
Public Comment Summary Report

Additional Script-Based Reference Label Generation Rules and Related Updates

Open for Submissions Date:

Thursday, 19 January 2023

Closed for Submissions Date:

Friday, 03 March 2023

Summary Report Due Date:

Friday, 14 April 2023 (Extended from Wednesday, 08 March 2023)

Category: Technical

Requester: ICANN org

ICANN org Contact(s): pitinan.koo@icann.org, sarmad.hussain@icann.org

Open Proceeding Link:

<https://www.icann.org/en/public-comment/proceeding/additional-script-based-reference-label-generation-rules-and-related-updates-19-01-2023>

Outcome:

ICANN org appreciates the comments submitted by the community on the additional script-based Reference Label Generation Rules (LGRs) and related updates.

ICANN org received 11 comments. Ten comments agree that these Reference LGRs meet the goal of defining second-level Internationalized Domain Names (IDNs). Of these, six comments also provide additional queries or suggestions. One comment does not support the Reference LGRs for the Latin script. After analyzing the comments and incorporating them as needed by ICANN org, the final version of Reference LGRs will be posted on the [Reference LGR web page](#).

Section 1: What We Received Input On

ICANN org has previously published [Reference LGRs for the second level](#) for multiple scripts and languages. The additional Reference LGRs that are the subject of this Public Comment proceeding were developed based on the detailed analysis and solutions conducted by the script community through the work on [Root Zone Label Generation Rules \(RZ-LGRs\)](#). The following Reference LGRs were released for Public Comment to gather community feedback.

- Seven new script-based Reference LGRs: Armenian, Cyrillic, Greek, Latin, Japanese, Korean, and Myanmar.

- Fifteen Reference LGRs with normative changes: Belarusian language, Bosnian (Cyrillic) language, Bulgarian language, English language, French language, German language, Hebrew language, Hebrew script, Khmer script, Macedonian language, Montenegrin language, Russian language, Serbian language, Sinhala script, and Ukrainian language.
- The remaining 31 Reference LGRs had the Unicode version updated without normative changes. The Unicode version of all Reference LGRs has been updated to version 11.0.0 to match the Maximal Starting Repertoire version 5 ([MSR-5](#)). MSR is the starting point for the work done by community-based [Generation Panels](#) that develop RZ-LGRs proposals for relevant scripts.

Section 2: Submissions

Organizations and Groups:		
Name	Submitted by	Initials
Registries Stakeholder Group (RySG)		RySG
Armenian Generation Panel	Igor Mkrtumyan	Armn-GP

Individuals:		
Name	Affiliation (if provided)	Initials
Jay Puadyal	Neo-Brahmi Generation Panel	JP
Nitin Walia		NW
Raymond Doctor	Neo-Brahmi Generation Panel	RD
Bill Jouris	At-Large	BJ
Hiro Hotta	Japan Registry Services (JPRS)	HH
Ehsanul Kabir	Bangladesh Telecommunication Regulatory Commission (BTRC), GAC Member from Bangladesh	EK
Mohammad Abdul Haque	Bangladesh Internet Governance Forum	MAH
Nazmul Hasan Majumder	Bangladesh School of Internet Governance (bdSIG)	NHM
Doron Shikmoni	Hebrew Generation Panel	DS

Section 3: Summary of Submissions

ICANN org received 11 comments on the additional script-based Reference LGRs and updated LGRs. Four comments fully agree that these reference LGRs meet the goal of defining

Internationalized Domain Name (IDN) labels for the language or script that are suitable for the second level. Six comments are in support and provide additional queries or suggestions. One comment does not support the Reference LGR for the Latin script.

Armn-GP, JP, NW, and RD support these reference LGRs and provide affirmation on Armenian script, Devanagari script, Hindi language, and Gujarati script Reference LGRs.

There are six supporting comments from organizations and individuals, which include additional queries or suggestions for the finalized Reference LGRs.

The RySG supports the process that creates the Reference LGRs and appreciates that ICANN org explicitly seeks experts and engages with them to ensure that production of Reference LGRs is made with a certain level of quality. In addition, the RySG requests ICANN org to continue developing Reference LGRs for additional languages and scripts.

The RySG questions the rationale for the solution of three code points U+006C U+00B7 U+006C representing Ela Geminada used in the Catalan orthography. The proposed solution is different from the current practice by relevant registry operators.

HH supports the Reference LGRs and appreciates that the Japanese language (standalone LGR) does not include integration affects with Chinese or Korean languages. In addition, HH provides two additional contextual rules for the Japanese language (standalone LGR) and Japanese script Reference LGRs to make the solution more secure and align with the natural usage of daily Japanese.

EK, MAH, and NHM support the Reference LGRs and request to incorporate the atomic version of three decomposed characters in the Bangla Reference LGR, namely য় YA (U+09DF), ড় RRA (U+09DC), and ঠ RRHA (U+09DD), instead of the sequence versions, "য" YA (U+09AF) + "়" Nukta (U+09BC), "ড" DDA (U+09A1) + "়" Nukta (U+09BC), and "ঠ" DDHA (U+09A2) + "়" Nukta (U+09BC).

DS supports the Reference LGRs and notes that defining block variant code points for "i" to "VAV" and "o" to "SAMECH" may not solve the real problem. However, there is no harm in including these variant definitions in the Hebrew script Reference LGR.

There is one comment does not support the Latin script Reference LGR.

BJ comments that the Root Zone LGRs were developed under the assumption that applied-for gTLDs are subject to manual review to avoid confusion while the second level domain name registrations are not, and cannot be, manually reviewed. In the interest of reducing confusion, and of avoiding enabling DNS abuse, the threshold of defining variant code points should be adjusted to include larger cases than the top-level consideration.

Section 4: Analysis of Submissions

ICANN org thanks all the contributors for their valuable input and feedback. All comments have been taken into consideration.

In response to the RySG’s query, the solution for Ela Geminada (U+006C U+00B7 U+006C) will be updated to be a variant code point sequence with Latin Small Letter L – Hyphen – Small Latin Letter L (U+006C U+002D U+006C) to match the current implementation by registry operators. The existing variant to U+006C U+006C will be removed. These updates are aligned with the existing practices by relevant registry operators and do not affect existing registrations.

The additional contextual rules suggested by HH will be included in the Reference LGR for the Japanese language and Japanese script.

Based on the definition of IDNA in RFC5982, IDN labels are encoded using the Unicode Normalization Form C (NFC). Following the Unicode standard, the NFC form of Bangla character YYA, RRA, and RRHA are the sequences with Nukta. Therefore, no further update is required for the Bangla Reference LGR.

In response to BJ’s comment, some of the examples provided may fall into the area of string similarity which is addressed in the IDN Implementation Guidelines version 4.1.

The Recommendation 14 states that “TLD registries are encouraged to consider IDN policies to minimize confusion of IDN labels with other labels within the same script, specifically arising due to homoglyph character”

The Additional Note IV further states that “It is important to understand that not all visual similarity issues can be addressed by IDN Tables and IDN policies. Other policies such as dispute resolution policies may be necessary to mitigate against abusive registrations exploiting visually similar characters.”

ICANN org will finalize the Latin script Reference LGR, based on the input available by the Latin script community in the RZ-LGR.

ICANN org will look into string similarity for the root zone. Based on the findings, ICANN will discuss with the relevant panels and the community if there are any additional updates needed in the relevant Reference LGRs.

Section 5: Next Steps

After consideration of the comments, and modification as needed, the final version of the Reference LGRs will be posted on the [Reference LGR web page](#).