

At-Large IDN TLDs - LGR Procedure Implementation - Maximal Starting Repertoire Version 2 Workspace

Comment Close Date	Statement Name	Status	Assignee (s)	Call for Comments Open	Call for Comments Close	Vote Open	Vote Close	Date of Submission	Staff Contact and Email	Statement Number
16.03.2015	IDN TLDs - LGR Procedure Implementation - Maximal Starting Repertoire Version 2	<div>ADOPTED</div> 9Y, 0N, 1A Reason of abstention > Rafid Fatani - I did not understand the Statement.	Satish Babu	03.03.2015	10.03.2015 23:59 UTC	11.03.2015 23:59 UTC	16.03.2015 20:00 UTC	16.03.2015	Sarmad Hussain sarmad.hussain@icann.org	AL-ALAC-ST-0315-02-00-EN

For information about this PC, please click [here](#) >>

- [Comments Forum](#)

Brief Overview

ICANN is releasing for public comment version 2 of the Maximal Starting Repertoire (MSR-2). This upwardly compatible version of the [MSR](#) adds six additional scripts to the repertoire. The MSR is the first deliverable under the "[Procedure to Develop and Maintain Label Generation Rules \(LGR\) for the Root Zone With Respect to IDN Labels](#)" [PDF, 772 KB] (the Procedure) and the starting point for the work by community based Generation Panels to develop their LGR proposals. The contents of MSR-2 and the detailed rationale behind its development are described in "[Maximal Starting Repertoire - MSR-2-Overview and Rationale](#)" [PDF, 1.12 MB]. Community members are invited to provide feedback on the contents of MSR-2.

Comment Period: 15 Dec 2014 23:59 UTC - 23 Feb 2015 23:59 UTC

Reply Period: 24 Feb 2015 23:59 UTC - 16 Mar 2015 23:59 UTC

Section I: Description, Explanation, and Purpose

As a deliverable under the "[Procedure to Develop and Maintain Label Generation Rules for the Root Zone With Respect to IDN Labels](#)" [PDF, 772 KB] (the Procedure), ICANN is providing for public comment version 2 of the Maximal Starting Repertoire (MSR-2). The contents of MSR-2 and the detailed rationale behind its development are described in "[Maximal Starting Repertoire - MSR-2-Overview and Rationale](#)" [PDF, 1.12 MB].

Summary of Contents

MSR-2 is a subset of IDNA 2008 PVALID code points for Unicode 6.3 (latest version of the Unicode Standard for which IANA provides IDNA 2008 tables). As stated in the Procedure, the code points included are not restricted from use in identifiers (as defined in Table 1 of UTS#39) and must not be used for writing an excluded script. The MSR was further adjusted by following the prescriptions of [the Procedure](#) [PDF, 772 KB] in eliminating code points not eligible for the root zone.

MSR-2 covers the following 28 scripts, of which six (marked with *) had been previously deferred: 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-2 contains 33,492 code points short-listed from 97,973 PVALID/CONTEXT code points of Unicode version 6.3.

The mere presence of a code point in the MSR does not indicate that the Integration Panel considers it acceptable for inclusion in the LGR. Where the Integration Panel was not able to resolve the status of a code point, it has tended to retain it in the MSR, with the aim of allowing Generation Panels to perform a more thorough review, and where appropriate to present a justification of the inclusion of such code points in the LGR.

In contrast, the absence of a code point affirms that the Integration Panel has determined that the code point is not appropriate for the DNS root, or, in certain situations, the panel has decided to defer it to a future version of the MSR.

Instructions for Reviewers

Reviewers are encouraged to carefully review the MSR-2 documents listed below. Communities that disagree with the choices that the Integration Panel has made in MSR-2 as presented here are advised to raise any issues during the public comment period, and to provide a rationale for adding or removing specific code points. Note that MSR-2 represents an upwardly compatible replacement of MSR-1, not merely an addition. As such reviewers are encouraged to comment on the entire content of MSR-2. In particular, communities for scripts that are about to begin work on an LGR for the first time are strongly encouraged to review the MSR. Note that MSR-2 did not add new content for the scripts that were already included in MSR-1. After the public comment period, the MSR-2 will be frozen for the purposes of developing the version of the Root Zone LGR based on it (LGR-1).

