IDN Table Review Tool
(Pilot Version)

IDN Program

June 2021

ICANN

Table of Content

IDN Tables Review Tool

Agenda Item #1

Background

- A generic top-level domain (gTLD) registry operator that intends to offer registrations in different languages and scripts must:
    - Define an IDN table for each language and script.
    - Get it approved by ICANN during:
        - IDN RSEP Service.
        - Registry System Testing (RST) for Registry Service Provider (RSP) Change.
        - RST for Pre-Delegation Testing (PDT).

- ICANN reviews the IDN table(s) for security and stability considerations.

IDN Table Review Tool

- ⊙ The IDN Table Review Tool was developed to:
  - ○ Increase efficiency in reviewing IDN tables.
  - ○ Promote consistency and transparency of reviews.

- ⊙ The tool is available online for ICANN org and registry operators.
  - ○ Input: IDN table and relevant reference LGR.
  - ○ Output: IDN table review report in HTML format.

- ⊙ The same review analysis is available for ICANN org and contracted parties which helps make the process transparent.

- ⊙ Comparison is always done with the published reference LGR which helps make the process consistent.
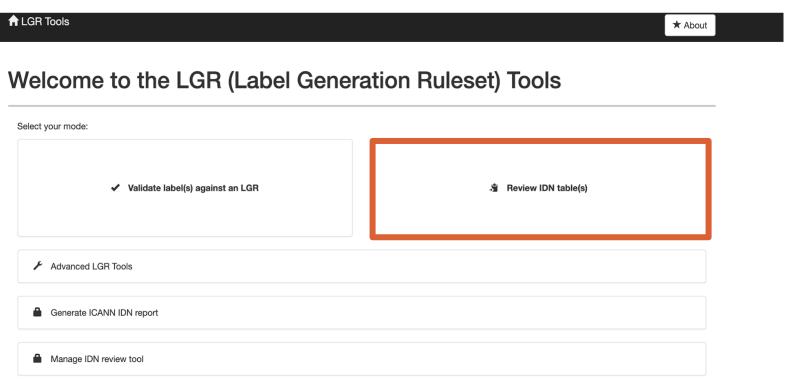
How to Use the IDN Table Review Tool

- ⊙ Step 1: Upload the IDN table(s).

- ⊙ Step 2: Select the reference LGR for comparison.

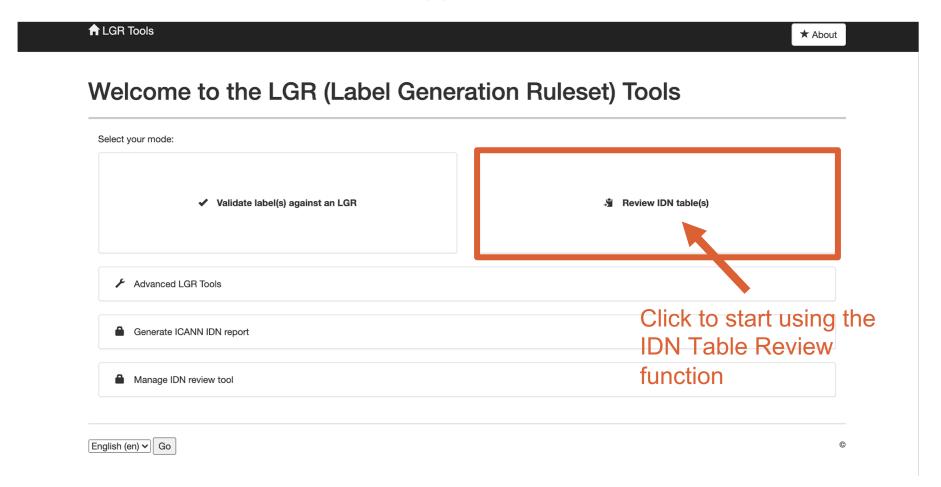- ⊙ Step 3: Review the HTML output report.

