

Myanmar, Neo-Brahmi and Sinhala GPs Meeting with Integration Panel

Thin Zar Phyo
Myanmar GP Chair

Ajay Data
NBGP Co-chair

Harsha Wijayawardhana
Sinhala GP Co-Chair

Agenda

- ⦿ Myanmar script LGR proposal
- ⦿ Neo-Brahmi script LGR proposals
- ⦿ Sinhala script LGR proposal
- ⦿ Joint discussion with the Integration Panel

Myanmar Generation Panel and Integration Panel Discussion

Thin Zar Phyo
Myanmar GP Chair

Agenda Overview

1

Scripts Covered and
Where They are
Used

2

Summary of
Myanmar LGR
Proposal

3

Discussion with IP

Myanmar GP – Languages Using Myanmar Script

- Myanmar script have been derived from the Brahmi script which has flourished in the Indian subcontinent between 5th Century B.C and 3rd Century A.D.
- Myanmar Script is written from left to right and requires no spaces between words
- Languages covered by the LGR:

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
Rakhine	[rki]	Myanmar	ရခိုင်	3	2,020,000
Karen, Sgaw	[ksw]	Myanmar, Thailand	ꨀꨃ	3	1,560,000
Mon	[mnw]	Myanmar, Thailand	မန်	5	851,000
Pa'O Karen	[blk]	Myanmar	ပအိုဝ်း	5	560,740

Code Point Repertoire

- Starting from MSR-3, the repertoire includes:
 - 87 code points
 - 15 sequences
 - 12 code points identified to proposed for next version of MSR
- The repertoire excludes:

#	Unicode Code Point	Glyph	Character Name	Reason for exclusion
1	108B	◌◦	Myanmar sign Shan Council Tone-2	Never used in colloquial Shan
2	108C	◌◦◦	Myanmar sign Shan Council Tone-3	Never used in colloquial Shan
3	108D	◌◦◦ —	Myanmar sign Shan Council Emphatic Tone	Never used in colloquial Shan

In-Script Variant Analysis

- Myanmar GP defines the following are in-script variant code points due to the nearly identical glyph

Set#	Unicode Code Point	Glyph	Unicode Code Point	Glyph
1	u1023	ꨀ	u1000 + u1039 + u1000	ꨀ
2	u1029	ꨁ	u101E + u103C	ꨁ
3	u102A	ꨂ	U1029 + u1031+u102C+u103A	ꨂ
4	u102A	ꨂ	u101E+ u103C + u1031+ u 102C+ u103A	ꨂ
5	u1061	ꨆ	u101B + u103E	ꨆ
6	u1009 + u103A	ꨃ	u1025 + u103A	ꨃ
7	u1009 + u1037 + u103A	ꨃ	u1025 + u1037 + u103A	ꨃ
8	u107E	ꨏ	u107D + u103E	ꨏ

Cross-Script Variant Analysis

⦿ Myanmar-Malayalam

No.	Glyph	Code Point	Myanmar Character Name	Glyph	Code Point	Malayalam Character Name
1	၀	U+1002	MYANMAR LETTER GA	റ	U+0D31	MALAYALAM LETTER RRA
2	၁	U+101D	MYANMAR LETTER WA	ത	U+0D20	MALAYALAM LETTER TTHA

⦿ Myanmar-Oriya

No.	Glyph	Code Point	Myanmar Character Name	Glyph	Code Point	Oriya Character Name
1	၁	U+101D	MYANMAR LETTER WA	ତ	U+0B20	ORIYA LETTER TTHA
2	၆	U+1031	MYANMAR VOWEL SIGN E	ଌ	U+0B47	ORIYA VOWEL SIGN E

Confusable Code Point Analysis (1/2)

⊙ In-Script confusable code points

No.	Glyph	Code Point	Glyph	Code Point	Note
1	၈	u1008	၈	U+0D31	The sequence U+1005 U+103B is invalid
2	၉	u1009 + u102C	၉	U+0D20	The sequence u1025+ u102C is invalid

