

Report of Public Comments

Title:	LGR Procedure Implementation - Maximal Starting Repertoire Version 2		
Publication Date:	08 April 2015		
Prepared By:	IDN Program Team		
Comment Period:		Important Information Links	
Comment Open Date:	15 December 2014	Announcement	
Comment Close Date:	16 March 2015	Public Comment Box	
		View Comments Submitted	
Staff Contact:	Sarmad Hussain	Email:	sarmad.hussain@icann.org
Section I: General Overview and Next Steps			
<p>To support IDN labels in the root zone, the ICANN community, at the direction of the Board, undertook several projects to study and make recommendations on their viability and delegation. One of these projects is the implementation of the Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels (the Procedure) 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.</p> <p>In the context of the implementation of the Procedure, ICANN released the second version of the Maximal Starting Repertoire (MSR-2) for public comment. The MSR is the first deliverable from the Integration Panel under the Procedure and will serve as a fixed collection of code points from which Generation Panels may make a selection in constructing the repertoire for their respective LGR proposals.</p> <p>The 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.</p> <p>This release of MSR-2 sets the stage for the work by Generation Panels. In addition to selecting their repertoire from within the MSR for developing LGR proposals, Generation Panels will also evaluate whether any such code points are variants and if any rules are needed to further constrain the labels generated using these code points. The resulting LGR proposals by the Generation Panels will be released for public comment before they are reviewed by the Integration Panel for integration into the Root Zone LGR. 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.</p> <p>The review of scripts of eligibility is ongoing. While MSR-2 covers all scripts that have been eligible, another version of the MSR will be developed to include repertoires for newly eligible scripts, any extended repertoires of PVALID code points based on future versions of the Unicode Standard, as well as make any other additions to the repertoire as warranted. All future versions of the MSR and all</p>			

versions of the LGR must retain full backwards compatibility.

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.

Organizations and Groups:

Name	Submitted by	Initials
ALAC	ICANN Staff on behalf of ALAC	ALAC

Individuals:

Name	Affiliation (if provided)	Initials
Dr. Andreas Wetter	Institut für Asien- und Afrikawissenschaften Humboldt-Universität zu Berlin	AW

Section III: Summary of Comments

General Disclaimer: This section is intended to broadly and comprehensively summarize the comments submitted to this Forum, but not to address every specific position stated by each contributor. Staff 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).

Summary of ALAC comments

1. The ALAC notes that the inclusion of the six scripts added in MSR-2 is expected to benefit several million end-users of the Internet, particularly from Developing Countries. The ALAC also notes that while some of the GPs are seated and active, others have been less active or inactive.
2. It is important that the IDN program is harmonized (in terms of parameters such as technology dissemination, capacity building and outreach) with the UAI. The ALAC also recommends that the UAI be utilized to ensure better community participation for the IDN program.
3. 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, 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. The ALAC recommends that ICANN clarifies the likely impact, if any, of changes to the underlying Unicode standard on MSR-2.
4. 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. The ALAC recommends extensive consultations with end-user and language communities to discuss the MSR-2 recommendations, as these have long-term ramifications.

5. The ALAC assures its support to the IDN team in stimulating participation of end-user communities. The ALAC would welcome joint activities that involve At-Large Structures in relevant geographies.
6. 0D4C ፬፯ 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.

Summary of AW comments

The comments contain an analysis of the use of the Ethiopic script for various languages.

1. Me'en language — There does not seem to be any usage of the Ethiopic script for the language. Therefore the signs 2D80 to 2D92 are not used by anyone currently.
2. Blin language — The language has been written in the Ethiopic script for more than a century, and is still written that way in the diaspora, while the government decided to introduce a Latin based orthography instead. Examples from other languages (e.g. Harari) show that the use of scripts can change quite rapidly. The probability that the Ethiopic script will be used for Blin in the future is therefore relatively high. Code points 2D93 to 2D96, are actually used in the official orthographies of two Agaw languages of Ethiopia which are not mentioned in the Unicode, namely Awngi and Khamtanga and should be included.
3. Sebatbeit language -- The signs so labeled in Unicode are not in use according to specialists of Addis Ababa University. Instead, redundant signs which already exist in the standard set of Ethiopia are used, and one additional new set has been created, which isn't yet part of Unicode.
4. Gamo-Gofa-Dawro – Those languages are written with a Latin-based orthography.
5. Basketo -- The signs for Basketo can be confirmed. But the signs AB11 to AB16 aren't used.