Future Development

The work that will be developed for integration in the first version of the LGR (LGR-1) will be based on MSR-2. If it becomes necessary to stage the release of the LGR, for example because not all Generation Panels are able to submit proposals at the same time, subsequent versions of the LGR may be released.

MSR-2 defers some code points that are already encoded in Unicode 7.0, because authoritative tables for IDNA 2008 are not yet available for Unicode 7.0. Unicode 8.0, due in 2015, is expected to further add code points that are potentially eligible for the root zone. In addition, the Integration Panel monitors any scripts not included in the MSR for indications that change in status is warranted. At regular annual intervals, another version of the MSR will be developed assuming that additional repertoire exists for which inclusion in the MSR is warranted. Until such a later version of the MSR is developed, MSR-2 would be the foundation for any LGR versions developed after its release.

All future versions of the MSR and all versions of the LGR must retain full backwards compatibility, such that they preserve the output of any label registration against the old LGR, when applied to an updated LGR or an LGR resulting from a later version of the MSR. Repertoire that has not been used for label registration is not required to be retained in future versions.

It is important to note that, while the expectation is that registrations predating the initial release of an LGR for the respective script will be allowed to remain in place even if they were to conflict, there is no requirement for an initial LGR to be compatible with them or to consider them precedents.

Section II: Background

To support IDN variants in the root zone, the ICANN community, at the direction of the Board, undertook several projects to study and make recommendations on their viability, sustainability and delegation. One of these projects is the implementation of [the Procedure](#) [PDF, 772 KB] allowing for the development of Label Generation Rules (LGR) for the Root Zone. The LGR for the Root Zone is a mechanism for creating and maintaining rules with respect to IDN labels for the root. This mechanism will be used to determine which Unicode code points are permitted for use in U-Labels in the root zone, what variants (if any) are allocatable and what variants (if any) are automatically blocked.

The MSR is the first deliverable from the Integration Panel under [the Procedure](#) [PDF, 772 KB] and will serve as a starting collection of code points from which Generation Panels may make a selection in constructing the repertoire for their respective LGR proposals. In accordance with [the Procedure](#) [PDF, 772 KB], "Generation panels must not include in their proposed repertoires any assigned code point that is not included in the maximal set of code points for the root zone defined by the integration panel."

Section III: Relevant Resources

[Maximal Starting Repertoire - MSR-2-Overview and Rationale.pdf](#) [PDF, 1.12 MB]

[MSR-2-Annotated-Han-Tables-20141114.pdf](#) [PDF, 43.3 MB]

[MSR-2-Annotated-Hangul-Tables-20141114.pdf](#) [PDF, 4.18 MB]

[MSR-2-Annotated-non-CJK-Tables-20141114.pdf](#) [PDF, 2.41 MB]

[MSR-2-Repertoire+WLE-Rules-20141204.xml](#) [PDF, 746 KB]