⊙ Myanmar-Malayalam confusable code points

No.	Glyph	Code Point	Myanmar Character Name	Glyph	Code Point	Malayalam Character Name
1	က	U+1000	MYANMAR LETTER KA	ന	U+0D28	MALAYALAM LETTER NA
2	ഖ	U+101A	MYANMAR LETTER YA	ഘ	U+0D27	MALAYALAM LETTER DHA
3	က	U+1000, U+102C	MYANMAR LETTER KA , MYANMAR VOWEL SIGN AA	ന	U+0D28, U+0D4D, U+0D28	MALAYALAM LETTER NA, MALAYALAM SIGN VIRAMA, MALAYALAM LETTER NA

Confusable Code Point Analysis (2/2)

⦿ Myanmar – Sinhala confusable code points

No.	Glyph	Code Point	Myanmar Character Name	Glyph	Code Point	Sinhala Character Name
1	꠵	U+1025	MYANMAR LETTER U	උ	U+0D8B	SINHALA LETTER UYANNA
2	꠵꠴	U+102C	MYANMAR VOWEL SIGN AA	උඳ	U+0DCF	SINHALA VOWEL SIGN AELA-PILLA
3	꠵꠴꠴	U+1031	MYANMAR VOWEL SIGN E	උඳඳ	U+0DD9	SINHALA VOWEL SIGN KOMBUVA
4	꠵꠴꠴꠴	U+1031, U102C	Sequence: MYANMAR VOWEL SIGN AA, MYANMAR VOWEL SIGN E	උඳඳඳ	U+0DDC	SINHALA VOWEL SIGN KOMBUVA HAA AELA-PILLA

Whole Label Evaluation Rules

⦿ Code point categories

C	→	Consonant
IV	→	Independent Vowel
DVS	→	Dependent Vowel Sign
ANUSVARA	→	1036 ँ
T_SHORT	→	1037 ऌ
T_LONG	→	1038 ऍ
K	→	Killer: 103A ऌ
VIRAMA	→	1039
M	→	Dependent Consonant Sign
C1	→	103F ೀ
LV	→	Long Vowel: 102B, 102C, 102E, 1030, 1031, 1032, 1036
LVS	→	Long Vowel Sequence: 102D+102F, 1031+102B, 1031+102C , 102F+1036
SV	→	Short Vowel: 102D, 102F

Whole Label Evaluation Rules

⦿ Whole Label Evaluation Rules

1. [DVS or ANUSAVARA] must follow C or M_set
2. M_set must follow C
3. (C+ K) or (C +1037+K) must follow C or [C+M_Set]
4. C1 must follow C
5. 1025 ρ can only be followed by 102F \$ \circ_L + (1036 \$ or 1038 \$ $\circ_:$)
6. 1026 ρ can only be followed by 1038 \$ $\circ_:$
7. T_SHORT must not follow M or 102D or 102F
8. T_SHORT can follow S12
9. T_LONG must not follow C or M or 102D or 102F
10. T_LONG can follow S12
11. K must follow [2C or 1023 ρ] or 1037 \$ \circ
12. S11 must be followed by C
13. VIRAMA cannot be at the end
14. T_LONG must follow LV or S12 or S13 or 103A
15. T_SHORT must follow LV or S12 or S13 or S14 or S15 or 1038

