# Harmonization of IDN Tables

3 May 2023

## 1 What is an IDN Table?

IDN Tables are used by top-level domains (TLDs) to encode rules for determining what are valid internationalized domain name (IDN) labels and what are their variant labels for registration at the second level.

A TLD may offer multiple IDN Tables covering a variety of languages and scripts. For example, for a list of IDN Tables offered by the various TLDs see https://www.iana.org/domains/idn-Tables.

## 2 Why is IDN Table Harmonization Needed?

While designing and before implementing, each IDN Table is checked for security and stability considerations. In the current process, each IDN Table is checked individually without looking at additional IDN Tables being implemented by the Registry Operator for the same TLD. This process is sufficient to identify and address any security and stability issues, if the IDN Table was being used in isolation.

However, in practice, an IDN Table is not used in isolation and multiple IDN Tables may be implemented under a TLD. As explained below, there are additional security considerations which may arise in using an IDN Table with other IDN Tables within a TLD. Such considerations are currently not checked. Harmonization of IDN Tables will allow these security considerations to be checked and will address any cross-IDN-Table security concerns as well.

## 3 Security Issues Solved by Harmonization of the IDN Tables

Harmonization is the process of collectively designing and implementing IDN Tables offered by a TLD for a secure solution. Harmonization ensures that any two second-level variant labels s1 and s2 (as determined by the community work on label generation rules (LGRs)) are consistently identified as variant labels, irrespective of the specific IDN Table used to create either s1 or s2 under a TLD.

For example, a top-level domain (TLD) registry may implement ASCII and IDN Tables for Russian, Arabic and Urdu languages for second level registrations.

The TLD registration system may use the Urdu language IDN Table to register label s1: مکة “Mecca” (U+0645 U+06A9 U+06C3) in Arabic script. Separately, it can use the Arabic language IDN Table to register label s2: مكة “Mecca” (U+0645 U+0643 U+0629) in Arabic script. Generally, the Urdu language IDN Table will not contain the Arabic language letters U+0643 and U+0629; and the Arabic language IDN Table will not contain the Urdu language letters U+06A9 and U+06C3. The Arabic script community has identified {ک U+06A9, ك U+0643} and {ۃ U+06C3, ة U+0629} as variants.

Without harmonization of the two IDN Tables, the variant relationships are not captured in the Urdu language and the Arabic language IDN Tables. So, the labels s1 and s2 cannot be identified as variant labels, and can be registered as distinct labels under the TLD:

- S1: مكة.TLD
- S2: مكة.TLD

Similarly, the TLD registration system may register English/ASCII label s1: epic (U+0065 U+0070 U+0069 U+0063) in the Latin script. Separately, it can use the Russian language IDN Table to register label s2: epic (U+0435 U+0440 U+0456 U+0441) in the Cyrillic script. Again, the Latin and Cyrillic script communities have identified the following cross-script variant sets {e U+0065, e U+0435}, {p U+0070, p U+0440}, {i U+0069, i U+0456}, {c U+0063, c U+0441}.

Using these variant relationships, s1 and s2 should be variant labels. But, these labels are not identified as variant labels using the IDN Tables and registered as distinct labels under the TLD:

- S1: epic.TLD
- S2: epic.TLD

If these labels are not considered variant labels, and can be registered independently by different registrants, they cause security concerns for end users. This is because the variant labels are visually or otherwise “same” as considered by the relevant script communities.

Harmonization prevents such cases by ensuring whichever IDN Table is used to register one label, like s1, the process automatically identifies and so may withhold registration of its corresponding variant label s2 under the TLD, even across IDN Tables.

# 4 Scope of Harmonization of the IDN Tables

Harmonizing is done over a set of IDN Tables. IDN Guidelines 4.0 suggests harmonizing all IDN Tables within a TLD. Further work through IDN Variant TLD Recommendations extends the scope of harmonization to include all IDN Tables under all variant TLDs. Details are provided below.

IDN Guidelines 4.0 Guideline 13 says: “TLD registries must ensure that all applicable IDN Tables with an IDN variant policy for a particular TLD have uniform IDN variant code points that properly account for symmetry and transitivity properties of all IDN variant code point sets across these IDN Tables.”

The Additional Note II explains: “For Guideline 13: The use of “uniform” here means that

i. two IDN variant code points or IDN variant code point sequences in one IDN Table cannot be non-IDN-variant code points or non-IDN-variant code point sequences in another IDN Table implemented under the same TLD, and
ii. all code points in all the IDN Tables under the same TLD must be collectively considered for analysis of IDN variants of code points for each of these IDN Tables.

These two measures are suggested to prevent cases of IDN Variant Labels being generated by different IDN Tables under the same TLD to be allocated to different registrants."

IDN Variant TLD Recommendations extend the scope of harmonization of all IDN Tables to all variant TLDs, suggesting: "Second-level IDN Tables offered under IDN variant TLDs [are] harmonized." This means:

"Second-level IDN Tables applicable for an IDN variant TLD set must be mutually coherent but not necessarily identical. For two second-level variant labels s1 and s1v1 under any TLD t1 generated using the applicable IDN Table for t1, these must also be variant labels under TLD t1v1 if generated by the applicable IDN Table for t1v1 …".

## 5 The Process for Harmonizing the IDN Tables

There is no standard process for harmonizing an IDN Table. There are at least two ways of doing it:

1. **Extend each IDN Table:** Fully specify each IDN Table with all cross-language and/or cross-script variant code points to help identify the complete set of variant labels against a label. This option allows the use of the existing process, but makes the IDN Table more complex.
    a. Check a label against an updated IDN Table for validity.
    b. Identify all its variant labels and their dispositions using the IDN Table.

2. **Extend label check to a two-step process:** Keep the IDN Table relevant for the particular language or script and use a "Common" IDN Table in an additional step in the label evaluation process to identify the complete set of variant labels. This option makes the process for label checking a bit more complex but keeps the IDN Tables simpler.
    a. Check a label against the IDN Table for validity.
    b. Identify all its variant labels using the "Common" IDN Table and finalize the label dispositions.

For example, in the recent public comment on Reference LGRs for the second level, a Common LGR has also been published to identify such cross-language and cross-script variant code points [HTML, XML] (which is being finalized based on community input). Such an LGR can be used to identify additional variant relationships across languages and across scripts for the IDN Tables.

Each of the options 1 and 2 above can be automated using LGR processing tools, in case LGR format (RFC 7940) is used for encoding IDN Tables.