Internationalized Domain Names Expedited Policy Development Process

C4, C5, C6

ICANN

IDN-EPDP Team Meeting #81 | 11 May 2023

Agenda

1. Roll Call and SOI Updates (2 mins)

2. Welcome and Chair Updates (5 mins)

3. IDN table harmonization (charter questions C4, C5, C6) (110 mins)

- Background and presentation on what harmonization is

- Why it is important

- Deliberate on future tables

- Deliberate on existing tables (time-permitting)

4. AOB (3 mins)

Charter Questions: C4, C5, C6

==Mutually Coherent = Harmonized==

Harmonization <u>Core Question</u>

C4) Should the second-level IDN tables offered under a TLD, including IDN variant TLDs, be required to be mutually coherent? If yes, how should existing registrations which may not meet the "mutually coherent" requirement of second-level IDN tables be addressed?

Harmonization <u>Mechanism</u>

C5) The Staff Paper suggests maintaining a common set of harmonized second-level IDN tables for all IDN variant TLDs and then (a) choosing all these IDN tables to offer for all IDN variant TLDs, or (b) choosing a relevant different subset of IDN tables to offer for each different IDN variant TLD. Are the above suggested methods in the Staff Paper sufficient for IDN table harmonization purposes? Should any additional implementation guidance be provided for a registry?

Harmonization <u>Mechanism</u> -

IDN Table Format

C6) Should Registry Operators be required to use the machine readable LGR format as specified in RFC 7940 for their second-level IDN tables? Or should Registry Operators have the flexibility to resolve the harmonization issue so long as it can predictably and consistently produce the same variant labels, albeit with different disposition values, across the same-script IDN tables?
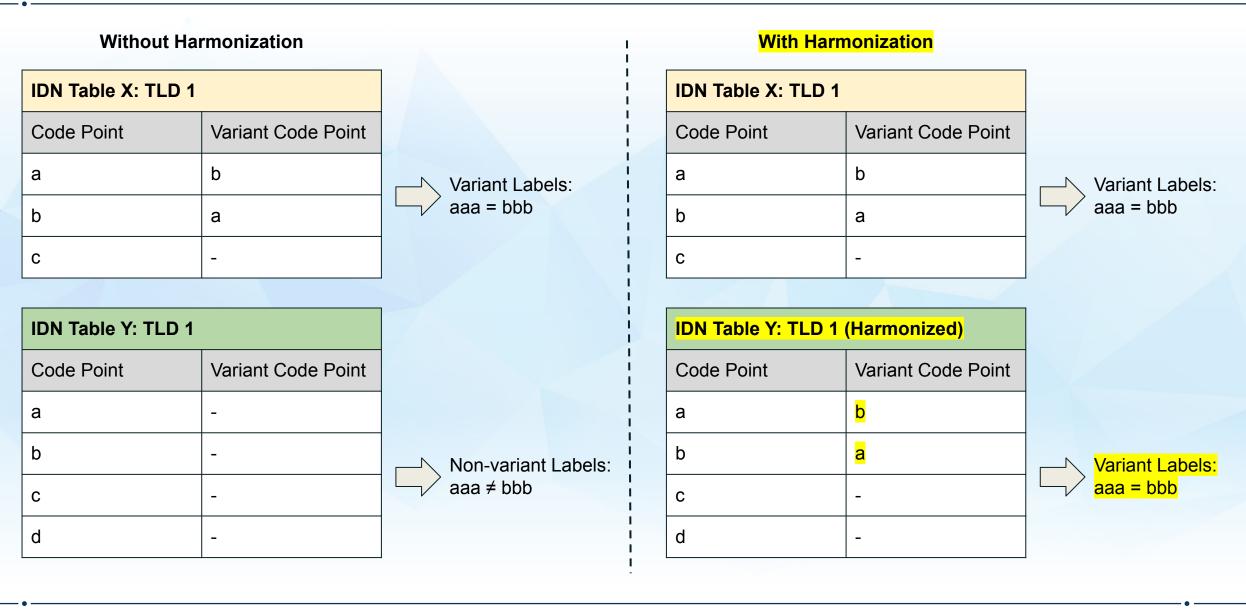
What is an IDN Table

- An IDN Table is used by a registry operator to represent **second-level rules** under a gTLD for:

    - validating IDN labels for registration
    - calculating variant labels
    - determining disposition values of variant labels

- A gTLD may offer **multiple IDN Tables** covering a variety of languages and scripts

- **Registry Operators develop IDN Tables** and **submit to ICANN org for review** for security / stability considerations