Timeline and Next Step



IP Discussion – Proposed code points for MSR-4

#	Unicode Code Point	Glyph	Character Name	Category	Language(s), with EGIDS Value
1	U+1028	ꨂ	MYANMAR LETTER MON E	Independent Vowel	5 Mon
2	U+1033	ꨃ	MYANMAR VOWEL SIGN MON II	Dependent Vowel Sign	5 Mon
3	U+1034	ꨄ	MYANMAR VOWEL SIGN MON O	Dependent Vowel Sign	5 Mon
4	U+105A	ꨆ	MYANMAR LETTER MON NGA	Consonant	5 Mon
5	U+105B	ꨇ	MYANMAR LETTER MON JHA	Consonant	5 Mon
6	U+105C	ꨈ	MYANMAR LETTER MON BBA	Consonant	5 Mon
7	U+105D	ꨉ	MYANMAR LETTER MON BBE	Consonant	5 Mon
8	U+105E	ꨊ	MYANMAR CONSONANT SIGN MON MEDIAL NA	Medial	5 Mon
9	U+105F	ꨋ	MYANMAR CONSONANT SIGN MON MEDIAL MA	Medial	5 Mon
10	U+1060	ꨌ	MYANMAR CONSONANT SIGN MON MEDIAL LA	Dependent Consonant Sign	3 S'gaw Karen, 5 Mon
11	108F	ꨍ	MYANMAR SIGN RUMAI PALAUNG TONE-5	Tone mark	5 Pa'O Karen
12	AA7B	ꨎ	MYANMAR SIGN PAO KAREN TONE	Tone mark	5 Pa'O Karen

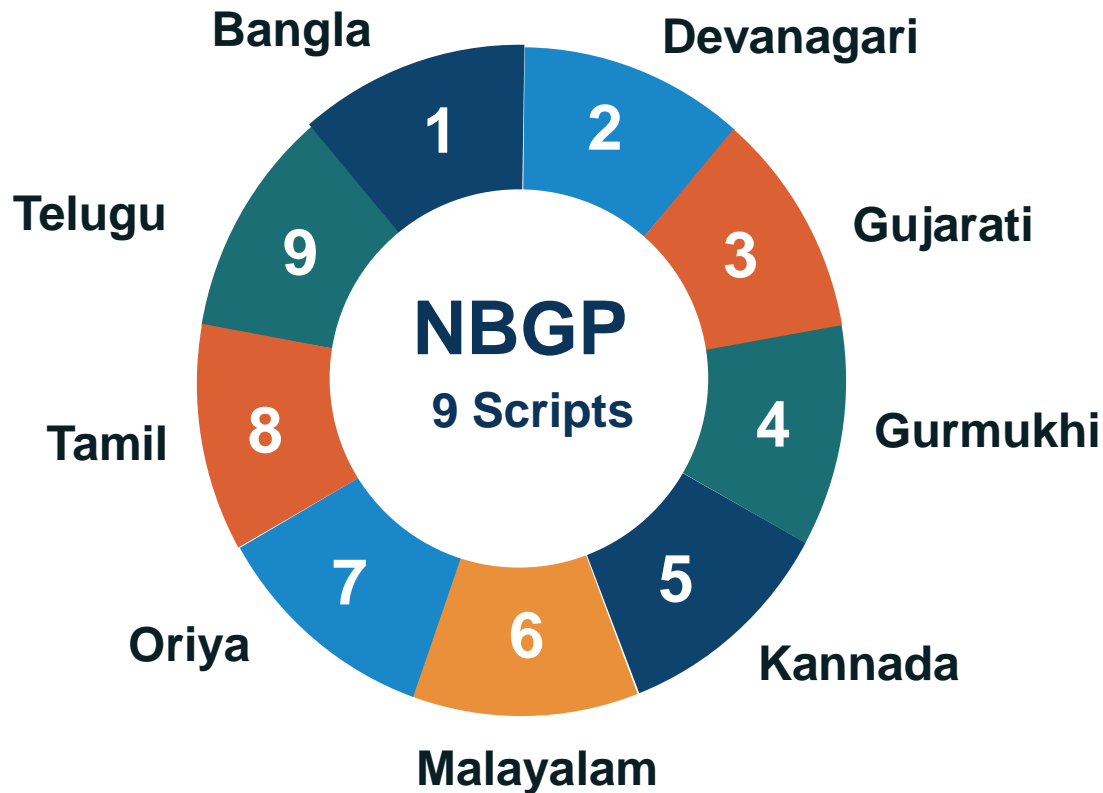
Neo-Brahmi Generation Panel (NBGP) and Integration Panel Discussion

