

# Proposal for the Generation Panel for the Myanmar Script Label Generation Ruleset for the Root Zone

## 1 General Information

### 1.1 Background

The letters of the alphabet used in Myanmar script are derived from the Brahmi script which flourished in the Indian subcontinent between 5<sup>th</sup> Century B.C and 3<sup>rd</sup> Century A.D. Existing records of original Myanmar script from votive tablets, terracotta plaques, stone inscriptions, painted glosses on murals, palm leaf inscriptions, bell inscriptions and inscriptions made on writing tablets made of handmade paper, gold or copper foil, show that earliest surviving records can be dated as far back as 5<sup>th</sup> Century of the Myanmar Era (12<sup>th</sup> Century AD).

The Myanmar script is used for a number of languages. This means that when considering the script as a whole, care must be taken not to over specify constraints on what character sequences should be considered valid or in error. The temptation is to use script level sequence constraints as a form of spell checking. But spell checking is inherently language specific. The result is that script constraints need to be the lowest common denominator of all the orthographies supported by the script. The orthography list is not closed: we have not described all the existing orthographies yet; languages change and develop and their orthographies with them. As a result, script constraints cannot simply be the intersection of all known writing system constraints, but must take a more intentional approach. The basic principle used here is not to try to constrain what users can generate, but to ensure that the label generation rulesets enable Myanmar scripts label in the Root Zone and maintain the security and stability of the Domain Name System using a conservative mechanism, as specified by the procedure to develop Label Generation Rules (LGR) for the Root Zone. E.g. the historical characters are not allowed, any variant characters are identified, and any ill-formed labels based on script-based rules are excluded. We do this by specifying a valid string as being a sequence of slots. Each slot may be empty or contain a character (or sequence as specified by the slot). Implementations may well add further, language specific, constraints to help their users.

A further concern when reading a developing document such as this is the stability criteria. What can we be sure about going into the future? The approach taken in this document follows the core principle of stability in Unicode: Any valid data today will always remain valid. This requires that any changes to the sequence order, for example, will always be to loosen it. Thus more sequences will be allowed rather than less. This means that invalid data today may not always remain invalid in future versions of this document. It should also be borne in mind that while the unity of the script as a whole may well be affected by the addition or changes in a single language, each language stands alone in its encoding and needs its own consistency. Care is taken that any changes that a difference in language may cause on the script as a whole (adding more legal

sequences), do not cause any changes in other language encodings. This may result in some decisions made for a particular language, looking different from those for another language and the temptation to try to over unify languages should be avoided.

## 1.2 Target Script

As per the Procedure to Develop and Maintain the Label Generation Rules for the DNS Root Zone in Respect of IDNA Labels (referred to simply as [Procedure] in the following), only code points included in Maximal Starting Repertoire [MSR-3] will be considered.

The Myanmar script has the following specifications:

ISO 15924 code: Mymr

ISO 15924 no.: 350

English Name: Myanmar (Burmese)

The complete set of characters in the Myanmar script fall in the following Unicode ranges:

Myanmar: U+1000 – U+109F

## 1.3 Principal Languages Using the Script

The languages that currently use Myanmar script and which either have EGIDS scale from 1 to 4, or have EGIDS 5 languages with substantive users, will be included in the analysis of the Generation Panel (GP), for example, the languages listed in Table 1. The final list of languages covered will be finalized in the LGR Proposal.

Language	ISO 639-3 Code(s)	Countries	Local Name of the Script	EGIDS Scale	Total Users in All Countries
Burmese	[mya]	Myanmar	မြန်မာ	1	42,906,490
Shan	[shn]	Myanmar, China, Thailand	လိၵ်ႈတႆး	3	3,295,000.
Karen, Sgaw	[ksw]	Myanmar, Thailand	စီၤ	3	1,560,000
Mon	[mnw]	Myanmar, Thailand	မန်	5	851,000.

Table 1: Sample list of languages in consideration

## 1.4 Related Scripts

Malayalam, Kannada, Sinhala and Telugu scripts can in some cases be similar to the Myanmar script block. Therefore, Myanmar GP will coordinate with the Neo-Brahmi and Sinhala GPs as needed, to undertake its work.

