

Maximal Starting Repertoire Version 5 (MSR-5)

Publication Date: 21 June 2021

Prepared By: Pitinan Kooarmornpatana

Public Comment Proceeding

| | |
|------------------------|---------------|
| Open Date: | 27 April 2021 |
| Close Date: | 8 June 2021 |
| Staff Report Due Date: | 24 June 2021 |

Important Information Links

| |
|-------------------------------------------|
| Announcement |
| Public Comment Proceeding |
| View Comments Submitted |

Staff Contact: Pitinan Kooarmornpatana

Email: pitinan.koo@icann.org

Section I: General Overview and Next Steps

Under the [Procedure to Develop and Maintain Label Generation Rules for the Root Zone with Respect to IDN Labels](#), the Maximal Starting Repertoire (MSR) is the starting point for the work done by community-based [Generation Panels](#) that develop Root Zone Label Generation Rules ([RZ-LGR](#)) proposals for relevant scripts. The Maximal Starting Repertoire Version 5 (MSR-5) is upwardly compatible with [MSR-4](#), adding one code point each to the Latin and Devanagari script repertoires, and adding two code points to the Arabic script repertoire. MSR-5 also changes the Unicode base version from 6.3 to 11.0. Code points added between Unicode Version 7.0 and 11.0 which meet the criteria for inclusion are added in the MSR-5.

The contents of MSR-5 and a detailed rationale behind its development are available in the MSR-5 [Overview and Rationale document](#). Based on the feedback received by the community, the Integration Panel will finalize the code point repertoire for MSR-5. After the release of MSR-5, Generation Panels, will be able to use the updated contents as a starting point for their analysis. If additional code points do not meet the criteria for inclusion in the affected Root Zone LGR, the LGRs would remain unchanged.

Section II: Contributors

At the time this report was prepared, a total of four (4) submissions from one (1) community member had been posted to the forum. The contributor is listed below. To the extent that quotations are used in the foregoing narrative (Section III), such citations will reference the contributor's initials.

Organizations and Groups:

None

Individuals:

| Name | Affiliation (if provided) | Initials |
|-------------|---------------------------|----------|
| Bill Jouris | | BJ |

Section III: Summary of Comments

General Disclaimer: *This section intends to summarize broadly and comprehensively the comments submitted to this public comment proceeding but does not address every specific*

position stated by each contributor. The preparer recommends that readers interested in specific aspects of any of the summarized comments, or the full context of others, refer directly to the specific contributions at the link referenced above (View Comments Submitted).

BJ submitted four individual comments which can be consolidated as follows:

The Maximal Starting Repertoire – [MSR-5 Overview and Rationale](#), Section 5.7.5, says “U+01C0 and U+01C1 are indistinguishable from the punctuation marks U+007C and U+2016 in certain user interface fonts....”

Yet the MSR-5 currently includes the following appear to have exactly the same conflict.

- Cyrillic Small Letter Palochka (U+04CF)
- Arabic Letter Alef (U+0627)
- CJK Unified Ideograph-4E28 (U+4E28)
- Latin Small Letter L (U+006C)

There should at least be a footnote indicating that an exception has been made in this case and explaining why it was made.

For letter small L the comment further notes: Since this symbol has been part of the Domain Name System from the beginning (e.g. .mil), it obviously must remain grandfathered in to the MSR.

Section IV: Analysis of Comments

General Disclaimer: This section intends to provide an analysis and evaluation of the comments submitted along with explanations regarding the basis for any recommendations provided within the analysis.

ICANN org thanks the contributor for their valuable input and feedback. The comment has been forwarded to the Integration Panel (IP) and the IP provides analysis as follows:

The code points cited as excluded from the MSR in the feedback, and the punctuation marks that they are indistinguishable from are:

- Latin Letter Dental Click (U+01C0) – Vertical Line (U+007C)
- Latin Letter Labial Dental Click – Double Vertical Line (U+2016)

The following shows all the code points mentioned in the feedback with typical glyphs:

2016 01C1 007C 01C0 4E28 0627 04CF 006C



As can be seen from the above, the excluded code points are identical in appearance and positioning on the line to the two punctuation marks. They are indistinguishable as claimed in the MSR-5 overview. The other four code points each have different length, width, or side

bearings (space around the “ink” in a glyph). In two cases, they also have variations in stroke thickness or termination.

While these four examples are highly similar, they are not indistinguishable from each other or from the Vertical Line (U+007C) punctuation mark. The IP therefore did not make an “exception”, given that the intent was to only exclude marks that are indistinguishable from punctuation marks. Neither was Latin Small Letter L (U+006C) singled out for “grandfathering”.

Nevertheless, a clarifying remark along these lines might be added to the Overview:

In particular, U+01C0 and U+01C1 are indistinguishable from the punctuation marks U+007C and U+2016 in certain user interface fonts. There are other code points with glyphs that look more or less like a straight line, but their glyphs show some variation in length, width and distance from the base line: the Integration Panel views the absolutely indistinguishable appearance as the relevant criterion in this instance. U+01C2 has a more distant resemblance to a line-drawing symbol U+256A; it is included here for consistency. U+01C3 is always indistinguishable from an exclamation mark.

Note that another member of the set, Latin Letter Retroflex Click (U+01C3) is also indistinguishable from a punctuation mark. Only the exclusion of Latin Letter Alveolar Click (U+01C2) is not based on the same principle, but on consistency with the other click letters, as noted already in the MSR.

The IP will update the text of the Overview to reflect this analysis.