Ajay Data
NBGP Co-chair

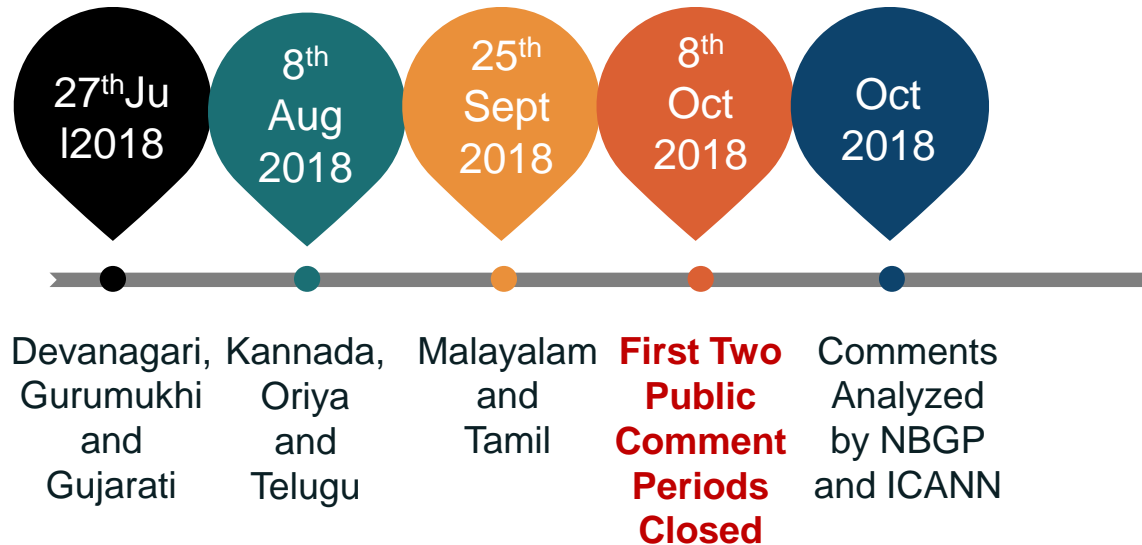
Members

Co-chairs: Dr. Ajay Data, Dr. Mahesh D. Kulkarni, Prof. Udaya Narayana Singh

Members: 60+ members from India, Nepal, Sri Lanka & Bangladesh



Public Comment Announcements



Summary

Google ICANN DEVANAGRI public comment and you will get the first link for public comment. Or visit <https://www.icann.org/public-comments>

Discussion with IP

- ⦿ Public Comment Feedback
 - Different views on Nukta inclusion in Devanagari
 - Different views on Nukta usage in Oriya
 - Cross-script variants between Devanagari and Bangla script needs to be completed
 - Cross-script variants between Oriya – Myanmar and Malayalam - Myanmar
- ⦿ Next steps after public comment of LGR proposals for Devanagari, Gurmukhi, Gujarati, Kannada, Telugu, Oriya, Malayalam, and Tamil
- ⦿ Cross-script variant analysis with Myanmar script and Sinhala script

Sinhala Generation Panel and Integration Panel Discussion

Harsha Wijayawardhana
Sinhala GP Co-Chair

Agenda Overview

1

Scripts Covered and
Where They are
Used

2

Summary of the
LGR Proposal

3

Discussion with IP

Sinhala GP – Languages Using Sinhala Script

- ⦿ Sinhala Script is primarily used in Sri Lanka to write Sinhala language which belongs to Indo-European Language family and Indo-Arya Sub family
- ⦿ The Script is Abugida Script which sprang from family of Southern Brahmi Script to which Telugu, Malayalam and Tamil belong to
- ⦿ Languages covered by the script:
 - Sinhala
 - Pali
 - Sanskrit