Section IV: Additional Information

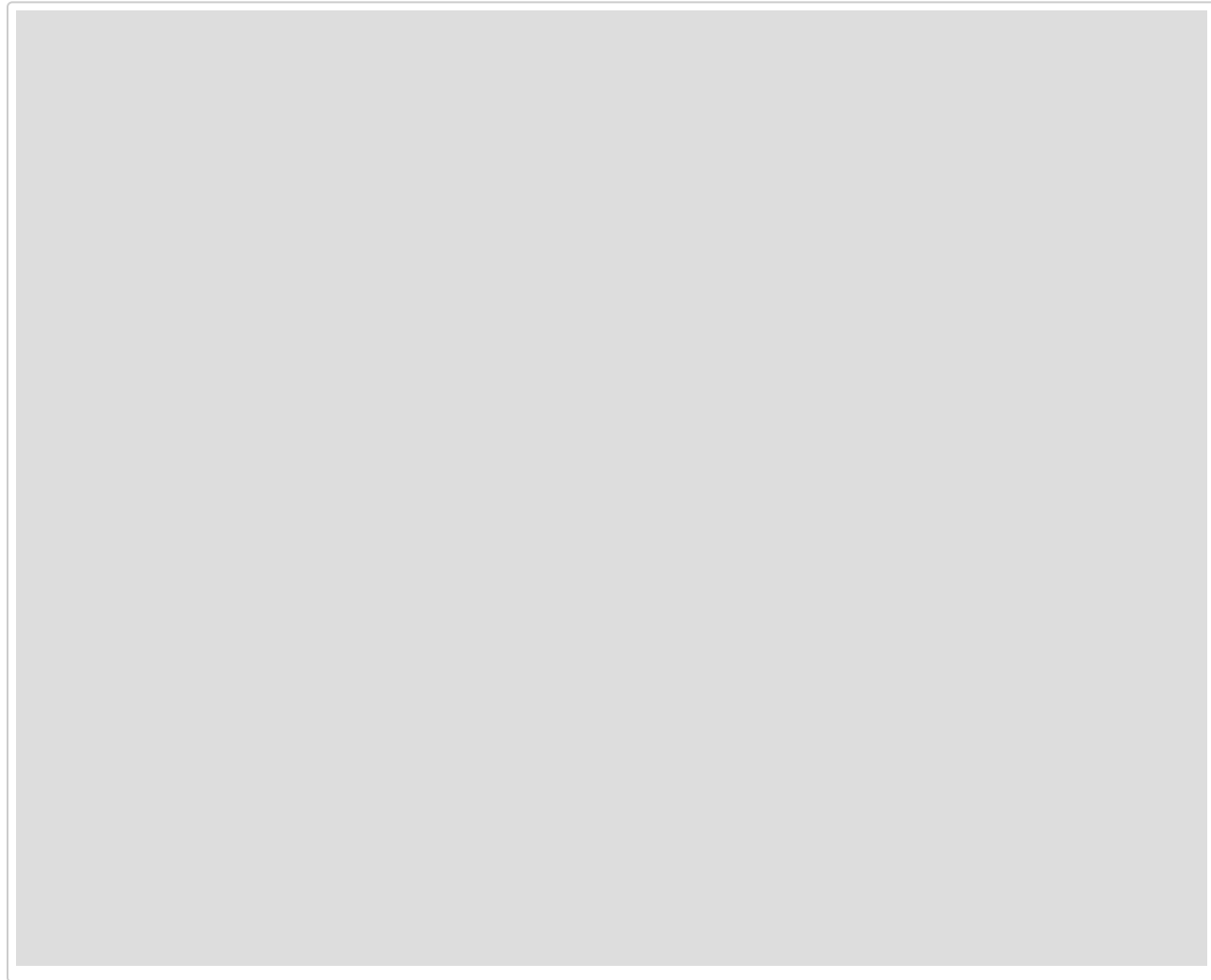
Section V: Reports

Staff Contact

Sarmad Hussain
sarmad.hussain@icann.org

FINAL VERSION TO BE SUBMITTED IF RATIFIED

Please [click here](#) to download a copy of the pdf document below.



FINAL DRAFT VERSION TO BE VOTED UPON BY THE ALAC

ALAC notes the progress made in the LGR Procedure Implementation project through the release of the Maximal Starting Repertoire Version 2 (MSR-2). The release of MSR-2 is an important step towards the creation of LGRs, which will, in turn, provide a consistent and predictable set of permissible code points for IDN TLDs, besides also providing ways to detect variant forms of labels. ALAC commends ICANN and the teams which have worked for this milestone.

MSR-2 adds six scripts (Armenian, Ethiopic, Khmer, Myanmar, Thaana and Tibetan) to the 22 scripts covered earlier in MSR-1, resulting in 33,492 valid Unicode code points (out of the total 97,973) that may be used by the Generation Panels (GPs). ALAC notes that the inclusion of these scripts is expected to benefit several million end-users of the Internet, particularly from Developing Countries.

ALAC also notes that while some of the GPs are seated and active, others have been less active or inactive, reflecting the varying nature of community participation in the program.