    - ICANN org reviews IDN Tables for significant security and stability or competition issues, e.g. through Registry Services Evaluation Policy (RSEP) process
    - Registrars rely on Registry Operators to check the requested label against IDN Tables

- Registry Operators may refer to the **Reference LGR**, which is developed in consultation with the script communities (i.e., Generation Panels), when developing their IDN Tables (*Reference LGR doesn't have the same weight as RZ-LGR*)

- Second-level IDN variant labels, as defined in a Registry Operator's IDN Tables and IDN Registration Rules **may be blocked or activated for registration** (see Exhibit A of RA)

- List of IDN Tables: https://www.iana.org/domains/idn-tables

What is Harmonization

> **Staff Paper Section 3.5.1: Harmonized IDN Tables for the Second-Level**
>
> - In case multiple IDN tables are offered, these be harmonized to produce a consistent set of second-level variant labels.
>
> - Second-level variant labels are ALSO required to be consistent across the IDN tables used for the TLD's variant labels.
>
> - The set of IDN tables **do not need to be exactly** the same under the TLD variants, **as long as they do not produce a conflicting set of second-level variant labels**.
>
> - If s1v1 and s1v2 are second level variant labels under t1 then these labels must not be non-variants under t1v1.

- **Summary**: Ensure that the **variant relationship** between any two given second-level labels is **consistently defined** across **all of the IDN Tables** offered by a gTLD and its variant gTLD(s)

- **End Goal**: No matter which IDN Table is used to calculate the variant labels of a requested second-level label, **the variant label set produced for the requested label must be consistent and include all of its variant labels identified in all of the IDN Tables** offered by that gTLD and its variant gTLD(s)

Example Output

Without Harmonization

With Harmonization

IDN Table X: TLD 1

Code Point	Variant Code Point
a	b
b	a
c	-

➡ Variant Labels:
aaa = bbb

IDN Table X: TLD 1

Code Point	Variant Code Point
a	b
b	a
c	-

➡ Variant Labels:
aaa = bbb

IDN Table Y: TLD 1

Code Point	Variant Code Point
a	-
b	-
c	-
d	-

➡ Non-variant Labels:
aaa ≠ bbb

IDN Table Y: TLD 1 (Harmonized)

Code Point	Variant Code Point
a	b
b	a
c	-
d	-

➡ Variant Labels:
aaa = bbb

Why is Harmonization needed

Current Practice:

- For a given code point in an IDN Table offered by a gTLD, there is no requirement to include that code point's other variant labels that are identified in the other IDN Tables offered by that gTLD

- A requested second-level label is only checked against a given IDN Table for variant labels without the requirement for being checked against the other IDN Tables offered by the same gTLD

- Each IDN Table is submitted to ICANN org for review / approval individually without the requirement for cross checking with the other IDN Tables offered by the same gTLD

Potential Consequence:

- The **second-level variant label set** produced for a requested second-level label **may be inconsistent** when different IDN Tables, which are offered by the same gTLD, are used

- Due to this inconsistency, **variant labels may be permitted for registration by different registrants as distinct labels** under the same gTLD

Example Output: Consequence Without Harmonization

Requested Label	IDN Table Language / Script	Unicode
Example A		
مكة	IDN Table 1: Arabic	U+0645 U+0643 U+0629
مكة	IDN Table 2: Urdu	U+0645 U+06A9 U+06C3
Example B		
epic	IDN Table 3: Latin	U+0065 U+0070 U+0069 U+0063
epic	IDN Table 4: Cyrillic	U+0435 U+0440 U+0456 U+0441

✅ Registrant A holds: T1.مكة

✅ Registrant B holds: T1.مكة

✅ Registrant C holds: epic.T2v1

✅ Registrant D holds: epic.T2v2

How are IDN Tables Harmonized

- **Currently no standard process for harmonizing IDN Tables**

- Staff Paper suggested two methods:

	Method 1	Method 2
Summary	**Extend Each IDN Table**	**Extend Label Check Process**
Specifics	Update each IDN Table to include relevant cross-language and cross-script variant code points to help identify the complete set of variant labels against a given label	Create an additional step in the label check process to check a requested label against: 1) relevant IDN Table AND 2) the **"Common" IDN Table** which includes all cross-language and cross-script variant code points
Pros	No change to label check process	No change to individual IDN Tables
Cons	More work for registries to update each IDN Table	More work for registries to update the label check process; impact on existing operations

- **"Common" IDN Table:** A Common LGR has been developed by language communities to identify cross-language and cross-script variant code points; the Common LGR is being finalized after Public Comment