IDN Table Tool Interface

⊙ A user guide with more information is published on the LGR Tool webpage.

⊙ URL for IDN Table Review Tool https://lgrtool.icann.org/ (available in coming weeks).
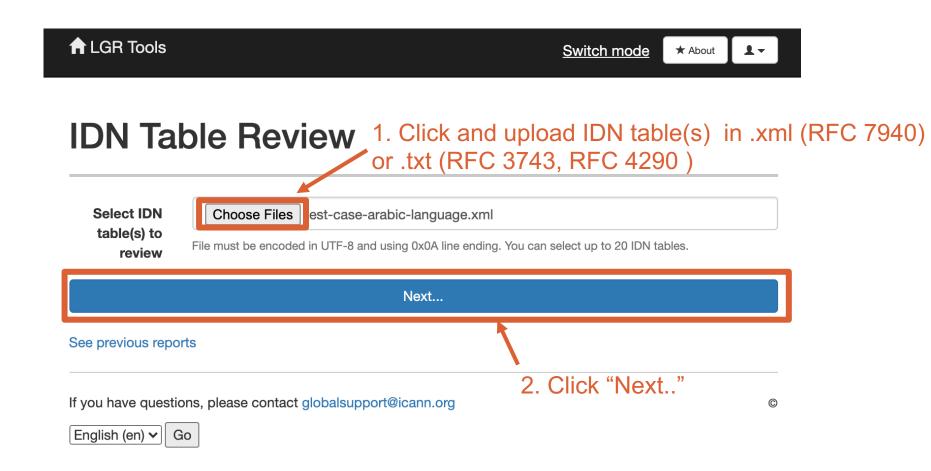
Step 1: Upload the IDN Table(s)

Step 1-1: Click the "Review IDN table(s)" menu.

Click to start using the IDN Table Review function
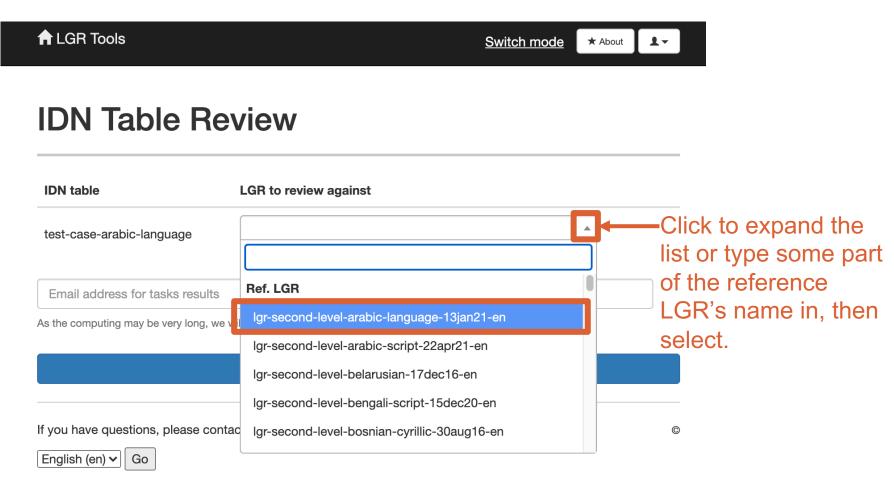
Step 1: Upload the IDN Table(s)

Step 1-2: Click "Choose Files" and select the IDN table(s) for review (the tool can review up to 20 IDN tables at a time), then click "Next…"

LGR Tools　　　　　　　　Switch mode　　★ About　　👤▾

IDN Table Review

1. Click and upload IDN table(s) in .xml (RFC 7940) or .txt (RFC 3743, RFC 4290)

Select IDN table(s) to review　　[Choose Files] est-case-arabic-language.xml

File must be encoded in UTF-8 and using 0x0A line ending. You can select up to 20 IDN tables.

Next...