Code Point Repertoire

- Starting from MSR-3, the repertoire includes:
 - 72 code points
 - 4 sequences
- The repertoire excludes:

#	Unicode Code Point	Glyph	Character Name	Reason for exclusion
1	0D8E	සෘෘ	SINHALA LETTER IRUUYANNA	Usage unknown
2	0D8F	ඌ	SINHALA LETTER ILUYANNA	Usage unknown
3	0D90	ඌඌ	SINHALA LETTER ILUUYANNA	Usage unknown
4	0D9E	ඛ	SINHALA LETTER KANTAJA	Not in modern usage
5	0DA6	ඌ	SINHALA LETTER SANYAKA	Only used in the word 'ඌඌ' (this word is used to call dogs)
6	0DDF	ඌ	SINHALA VOWEL SIGN GAYANUKITTA	Usage unknown
7	0DF3	ඌ	SINHALA VOWEL SIGN DIGA GAYANUKITTA	Usage unknown

In-Script Variant Analysis

- Sinhala GP decided the following are in-script variant code points due to similar shapes and characters which could be used interchangeably
 - ස (U+0DC3) and ස (U+0D9D)
 - බ (U+0DB6) and බ (U+0D9B)
 - භ (U+0DC4) and භ (U+0DB7)
 - ච (U+0DA0) and ච (U+0DC0)
 - ඔ (U+0D94) and ඔ (U+0DB9)
 - ඵ (U+0D91) and ඵ (U+0DB5)
 - සා (U+0D8D) and සා (U+0DC3 U+0DD8)
 - ඔඵ (U+0D93) and ඔඵ (U+0DB5 U+0DD9)
 - ඵ් (U+0D92) and ඵ් (U+0DB5 U+0DCA)
 - ඔ් (U+0D95) and ඔ් (U+0DB9 U+0DCA)

Cross-Script Variant Analysis

- Sinhala GP concluded there is no cross-script variant rules
- Following are confusable cases
 - U+0D82 (SINHALA SIGN ANUSVARAYA, ീ)

Sinhala	Telugu	Kannada	Malayalam
ീ (U+0D82)	ీ (U+0C02)	ೀ (U+0C82)	ഀ (U+0D02)

- U+0D83 (SINHALA SIGN VISARGAYA, ു)

Sinhala	Devanagari	Gujarati	Telugu	Kannada	Malayalam
ു (U+0D83)	ः (U+0903)	ઃ (U+0A83)	ృ (U+0C03)	ೃ (U+0C83)	ഃ (U+0D03)

Cross-Script Variant Analysis

- Following are confusable cases (cont.)

- Sinhala and Malayalam

Sinhala	Malayalam
ඞ (U+0D9C)	ඞ (U+0D17)
ඞ (U+0DC1)	ඞ (U+0D36)
ඞ (U+0DCF)	ඞ (U+0D3E)

