Proposed Process to Review gTLD IDN Tables Without Reference Label Generation Rules

Global Domains & Strategy
1 August 2024

Proposed Process to Review gTLD IDN Tables Without Reference Label Generation Rules

1. Background

ICANN's Mission asks "to adopt or implement policies or procedures that take into account the use of domain names as natural-language identifiers" while preserving the security and stability of the Domain Name System (DNS). In line with this, ICANN org has been working with the support and guidance of technical and language/script communities to develop the relevant rules to implement for Internationalized Domain Names (IDNs), to enable secure and broader access to the DNS. As a result, ICANN org now supports IDNs in hundreds of languages written in 26 scripts, as noted below.

These rules, which are called IDN tables or Label Generation Rules (LGRs), are defined by the generic top-level domain (gTLD) registry operator (RO) for determining unique and valid domain labels in a particular language or script under its gTLD. Depending on the business needs of the RO, these rules may vary by the gTLD, but must be capable of generating a secure and stable set of labels. If an RO intends to offer registrations in different languages and scripts under its gTLD, an IDN table for each language or script must be defined. All proposed IDN tables are reviewed and approved by ICANN org for security and stability considerations before implementation.

To assist the community, ICANN org has developed reference IDN tables in machine-readable format, called Reference Label Generation Rules for the second level (Ref. LGRs). The Ref. LGRs are developed in consultation with the relevant script community and linguistic experts, while addressing technical security and stability concerns.

2. Role of Ref. LGRs in IDN Table Review Process for gTLD Registries

Ref. LGRs are used in reviewing IDN tables submitted by the gTLD registries through the IDN services or Registry System Testing (RST). ICANN org has also developed and launched the IDN Table Review Tool to further improve transparency and consistency of the gTLD IDN table review process. This self-service tool allows ROs and Registry Service Providers (RSPs) to evaluate their IDN tables using the Ref. LGRs before submitting them to ICANN org for review, as explained in the blog.

In total, 26 script-based and 31 language-based Reference LGRs have now been published:

- **26 script-based Ref. LGRs**: Arabic, Armenian, Bangla (Bengali), Chinese (Han), Cyrillic, Devanagari, Ethiopic, Georgian, Greek, Gujarati, Gurmukhi, Hebrew, Japanese (Hiragana, Katakana, Kanji [Han]), Kannada, Khmer, Korean (Hangul, Hanja [Han]), Lao, Latin, Malayalam, Myanmar, Oriya, Sinhala, Tamil, Telugu, and Thai.

- **31 language-based Ref. LGRs**: Arabic, Belarusian, Bosnian (Cyrillic), Bosnian (Latin), Bulgarian, Chinese, Danish, English, Finnish, French, German, Hebrew, Hindi, Hungarian, Icelandic, Italian, Japanese, Korean, Latvian, Lithuanian, Macedonian, Montenegrin, Norwegian, Polish, Portuguese, Russian, Serbian, Spanish, Swedish, Thai, and Ukrainian.

If a specific language-based Ref. LGR is not available, the IDN table can still be reviewed by using a relevant script-based Ref. LGR. For example, a Dutch language IDN table can be reviewed by using the Latin script Ref. LGR. The current script-based Ref. LGRs cover more than 386 actively used languages.

3. Supporting Additional Unicode Scripts for IDNs

The Unicode standard encodes 161 scripts in its latest version 15.1. While ICANN org supports 26 of these scripts (listed above), there are still many scripts which are not covered by the current set of Ref. LGRs.

The Unicode Standard Annex # 31: Unicode Identifier and Pattern Syntax (UAX#31), which states *"forms an integral part of the Unicode Standard … describes specifications for recommended defaults for the use of Unicode in the definitions of general-purpose identifiers,…".* UAX#31 categorizes the 161 scripts encoded by the Unicode standard as

Recommended, Limited Use, or Excluded for their use in identifiers, providing the following details:

- **Recommended Scripts:** "generally recommended for use in identifiers" because "these are in widespread modern customary use or are regional scripts in modern customary use by large communities."
- **Limited Use Scripts:** "in more limited use," so "to avoid security issues, some implementations may wish to disallow the limited-use scripts in identifiers."
- **Excluded Scripts:** "not in customary modern use, and thus implementations may want to exclude them from identifiers" because "these include historic and obsolete scripts, scripts used mostly liturgically, and regional scripts used only in very small communities or with very limited current usage. Some scripts also have unresolved architectural issues that make them currently unsuitable for identifiers."

ICANN org engaged experts to review the scripts and their Unicode categorization to determine the suitability of the scripts for IDNs. They analyzed the scripts based on their usage by language communities, documented their analysis and made some recommendations in the report on [Evaluating Unicode Scripts for Use in IDNs](#) based on the Unicode script categorization above.

In their feedback given during the [public comment](#) in January 2022 on this [report](#), the gTLD Registries Stakeholder Group (RySG) commented that the ROs should be allowed to support any scripts identified by the Unicode standard, "if a reasonable justification is made by a RO, this way, communities that use those scripts are not excluded from representation on the Internet from a domain name registration standpoint." The RySG noted that ICANN org may create Ref. LGR for additional scripts. However, the RySG further noted that ICANN org may want to do this reactively, rather than proactively, as the latter option offers diminishing returns since it would be difficult to assess the demand for scripts in Limited Use and Excluded scripts.

4. Proposed Process to Review IDN Tables without Ref. LGRs

There are three processes relevant to the IDN table review.

1. [IDN Service: Add, Modify, or Remove](#)
2. [IDN Service: IDN Tables: Update and Publish](#)
3. [Registry System Testing (RST)](#)

Based on further consultation with RySG, during the IDN services or Registry System Testing (RST) process, if an IDN table review request is submitted using a script for which the relevant Ref. LGR is not available, then ICANN org will take up the following process:

Step 1: ICANN org is not able to review an IDN table without a Ref. LGR for that language or script. If such an IDN table is submitted, ICANN org will ask the requesting RO to withdraw the IDN table from the request. The remaining IDN tables can proceed for review.

Step 2: ICANN org will work with the RO and the relevant language and script experts (as identified by the RO) and will work with them to develop the Ref. LGR.

Step 3: ICANN org will release the newly developed draft Ref. LGR for public comment and then finalize and publish it.

Step 4: The requester can then resubmit the request for the IDN table for review by ICANN org.

A Ref. LGR may take several months to complete as it is dependent on the complexity and breadth of the script, the availability of resources and documentation for the language and/or script, and the time needed for the community panel to complete the work. As such, it is not possible to set timelines for the steps noted.

ICANN org's role is limited to providing support to this community-led process, guiding the language/script Ref. LGR panel of volunteers, to consider all the technical and linguistic aspects. This includes testing the draft Ref. LGR against linguistic data, taking it through the public comment process and publishing the finalized Ref. LGR.