See previous reports

2. Click "Next.."

If you have questions, please contact globalsupport@icann.org　　©

English (en) ▾　[Go]

Step 2: Select the Reference LGR for Comparison

Step 2-1: For each IDN table uploaded, select the relevant reference LGR from the dropdown list.

Click to expand the list or type some part of the reference LGR's name in, then select.

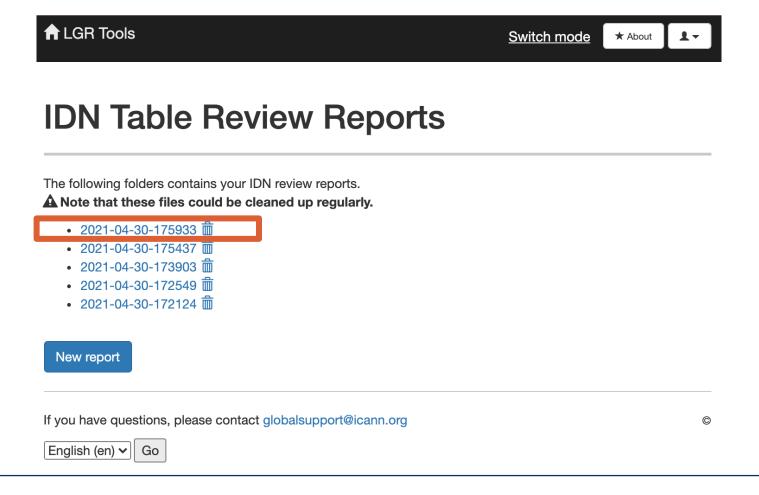# Step 2: Select the Reference LGR for Comparison

Step 2-2: Input your email address to receive a notification when the processing is complete.

Step 2: Select the Reference LGR for Comparison

Step 2-3: After getting a notification email, the HTML output reports can be viewed online or downloaded to a user's local machine.

Step 3: Review the HTML Output Report

⊙ The IDN Table Review Tool generates a report in HTML format for each uploaded IDN table.

⊙ Each report includes:
- Report Header: details of the IDN table and the reference LGR.
- Overall Summary: a quick summary of results.
- Detail Analysis for Each Section:
  - Language Tag
  - Repertoire
  - Variant Sets
  - Classes
  - WLE Rules
  - Actions

Structure of HTML Output Report

IDN Table Review Report

Table of Contents

Date: Feb. 23, 2021

Waiting for ICANN disclaimer

This document presents the review report for IDN Table:
aaa_da_3.0.xml Version: 3.0
with the Reference LGR:
lgr-second-level-danish-30aug16-en Version: 1

Report Header

Overall Summary

#	Comparing Item	Overall Result	Summarized Remarks	Remark occurence
1	Language Tag	MATCH	The language tag in IDN Table relevant to the script tag in Reference LGR	1
2	Repertoire	NOTE	Tags not required in Reference LGR Rules not required in Reference LGR	1
			Matches code point (including tags, context rule)	43
3	Variant Sets	MATCH	-	-
4	Classes	MATCH	-	-
5	Whole label evaluation rules	MANUAL CHECK	Match as a subset (for the rules missing in IDN Table, applicable code points in Ref. LGR are not in IDN Table)	2
			Mismatch (WLE rule only exists in Ref. LGR)	1
			Mismatch (WLE rule only exists in IDN Table)	1
6	Actions	MANUAL CHECK	Mismatch (missing action)	2

Overall Summary

Language Tag

#	IDN Table language tag	Ref. LGR language tag	Result	Remark
1	da	dan-Latn	MATCH	The language tag in IDN Table relevant to the script tag in Reference LGR