Anecdote on Harmonization practice

Anecdote shared by Michael Bauland (experience as a backend operator for several registries)

➔ For each requested second-level label, **calculate its "canonical" form** based on **ALL ACTIVE IDN Tables** of a given gTLD

◆ Canonical = a given label's **variant code point of lowest unicode number**

● Example: U+0127 has variant code points U+0068 and U+0125; U+0068 is the "canonical" code point

◆ Canonical code point can derive from the same IDN Table of the requested label, or derive from a different IDN Table

● Example: For the Cyrillic letter U+0430, its canonical code point is U+0061, which is derived from the Latin IDN Table

➔ **Check whether the canonical form of a requested label is the same as the canonical form of any existing label**

➔ If the canonical form is the same, registration is blocked by default

If Harmonization becomes a policy requirement

Potential Outcome:

- ICANN org be authorized to **review ALL IDN Tables offered by a gTLD and its variant gTLDs in a <u>holistic manner</u>**

- ICANN org may reject an IDN Table **<u>if the variant label set of a given code point is not consistently produced</u>**

IDN Table formats

- **Evolution of IDN Table Formats:**

RFC 3743	RFC 4290	RFC 7940 (2016)
2004	2005	2016
Informational	Informational	Standards Track
TXT format	TXT format	XML format
List of code points	List of code points	List of code points
Variants for code points using semicolons (;)	Variants for code point using pipe symbol (|)	Variant definition on each code point
Rules described in comment section	Rules described in comment section	Rules are machine readable

- **Formats of IDN Table stored in IANA Repository (as of 5 Oct 2021)**
  - TXT: 12,985
  - XML: 1,113
  - HTML: 61
  - PDF: 1

- Reference LGR, including the Common LGR, use the **XML format** recommended by RFC 7940

- **LGR processing tools** can be developed to help registries **automatically harmonize IDN Tables in XML format**

Anecdote on IDN Tables

Anecdote shared by Zuan Zhang

- Chinese community uses **IDN Table developed by the Chinese Domain Name Consortium (CDNC)**

  - An independent non-profit organization

  - Jointly founded by four NICs of China Mainland, Hong Kong, Macau, and Taiwan (i.e., CNNIC, HKIRC, MONIC, TWNIC)

  - Coordinate and develop a **consistent Chinese IDN Table**

- CDNC Table includes

  - Simplified Chinese

  - Traditional Chinese

  - Variant labels of simplified and traditional Chinese

- CDNC Table follows RFC 3743 format

- Only the code points within the CDNC IDN Table can be available for registration

- To update the CDNC Table:

  - A CDNC member submits an a request to CDNC Secretary

  - CDNC Secretary refers the request to an Expert Panel for evaluation

  - If Expert Panel approves the request, CNDC Secretary solicits comments from other CDNC members

  - CDNC Board reviews request, and if approved, CDNC Table is updated accordingly

Discussion Questions - Future Applicants

For new IDN Tables <u>to be submitted</u> by future applicants:

1. Should harmonization be a requirement?

    a. In other words, should it be a requirement that the variant relationship between any two given second-level labels is consistently defined across all of the IDN Tables offered by a gTLD and its variant gTLD(s)

2. If the answer to question 1 is yes, should any specific harmonization mechanism be recommended?

    a. Specifically, should the Reference LGR, including the Common LGR, be recommended as a reference for developing IDN Tables by future applicants?

    b. Should the XML format, as recommended by RFC 7940, be required for IDN Tables to be submitted by future applicants?

Discussion Questions - Existing ROs

For existing IDN Tables <u>already implemented</u> by existing ROs:

1. Should harmonization be a requirement?

2. If the answer to question 1 is yes, should any specific harmonization mechanism be recommended?

3. If the answer to question 1 is yes, what would be the effect of the harmonization on existing registrations?

4. Should the XML format, as recommended by RFC 7940, be required retroactively for already implemented IDN Tables?

Discussion Questions - Existing ROs (Cont.)

For new IDN Tables <u>to be submitted</u> by existing ROs (e.g., as part of the variant gTLD application):

1. Should harmonization be a requirement?

    a. If the answer to question 1 is yes, how to manage any potential inconsistency with any of the ROs' already implemented IDN Tables?

2. If the answer to question 1 is yes, should any specific harmonization mechanism be recommended?

    a. Specifically, should the Reference LGR, including the Common LGR, be recommended as a reference for developing new IDN Tables to be submitted by existing ROs?

3. Should the XML format, as recommended by RFC 7940, be required for new IDN Tables to be submitted by existing ROs?