be utilizing “technical design” and “linguistic considerations” from work products that include RFC 6912 and the Root Zone Label Generation Ruleset program (RZ-LGR).
- b. As contracted parties the RySG appreciates the effort in improving predictability and consistency in administrative processes, such as the review of IDN tables.
- c. This being said, the RySG notes that contractual requirements regarding IDN implementation are clearly stipulated in the registry agreement and in the ICANN IDN Guidelines. Conformance to ICANN reference LGRs for the second-level can be encouraged but should not be required, because the work product was not the result of consensus policy. If ICANN org or the at-large ICANN community believes these issues need to be discussed, the RySG looks forward to working with ICANN org, and other stakeholders, to further develop the

---

<sup>2</sup> <https://idnworldreport.eu/>, accessed on 29 September 2020.

<sup>3</sup> [https://en.wikipedia.org/wiki/List\\_of\\_languages\\_by\\_number\\_of\\_native\\_speakers](https://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers), <https://www.internetworldstats.com/stats.htm>, accessed on 2 October 2020.

<sup>4</sup> “Data continues to show that the unique characteristics of Internationalized Domain Names (IDNs) are not being used to facilitate phishing in meaningful numbers”, Global Phishing Survey 2016: Trends and Domain Name Use, Page 25, URL: [https://docs.apwg.org/reports/APWG\\_Global\\_Phishing\\_Report\\_2015-2016.pdf](https://docs.apwg.org/reports/APWG_Global_Phishing_Report_2015-2016.pdf)

operational framework for IDN implementation in accordance with contractual and consensus policies.

3. ICANN policies for registries should not be expected to solve for hardware and/or software limitations which registry operators cannot control.
  - a. A reference LGR uses the principle that “the end user should be able to reach a website connected to his/her domain name regardless of location [...] input devices (language table) that the user may use to reach a domain name should be carefully considered when defining variants.” The RySG believes this feature (e.g. international reachability<sup>5</sup>), while a novel feature, it is something that should not be required, or expected, from registry operators to solve for because it can’t be hardly scalable across language tables. “International reachability” is not a problem specific to the Arabic Language, since the same issue can be found in other languages (e.g. French, German or Spanish users using an English-only keyboard).
  - b. The RySG opines that this could be a potential feature a registry operator might want to offer to address its market’s needs, but it should not be expected or required to be implemented.
4. The domain name system should not need to conform to the “nature”<sup>6</sup> of languages or scripts.
  - a. Domain names can be reasonable and memorable mnemonics without imposing complex rules that attempt to replicate the use of a language or script i.e. syllable structure of scripts.
  - b. Spelling preferences (e.g. “correct ways to write [words]”, Reference LGR for Arabic Language) can’t be expected to be mechanically implemented as a determination for variant relationship in a consistent manner across several LGRs. This practice is something a registry operator might want to offer as a value-added service, but can’t be expected to be implemented on security and stability grounds.

---

<sup>5</sup> “For example, if someone registered the domain name “مكة” (all characters from the Arabic language) and a user try to reach the website connected to this domain name from an Internet café or airport, say, in Pakistan. He/she will not be able to reach that website unless if the variant “مكة” (Urdu variant) is already allocated and activated. Thus, variants need to be studied from both similarity point of view (by language community) and reachability pointy of view (based on input devices used by other language communities).”, Reference LGR for Arabic Language.

<sup>6</sup> “A small number of the LGRs contain additional LGR-specific WLE rules, reflecting a further constraints on possible labels based on the nature of the language or script.”, Reference Label Generation Rules (LGR) for the Second Level – Overview and Summary, Page 6.

5. RFC 5891 does not prohibit ASCII only labels.
  - a. The reference LGRs propose a whole-label evaluation rule that determines a label invalid (within the definition of the target LGR) when the label is composed of ASCII-only code points. This rule is allegedly taken from RFC 5891.
  - b. RFC 5891 does not prohibit a label containing all ASCII code points. RFC 5891 merely states “Strings that contain at least one non-ASCII character [...] are U-labels.”. So, the protocol is defining what a U-label or IDN is, but it is not requiring blocking a label that does not meet that criterion.
  - c. We would like ICANN org to expand on the need of this rule. Typically, registries will not reject an ASCII-only label registration.

In sum, the RySG is of the opinion that IDN policies should:

1. Be simple and scalable to foster adoption and growth.
2. Be focused on addressing the most egregious forms of potential security and stability risks (i.e. in-label script mixing, whole-script confusable labels).
3. Be developed in accordance with contractual and consensus policies to ensure greater acceptance by the ICANN community.

The RySG stands ready and looks forward to continuing working with ICANN org and the community to further develop IDNs in a way that is consistent with the above objectives.

---