Repertoire

#	Code Point	Glyph	Name	IDN Table	Ref. LGR	Result	Remark
1	U+002D	-	HYPHEN-MINUS	✓	✓	NOTE	Tags not required in Reference LGR Rules not required in Reference LGR
2	U+0030	0	DIGIT ZERO	✓	✓	MATCH	Matches code point (including tags, context rule)
3	U+0031	1	DIGIT ONE	✓	✓	MATCH	Matches code point (including tags, context rule)
4	U+0032	2	DIGIT TWO	✓	✓	MATCH	Matches code point (including tags, context rule)
5	U+0033	3	DIGIT THREE	✓	✓	MATCH	Matches code point (including tags, context rule)
6	U+0034	4	DIGIT FOUR	✓	✓	MATCH	Matches code point (including tags, context rule)
7	U+0035	5	DIGIT FIVE	✓	✓	MATCH	Matches code point (including tags, context rule)

Detail Analysis

Report Header

⊙ Report Header includes:
  ○ Table of contents.
  ○ Date of comparison.
  ○ IDN table name and version, if available.
  ○ Reference LGR name and version.

IDN Table Review Report

Table of Contents

Date: Feb. 23, 2021

This document presents the review report for IDN Table:
aaa_da_3.0.xml Version: 3.0
with the Reference LGR:
lgr-second-level-danish-30aug16-en Version: 1

Overall Summary

◉ Overall Summary includes:
  ○ The result for each section.
    • Result categories: MATCH, SUBSET, NOTE, MANUAL CHECK, REVIEW.
  ○ Remarks and their occurrences.

Overall Summary

#	Comparing Item	Overall Result	Summarized Remarks	Remark occurence
1	Language Tag	MATCH	The language tag in IDN Table relevant to the script tag in Reference LGR	1
2	Repertoire	NOTE	Tags not required in Reference LGR Rules not required in Reference LGR	1
			Matches code point (including tags, context rule)	43
3	Variant Sets	MATCH	-	-
4	Classes	MATCH	-	-
5	Whole label evaluation rules	MANUAL CHECK	Match as a subset (for the rules missing in IDN Table, applicable code points in Ref. LGR are not in IDN Table)	2
			Mismatch (WLE rule only exists in Ref. LGR)	1
			Mismatch (WLE rule only exists in IDN Table)	1
6	Actions	MANUAL CHECK	Mismatch (missing action)	2

Output Result Interpretation

Result	Interpretation
MATCH	Data in IDN table exactly matches the data in the reference LGR.
SUBSET	Data in IDN table matches the data in the reference LGR and does not create security issue; e.g., IDN table code point repertoire is a subset of reference LGR code point repertoire.
NOTE	Data in the IDN table is more conservative than data in the reference LGR; e.g., IDN table applies some additional rules on the code points while reference LGR does not.
MANUAL CHECK	The tool finds a deviation from the reference LGR but cannot decide whether the deviation causes any security or stability issues. Such cases require further manual review to confirm; e.g., IDN table includes an extra code point.
REVIEW	The tool finds a deviation from the reference LGR which may potentially cause security or stability issues; e.g., IDN table omits a variant mapping present in the reference LGR.

Modifying an IDN Table Based on Detail Analysis

⊙ For "MANUAL CHECK" cases: the IDN table owner should make sure that the deviation does not create any security issues; supporting information should be included in the description section of the IDN table.

⊙ For "REVIEW" cases: the IDN table owner should consider if additional mechanisms should be added to address potential security issues.

Webinar and Demo

- For webinar session and demonstration of the tool please visit the webinar recording.

- The webinar includes:
  - IDN Table and Reference Label Generation Rules (LGRs)
  - IDN Table Review Tool features
  - How to use the IDN Table Review Tool
  - Demo of IDN Table Review Tool

Conclusion

- Link to the IDN Table Review Tool:
  https://www.icann.org/resources/pages/lgr-toolset-2015-06-21-en

- Link to reference LGRs:
  https://www.icann.org/resources/pages/second-level-lgr-2015-06-21-en

- This is currently a pilot release, and you are invited to use the tool and share feedback with us, which will lead to further improvements. Please contact IDNProgram@icann.org to report any bugs.

