# MAXIMAL STARTING REPERTOIRE VERSION 4 (MSR-4) FOR ROOT ZONE LABEL GENERATION RULES (RZ-LGR)

| | |
|---|---|
| **Publication Date:** | 23 January 2019 |
| **Prepared By:** | Sarmad Hussain (on behalf of the Integration Panel) |

<table>
<tr><td colspan="2"><b>Public Comment Proceeding</b></td></tr>
<tr><td>Open Date:</td><td>26 November 2018</td></tr>
<tr><td>Close Date:</td><td>10 January 2019</td></tr>
<tr><td>Staff Report Due Date:</td><td>25 January 2019</td></tr>
</table>

**Important Information Links**

[Announcement](#)
[Public Comment Proceeding](#)
[View Comments Submitted](#)

| | | | |
|---|---|---|---|
| **Staff Contact:** | Sarmad Hussain | **Email:** | sarmad.hussain@icann.org |

## Section I: General Overview and Next Steps

ICANN released for public comment version 4 of the Maximal Starting Repertoire (MSR-4: HTML, XML). This version is upwardly compatible with MSR-3 and adds three code points to the repertoire of Latin script and twelve code points to the repertoire of Myanmar script. Under the Procedure to Develop and Maintain Label Generation Rules for the Root Zone with Respect to IDN Labels, the MSR is the starting point for the work by community based Generation Panels developing the Root Zone Label Generation Rules (RZ-LGR) proposals for relevant scripts. The contents of MSR-4 and the detailed rationale behind its development are described in MSR-4 Overview and Rationale document. The Generation Panels currently use MSR-3, which covers 28 scripts: Arabic, Armenian, Bengali, Cyrillic, Devanagari, Ethiopic, Georgian, Greek, Gujarati, Gurmukhi, Han, Hangul, Hebrew, Hiragana, Kannada, Katakana, Khmer, Lao, Latin, Malayalam, Myanmar, Oriya, Sinhala, Tamil, Telugu, Thaana, Tibetan and Thai. MSR-4 will cover the same scripts, but adds 15 code points to contain 33,511 code points short-listed from 97,973 PVALID/CONTEXT code points of Unicode version 6.3. The Integration Panel will finalize the specification for MSR-4 based on the feedback received by the community. After the release of MSR-4, Generation Panels which are developing their RZ-LGR proposals will be able to use the updated contents as a starting point for their analysis.

## Section II: Contributors

*At the time this report was prepared, a total of two (2) community submissions had been posted to the forum. The contributors, both individuals and organizations/groups, are listed below in chronological order by posting date with initials noted. To the extent that quotations are used in the foregoing narrative (Section III), such citations will reference the contributor's initials.*

Individuals:

| Name | Affiliation (if provided) | Initials |
|---|---|---|
| Satish Babu | APRALO | SB |
| 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).*

SB1:
This comment requests that added code points be listed in a single place and rationale given.

BJ1:
The comment asks for additional clarification concerning the exclusion of Latin click letters in the range U+01C0-U+01C2 and suggests putative confusables that should be treated analogously as they are also vertical lines.

## 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.*

SB1: list additions explicitly
The Latin additions were listed already in the appropriate section; the Myanmar addition will likewise be listed in their appropriate section. The text of these sections addresses the rationale for inclusion.

BJ: clarify rationale for excluding click letters and vertical line examples
The Integration Panel (IP) reiterates that the reason for excluding the click letters is confusion with code points that are not PVALID and represent punctuation, e.g. U+007C VERTICAL LINE looks identical to U+01C0 LATIN LETTER DENTAL CLICK. Similar considerations apply to U+01C1, to a lesser degree to U+01C2, but the latter is best treated as part of a set, in the view of the IP.

The comment then notes that there are many PVALID code points that are depicted with a vertical line and that are not excluded. The argument made is that these should have been treated likewise, or that the security issues for all of them could be handled via a different mechanism, for example variants. For the MSR, there is usually no variant analysis performed, so that analysis is generally out of scope, except insofar as it might not be necessary to exclude a code point, if the LGR can handle security issues via the variant mechanism.

In the IP's analysis none of the example characters are suitable as variants for the click letter. In particular, the code point 4E28 (CJK Unified Ideograph) has a glyph that has wide left and right side bearings that introduce rather visible spacing on the left and right of the vertical line that makes up the ink in that glyph.

What makes the click letter a special concern is that it is often indistinguishable from a code point that is **not** PVALID, and therefore itself can never be a variant. Therefore, the only mitigation for this security issue is exclusion.

The comment also requests better documentation of the rationale and to this end the relevant text will be clarified.