- Sinhala and Myanmar

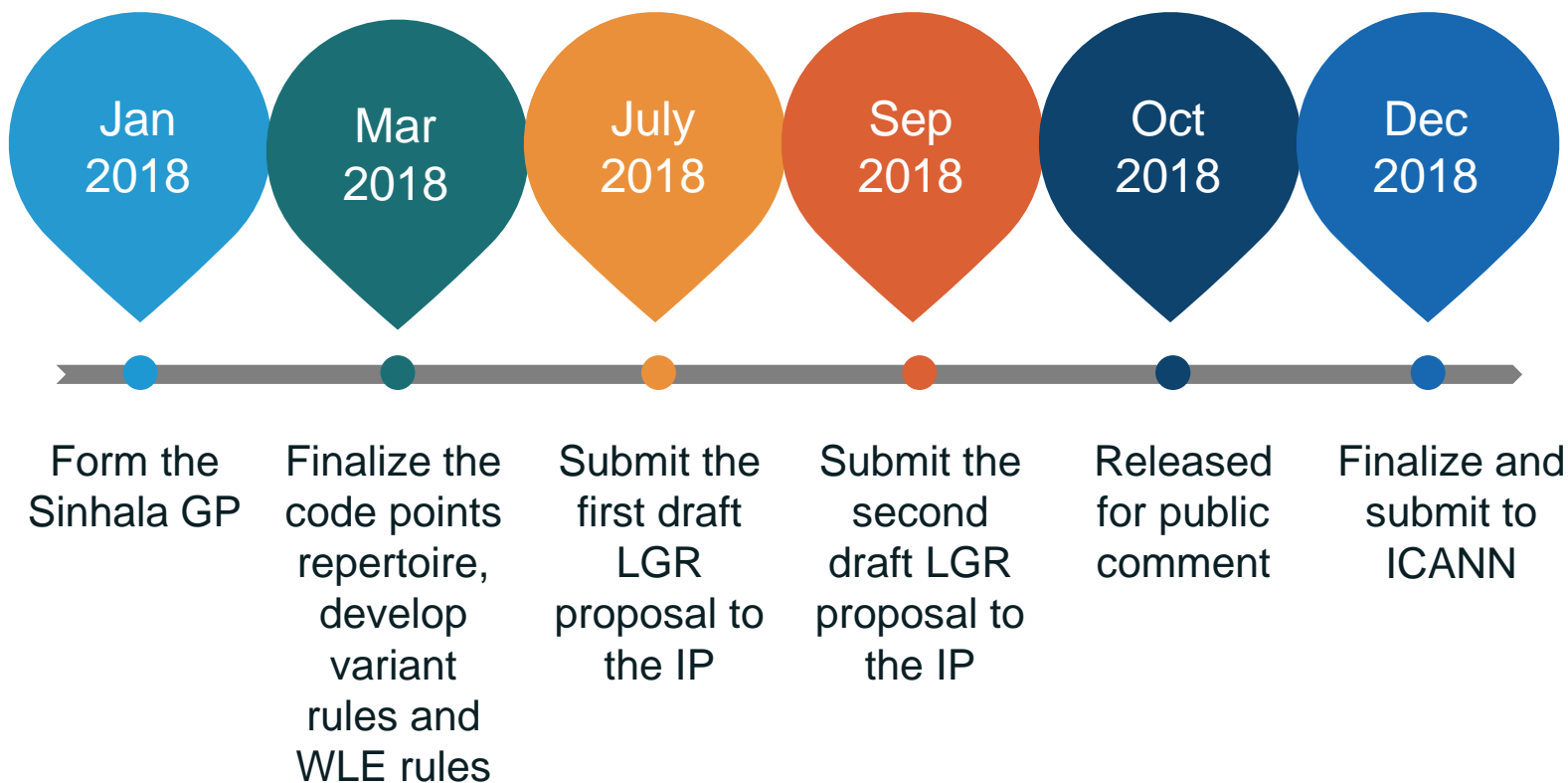
Sinhala	Myanmar
ඞ (U+0D9C)	ၞ (U+1010)
ඞ (U+0DC1)	ၞ (U+107B)

Whole Label Evaluation Rules

- ⊙ Code point category
 - C → Consonant
 - V → Vowel
 - M → Matras / Vowel Signs
 - B → Anusvara (Bindu)
 - X → Visarga
 - H → Halanta / Virama
 - J → Sannjakas

- ⊙ Whole Label Evaluation Rules
 - H: must be preceded by C
 - M: must be preceded by C or J
 - X: must be preceded by either V, C, or M
 - B: must be preceded by either V, C, J or M

Timeline and Next Step



Sinhala GP and IP Discussion

- ⦿ Inclusion of the additional code points

Joint Discussion