- Please remember: The IDN table review function of the LGR tool compares IDN tables with reference label generation rulesets. The purpose of this report is to assist the user in identifying potential issues that exist in IDN tables. The IDN Table Review Tool report is for information only – it is not a warranty or guarantee of an ICANN IDN table review process.

Appendix A:

Introduction to IDN Tables

What is an IDN Table?

- ◉ IDN tables define the labels that can be registered for a particular language or script under a top-level domain (TLD).

  - ○ IDN tables help manage the security, stability, and usability of IDNs.

- ◉ IDN tables include Unicode code points, variants, and linguistic and technical rules to determine appropriate and secure domain labels.

  - ○ Repertoire: {U+0061 - U+007A, U+0030 - U+0039, U+002D}; a-z, 0-9, -.
  - ○ Variants: U+0629 (ة) and U+06C3 (ۃ) are variants in Arabic script.
  - ○ Rules:

    - • Linguistic: A tone mark follows the main consonant in the Lao script.
    - • Technical: The Unicode string MUST NOT contain a hyphen in the third- and fourth-character positions as per RFC 5891.

- ◉ IDN tables can be defined in multiple formats:

  - ○ Latest machine-readable XML format specified in RFC 7940.

    - • RFC 7940 refers to IDN tables as Label Generation Rules (LGRs).
  - ○ Previously, text-based formats in RFC 3743 and RFC 4290 were also used.

⊙ Repertoire is a subset of possible code points based on IDNA2008.

Variant Labels

⊙ Variant labels are generated by code point (or sequence of code points) that could be substituted for a code point (or sequence of code points) that is considered the "same" in some measure by a given community of Internet users. They can be defined for **Security** and/or **Usability**

fit
0066 0069 0074

ƒit
0192 0069 0074

.澳门
6FB3 95E8

.澳門
6FB3 9580

السعودية
0627 0644 0633 0639 0648 062F 064A 0629

السعودية
0627 0644 0633 0639 0648 062F 06CC 06C3

Rules

◉ Linguistic: A tone mark follows the main consonant in the Lao script.

ຊ + ◌̃   rendered as ຊ̃ ✓

0E8A   0EC9

ຊ + ◌̃ + ◌̃   rendered as

0E8A   0EC9   0EC9

 ✗ Chrome v.90.0.4430.93 on Window 10 Home

 ✗ Chrome v.90.0.4430.93 on macOS Catalina v10.15.7

 ✗ Chrome v.12.4.9.23 on Android

◉ Technical: The Unicode string MUST NOT contain a hyphen in the third- and fourth-character positions as per RFC 5891.

    ○ niña-café ✓

    ○ ni--ñacafé ✗

Format of an IDN Table

- IDN tables include a Repertoire, Variant Mappings, and Rules.

IDN Table

Code Point Repertoire
Variant Mappings
Rules

- IDN tables can be defined in multiple formats:

  - Latest machine-readable XML format specified in RFC 7940.

    - RFC 7940 refers to IDN tables as Label Generation Rules (LGRs).

  - Previously, text-based formats in RFC 3743 and RFC 4290 were also used.