Recommendations

1. ICANN is in the process of rolling out a major initiative for Universal Acceptance (UAI) that is expected to be operational for the long-term. It is important that the IDN program is harmonized (in terms of parameters such as technology dissemination, capacity building and outreach) with the UAI so that the synergies between the two can be leveraged. In addition to such harmonization, ALAC also recommends that the UAI be utilized to ensure better community participation for the IDN program.
2. MSR-2 is based on Unicode 7, but is limited to the Unicode 6.3 subset. Given the fact that Unicode 8.0 is scheduled for release in 2015 (a draft has already been released), there may be questions from the community on the stability of the contents of MSR-2, particularly if the Generation Panels are to immediately commence their work based on MSR-2. ALAC recommends that ICANN clarifies the likely impact, if any, of changes to the underlying Unicode standard on MSR-2.
3. Once MSR-2 becomes operational and provides the basis of LGR-1, and once IDNs start getting registered, it would not be possible to change the once-registered names (or add more PVALID code-points to the MSR) without causing serious erosion of trust in the global Internet in general and IDNs in particular. This calls for sufficient interactions with communities to ensure that any contentious issues are addressed. For instance, *0D4C MALAYALAM VOWEL SIGN AU* is PVALID, but marked as 'Obsolete' in MSR-2, which has been pointed out by some users as not justified, as the sign in question is widely used. Whether this particular case is valid or not, it points to the need for more interactions. ALAC recommends extensive consultations with end-user and language communities to ensure that they get an opportunity to discuss the MSR-2 recommendations, as these have long-term ramifications.
4. Noting that end-user communities should be better integrated into the IDN Program, ALAC assures its support to the ICANN IDN team in stimulating participation of end-user communities. In particular, ALAC would welcome joint activities that involves At-Large Structures in the relevant geographies.

FIRST DRAFT SUBMITTED

ALAC notes the progress made in the LGR Procedure Implementation project through the release of the Maximal Starting Repertoire Version 2 (MSR-2). The release of MSR-2 is an important step towards the creation of LGRs, which will, in turn, provide a consistent and predictable set of permissible code points for IDN TLDs, besides also providing ways to detect variant forms of labels. ALAC commends ICANN and the teams which have worked for this milestone.

MSR-2 adds six scripts (Armenian, Ethiopic, Khmer, Myanmar, Thaana and Tibetan) to the 22 scripts covered earlier in MSR-1, resulting in 33,492 valid Unicode code points (out of the total 97,973) that may be used by the Generation Panels (GPs). ALAC notes that the inclusion of these scripts is expected to benefit several million end-users of the Internet, particularly from Developing Countries.

ALAC also notes that while some of the GPs are seated and active, others have been less active or inactive, reflecting the varying nature of community participation in the program.

Recommendations

1. ICANN is in the process of rolling out a major initiative for Universal Acceptance (UAI) that is expected to be operational for the long-term. It is important that the IDN program is harmonized (in terms of parameters such as technology dissemination, capacity building and outreach) with the UAI so that the synergies between the two can be leveraged. In addition to such harmonization, ALAC also recommends that the UAI be utilized to ensure better community participation for the IDN program.
2. MSR-2 is based on Unicode 7, but is limited to the Unicode 6.3 subset. Given the fact that Unicode 8.0 is scheduled for release in 2015 (a draft has already been released), there may be questions from the community on the stability of the contents of MSR-2, particularly if the Generation Panels are to immediately commence their work based on MSR-2. ALAC recommends that ICANN clarifies the likely impact, if any, of changes to the underlying Unicode standard on MSR-2.
3. Once MSR-2 becomes operational and provides the basis of LGR-1, and once IDNs start getting registered, it would not be possible to change the once-registered names (or add more PVALID code-points to the MSR) without causing serious erosion of trust in the global Internet in general and IDNs in particular. This calls for sufficient interactions with communities to ensure that any contentious issues are addressed. For instance, *0D4C MALAYALAM VOWEL SIGN AU* is PVALID, but marked as 'Obsolete' in MSR-2, which has been pointed out by some users as not justified, as the sign in question is widely used. Whether this particular case is valid or not, it points to the need for more interactions. ALAC recommends extensive consultations with end-user and language communities to ensure that they get an opportunity to discuss the MSR-2 recommendations, as these have long-term ramifications.
4. Noting that end-user communities should be better integrated into the IDN Program, ALAC assures its support to the ICANN IDN team in stimulating participation of end-user communities. In particular, ALAC would welcome joint activities that involves At-Large Structures in the relevant geographies.