## 2 Proposed Initial Composition of the Panel

### 2.1 List of Prospective Chair and Members

The Myanmar Script GP comprises of experts in policy, linguistics, DNS, IDNA, Unicode and Myanmar language user community. The composition may change over time as new members join the GP.

No.	Name	Role	Designation	Organization	Expertise
1	Dr. Myint Myint Than	Advisor	Executive Director & member of Myanmar NLP Working Committee	Myanmar Computer Federation	Unicode & Policy
2	Dr. Khin Aye	Advisor	Member	Myanmar Language Department	Linguistics
3	Mr. Ngwe Tun	Advisor	Director (Technical)	Trust Link Co. Ltd.	Unicode
4	Mr. Naing Win Oo	Member	DNS and System Expert	Yatanarpon Teleport Public Co. Ltd. (ISP)	DNS & IDNA
5	Ms. Thin Zar Phy	Chair	NLP researcher	Techno Zenith (IT Solution)	Unicode
6	Mr. Kaung Khant Zaw	Member	CEO	Techno Zenith (IT Solution)	Software Development and Unicode
7	Ms. Yin May Oo	Member	Digital Linguist	Teledirect, Singapore	Computational Linguistics
8	Mr. Ye Zarni Aung	Member	Language Experts	Myanmar Unicode Area	Unicode and IDNS
9	Mr. Min Paing Khant Oo	Member	Network Engineer	MyHongsa Telecom Office	Unicode and IDNS
10	Mr. Thura Soe	Member	Software Engineer	Techno Zenith (IT Solution)	Software Development and NLP
11.	Mr. Sai Zin Di Di Zone	Member	Editor in charge	Hsenpai News journal	Shan Language expert

### 3 Work Plan

#### 3.1 Suggested Timeline with Significant Milestones

The Myanmar GP intends to follow the plan provided below.

Activity	Description	Start Date	Duration
Start writing the introduction part of the proposal	Start a Google Document as a shared working space	16-Jun-18	2days
Develop principles	Principles to be used to determine code points to be included, variants code point and variant labels	18-June-18	4 weeks
Determine code points	Select the code points from MSR which are needed for Root Zone LGR (Write up the relevant part of the Root Zone LGR proposal + XML + Test file)	16-July-18	2 weeks
Determine (any) variants	From the codes points selected, determine if the end-user may confuse two code points (Write up the relevant part of the Root Zone LGR proposal + XML + Test file)	30-July-18	4 weeks
Determine label level rules	Determine if there are any label level constraints on the use of selected code points (Write up the relevant part of the Root Zone LGR proposal + XML + Test file)	27-Aug-18	5 weeks
Public Consultation	Seminar/meeting with community	1-Oct-18	1 day
Submit	Finalize and submit for Public Comments by ICANN	8-Oct-18	1 week

#### 3.2 Proposed Schedule for Teleconferences

Most of the work will be accomplished through the Myanmar GP email list [myanmargp@icann.org](mailto:myanmargp@icann.org) and face-to-face meetings. The GP will be holding regular teleconferences every two weeks, for the period of the work. In addition, the GP will also organize face-to-face meetings, tentatively during the discussions on variants (13-Aug-18) and then towards the end of the work.

### 3.3 Source of Funding and Logistics

Foremost, all the members of the GP will be volunteering their time for this purpose. ICANN will support the online logistics of the group (e.g. hosting the conference calls, assisting in GP coordination, developing the wiki page for posting information, maintaining the mailing list, etc.).

Members of the working group are encouraged to find sources of funding to attend face-to-face meeting(s) related to the MM-GP. ICANN, might be able to fund a limited number of those members who are active in the proceedings of the working group to present the MM-GP work progress at ICANN meeting(s), and will also provide support for remote participation for the members not able to attend the meeting(s) in person.

### Other Information

Following references will be used as the base to develop the Myanmar LGR Proposal.

#### References:

1. <http://www.ethnologue.com/country/MM/status>
2. <http://www.unicode.org/notes/tn11/tn11-4.html>
3. <http://www.mlc.gov.mm/index.php?page=bWVudWRldGFpbCZpZD03JnNpZD0xOA%3D%3D>