IDN Table Example in XML Format (RFC 7940)

```xml
▼<lgr xmlns="urn:ietf:params:xml:ns:lgr-1.0">
  ▼<meta>
    <version comment="Second Level Reference LGR">3</version>
    <date>2020-08-24</date>
    <language>und-Khmr</language>
    <unicode-version>6.3.0</unicode-version>
  ▶<description type="text/html">
    ...
    </description>
  ▶<references>
    ...
    </references>
  </meta>
  ▼<data>
    <char cp="002D" not-when="h
  ▼<char cp="0030" tag="Common
    <var cp="17E0" type="block
    </char>
  ▼<char cp="0031" tag="Common
    <var cp="17E1" type="block
    </char>
  ▼<char cp="0032" tag="Common
    <var cp="17E2" type="block
    </char>
  ▼<char cp="0033" tag="Common
    <var cp="17E3" type="block
    </char>
  ▼<char cp="0034" tag="Common-digit sc:Zyyy" ref="0" comment="DIGIT FOUR; ⊛">
    <var cp="17E4" type="blocked" comment="Khmer digit variant"/>
    </char>
  ▼<char cp="0035" tag="Common-digit sc:Zyyy" ref="0" comment="DIGIT FIVE; ⊛">
    <var cp="17E5" type="blocked" comment="Khmer digit variant"/>
    </char>
  ▼<char cp="0036" tag="Common-digit sc:Zyyy" ref="0" comment="DIGIT SIX; ⊛">
    <var cp="17E6" type="blocked" comment="Khmer digit variant"/>
    </char>
  ▼<char cp="0037" tag="Common-digit sc:Zyyy" ref="0" comment="DIGIT SEVEN; ⊛">
    <var cp="17E7" type="blocked" comment="Khmer digit variant"/>
    </char>
```

```xml
▼<char cp="0038" tag="Common-digit sc:Zyyy" ref="0" comment="DIGIT EIGHT; ⊛">
  <var cp="17E8" type="blocked" comment="Khmer digit variant"/>
  </char>
▼<char cp="0039" tag="Common-digit sc:Zyyy" ref="0" comment="DIGIT NINE; ⊛">
  <var cp="17E9" type="blocked" comment="Khmer digit variant"/>
  </char>
  <char cp="1780" tag="consonant sc:Khmr series-three" ref="3 203 205" comment="Khmer"/>
  <char cp="1781" tag="consonant sc:Khmr" ref="3 203" comment="Khmer"/>
  <char cp="1782" tag="consonant sc:Khmr" ref="3 203" comment="Khmer"/>
  <char cp="1783" tag="consonant sc:Khmr" ref="3 203" comment="Khmer"/>
  <char cp="1784" tag="consonant sc:Khmr series-three series-two" ref="3 203 205 210" c
                                              ref="3 203 205" comment="Khmer"/
                                              comment="Khmer"/>
                                              comment="Khmer"/>
                                              comment="Khmer"/>
                                              series-two" ref="3 203 205 210" c
                                              comment="Khmer"/>
                                              comment="Khmer"/>
                                              comment="Khmer"/>
                                              comment="Khmer"/>
                                              comment="Khmer"/>
                                              e" ref="3 203 205" comment="Khmer"/
                                              comment="Khmer"/>
                                              comment="Khmer"/>
                                              comment="Khmer"/>
                                              e" ref="3 203 205" comment="Khmer"/
                                              series-three series-two" ref="3 203
  <char cp="1796" tag="consonant sc:Khmr" ref="3 203" comment="Khmer"/>
  <char cp="1797" tag="consonant sc:Khmr" ref="3 203" comment="Khmer"/>
  <char cp="1798" tag="consonant sc:Khmr series-two" ref="3 203 210" comment="Khmer"/>
  <char cp="1799" tag="consonant sc:Khmr series-two" ref="3 203 210" comment="Khmer"/>
  <char cp="179A" tag="consonant sc:Khmr series-two" ref="3 203 210" comment="Khmer"/>
  <char cp="179B" tag="consonant sc:Khmr series-three" ref="3 203 205" comment="Khmer"/>
  <char cp="179C" tag="consonant sc:Khmr series-two" ref="3 203 210" comment="Khmer"/>
  <char cp="179F" tag="consonant sc:Khmr series-one series-three" ref="3 203 205 210" c
  <char cp="17A0" tag="consonant sc:Khmr series-one" ref="3 203 210" comment="Khmer"/>
  <char cp="17A1" tag="base-only consonant sc:Khmr" ref="3 203" comment="Khmer"/>
  <char cp="17A2" tag="consonant sc:Khmr series-one" ref="3 203 210" comment="Khmer"/>
  <char cp="17A5" tag="independent-vowel sc:Khmr" ref="3 206" comment="Khmer"/>
  <char cp="17A6" tag="independent-vowel sc:Khmr" ref="3 206" comment="Khmer"/>
```

```xml
▼<rule name="follows-consonant-robat-shifter" comment="WLE Rule 5: makes
  sure that dependent vowel follows a consonant or a shifter or a robat">
  ▼<look-behind>
    ▼<choice>
        <class by-ref="shifter"/>
        <class by-ref="consonant"/>
        <class by-ref="robat"/>
      </choice>
    </look-behind>
    <anchor/>
  </rule>
```