Section IV: Analysis of Comments

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

Response to ALAC-1, 2 and 5

These comments do not directly address the contents of the MSR. They have been noted. The support extended by ALAC to get community involved in the IDN program is appreciated and staff will continue to work with ALAC to support the process. IDN Program staff is involved in the Universal Acceptance Initiative. IDN Program staff will further follow up with ALAC to get further guidance on their recommendation for further harmonization of the two programs.

Response to ALAC-3

The difference between Unicode 7.0 and 6.3 is limited, and therefore the impact is likewise limited. In fact, continuing to base the MSR on 6.3 is the best way to guarantee that LGRs are the most stable. Nevertheless, the ICANN recommends that GPs take into account both pending and future addenda to the Unicode Standard and review them for potential forward compatibility issues, so that LGRs based on Unicode 6.3 can be updated to later versions of Unicode without problems.

Response to ALAC-4

Existing domain names must not be changed, but each iteration of IDNA and each iteration of the MSR will likely result in the addition of some new PVALID code points. Not all code points in the MSR are expected to make it into the Root Zone LGR. In fact, Generation Panels are advised to be conservative in their inclusion, leading to the expectation that many code points in the MSR will not, in fact, be included after GP review.

Likewise, if future iterations of the MSR add additional code points, GPs are advised to only include them in future editions of the LGR, if they pass a stringent review for compatibility issues in relation to the already existing Root Zone repertoire. The Generation Panels are the means, under the procedure, by which the end-user communities interact with the process.

Response to ALAC-6

Typically, in cases of reasonable doubt, the MSR errs on the side of including a code point, so as to allow the resolution of the status of the code point as part of the process. In this particular case there are additional issues beyond the degree of usage.

The Unicode Standard and other sources on Malayalam mention this vowel sign not only as “archaic”, but as superseded by another code point. A typical explanation of Vowel Sign Au, reads:

“Malayalam has two forms for Vowel Sign Au. The ancient form, which has vowel markers on both side of the consonant, as in കൗ ‘kau’. This is encoded in Unicode as കൗ U+0D4C MALAYLAM VOWEL SIGN AU. However, this is an archaic and obsolete method. The modern form has only one vowel marker, to the right side of the consonant, as in കൗ ‘kau’. This modern form is encoded as കൗ U+0D57 MALAYALAM AU LENGTH MARK.”

If the code point 0D4C was permitted in the Root Zone, it would represent a dual encoding that, at the minimum, would need to be handled by a variant relation with U+0D57. It would be unlikely that any dual representation (particularly if it includes an archaic form) would be acceptable for the root. It is best not to introduce such a risk into the root zone LGR to begin with. Therefore the conservative response must be to not admit the code point into the MSR, unless it is unavoidable. Finally, it is worth noting that there is a difference between a claim of a code point reportedly being ‘widely’ used and evidence that it is needed in modern writing, which is the criterion used for inclusion in the MSR.

Response to AW-1

The wording of comment implies some uncertainty. Given that changes to the use of Latin are recent enough that the situation may not necessarily be settled definitely, the best approach may be to retain these to allow an Ethiopic Generation Panel to provide further review. The code point 2D80 to 2D92 do not present any particular risks, such as confusability or dual encoding and can therefore remain, but with a recommendation to the GP to not include them in the LGR unless clear evidence of the required level of use can be supplied.

Response to AW-2

Code points 2D93-2D96 will be added to allow the Ethiopic Generation Panel to consider them support of official orthographies for the Awngi and Khamtanga languages.

Response to AW-3

No change needed for MSR-2. If the missing sign gets added to Unicode, it can be considered for a later version of the MSR.

Response to AW-4

Remove AB11-AB16 as they are only used in religious context and both current experts and written sources going back 20 years indicate that the main script in use has been Latin.

Response to AW-5

The ranges AB01-AB06 and AB09-AB0E are retained based on their use by Basketo.

The contents of the MSR-2 will be updated to reflect the accepted recommendations from these comments, as well as any further input received from expert advisors.