IDN Table Example in Text Format (RFC 4290)

```
# Info #################################################
#
# Registry:    Verisign Inc.
# Contact:     Registry Customer Service
# Email:       info@verisign-grs.com
# Phone:       +1-703-925-6999
# Script:      Khmer (khmr)
# Version:     2.5
# Date:        March 21st, 2017
# Unicode:     6.0.0
#
#################################################


# Applicable Rules #######  #################################################
#
# Code points: U+002D (HYPHEN-MINUS)
# Reference:   RFC 5891 4.2.3.1
# Rules:       Cannot begin or end a label
#              Cannot occupy 3rd and 4th position
#
# Code points: U+17B6..U+17D3, U+17DD (COMBINING MARKS)
# Reference:   RFC 5891 4.2.3.2
# Rules:       Cannot begin a label
#
# Code points: U+17D7 (MODIFIER LETTERS)
# Reference:   Predelegation Tester Request
# Rules:       Cannot begin a label
#
#################################################


# Points #################################################
U+002D  # Common        HYPHEN-MINUS
U+0030  # Common        DIGIT ZERO
U+0031  # Common        DIGIT ONE
U+0032  # Common        DIGIT TWO
U+0033  # Common        DIGIT THREE
U+0034  # Common        DIGIT FOUR
U+0035  # Common        DIGIT FIVE
```

```
U+0036  # Common        DIGIT SIX
U+0037  # Common        DIGIT SEVEN
U+0038  # Common        DIGIT EIGHT
U+0039  # Common        DIGIT NINE
U+1780  # Khmer KHMER LETTER KA
U+1781  # Khmer KHMER LETTER KHA
U+1782  # Khmer KHMER LETTER KO
U+1783  # Khmer KHMER LETTER KHO
U+1784  # Khmer KHMER LETTER NGO
U+1785  # Khmer KHMER LETTER CA
U+1786  # Khmer KHMER LETTER CHA
U+1787  # Khmer KHMER LETTER CO
U+1788  # Khmer KHMER LETTER CHO
U+1789  # Khmer KHMER LETTER NYO
U+178A  # Khmer KHMER LETTER DA
U+178B  # Khmer KHMER LETTER TTHA
U+178C  # Khmer KHMER LETTER DO
U+178D  # Khmer KHMER LETTER TTHO
U+178E  # Khmer KHMER LETTER NNO
U+178F  # Khmer KHMER LETTER TA
U+1790  # Khmer KHMER LETTER THA
U+1791  # Khmer KHMER LETTER TO
U+1792  # Khmer KHMER LETTER THO
U+1793  # Khmer KHMER LETTER NO
U+1794  # Khmer KHMER LETTER BA
U+1795  # Khmer KHMER LETTER PHA
U+1796  # Khmer KHMER LETTER PO
U+1797  # Khmer KHMER LETTER PHO
U+1798  # Khmer KHMER LETTER MO
U+1799  # Khmer KHMER LETTER YO
U+179A  # Khmer KHMER LETTER RO
U+179B  # Khmer KHMER LETTER LO
U+179C  # Khmer KHMER LETTER VO
U+179D  # Khmer KHMER LETTER SHA
U+179E  # Khmer KHMER LETTER SSO
U+179F  # Khmer KHMER LETTER SA
U+17A0  # Khmer KHMER LETTER HA
U+17A1  # Khmer KHMER LETTER LA
U+17A2  # Khmer KHMER LETTER QA
U+17A3  # Khmer KHMER INDEPENDENT VOWEL QAQ
U+17A4  # Khmer KHMER INDEPENDENT VOWEL QAA
```

```
U+17A5  # Khmer KHMER INDEPENDENT VOWEL QI
U+17A6  # Khmer KHMER INDEPENDENT VOWEL QII
U+17A7  # Khmer KHMER INDEPENDENT VOWEL QU
U+17A8  # Khmer KHMER INDEPENDENT VOWEL QUK
U+17A9  # Khmer KHMER INDEPENDENT VOWEL QUU
U+17AA  # Khmer KHMER INDEPENDENT VOWEL QUUV
U+17AB  # Khmer KHMER INDEPENDENT VOWEL RY
U+17AC  # Khmer KHMER INDEPENDENT VOWEL RYY
U+17AD  # Khmer KHMER INDEPENDENT VOWEL LY
U+17AE  # Khmer KHMER INDEPENDENT VOWEL LYY
U+17AF  # Khmer KHMER INDEPENDENT VOWEL QE
U+17B0  # Khmer KHMER INDEPENDENT VOWEL QAI
U+17B1  # Khmer KHMER INDEPENDENT VOWEL QOO TYPE ONE
U+17B2  # Khmer KHMER INDEPENDENT VOWEL QOO TYPE TWO
U+17B3  # Khmer KHMER INDEPENDENT VOWEL QAU
U+17B6  # Khmer KHMER VOWEL SIGN AA
U+17B7  # Khmer KHMER VOWEL SIGN I
U+17B8  # Khmer KHMER VOWEL SIGN II
U+17B9  # Khmer KHMER VOWEL SIGN Y
U+17BA  # Khmer KHMER VOWEL SIGN YY
U+17BB  # Khmer KHMER VOWEL SIGN U
U+17BC  # Khmer KHMER VOWEL SIGN UU
U+17BD  # Khmer KHMER VOWEL SIGN UA
U+17BE  # Khmer KHMER VOWEL SIGN OE
U+17BF  # Khmer KHMER VOWEL SIGN YA
U+17C0  # Khmer KHMER VOWEL SIGN IE
U+17C1  # Khmer KHMER VOWEL SIGN E
U+17C2  # Khmer KHMER VOWEL SIGN AE
U+17C3  # Khmer KHMER VOWEL SIGN AI
U+17C4  # Khmer KHMER VOWEL SIGN OO
U+17C5  # Khmer KHMER VOWEL SIGN AU
U+17C6  # Khmer KHMER SIGN NIKAHIT
U+17C7  # Khmer KHMER SIGN REAHMUK
U+17C8  # Khmer KHMER SIGN YUUKALEAPINTU
U+17C9  # Khmer KHMER SIGN MUUSIKATOAN
U+17CA  # Khmer KHMER SIGN TRIISAP
U+17CB  # Khmer KHMER SIGN BANTOC
U+17CC  # Khmer KHMER SIGN ROBAT
U+17CD  # Khmer KHMER SIGN TOANDAKHIAT
```

Appendix B:

Introduction to Reference LGRs

Reference Label Generation Rules (LGRs)

- Reference LGRs are based on the RZ-LGR, and designed for the second level:
  - Formed in consultation with the relevant script community.
  - Follow design principles in RFC 6912 and SSAC's SAC060.
  - Developed in machine-readable RFC 7940 format.
  - Pre-vetted for security and stability considerations.

- Registry operators are encouraged to consult the reference LGRs while they design their own IDN tables. However, strict adherence is not required.

- Reference LGRs are used as a baseline for reviewing IDN tables for generic top-level domain (gTLD) registries.

- Reference LGRs are published at https://www.icann.org/resources/pages/second-level-lgr-2015-06-21-en

Engage with ICANN – Thank You and Questions

One World, One Internet

Visit us at **icann.org**

@icann

facebook.com/icannorg

youtube.com/icannnews

flickr.com/icann

linkedin/company/icann

slideshare/icannpresentations

soundcloud/icann

instagram.com/icannorg