

ICANN
COMMUNITY FORUM

64

KOBE

9–14 March 2019



IDN Program Update



ICANN64
13 March 2019

Overview of Session Presentations

- ⦿ IDN Program Overview and Progress - Sarmad Hussain
- ⦿ Update by Integration Panel - Wil Tan
- ⦿ Community Updates
 - Hebrew GP - Yoram Hacoen
 - Latin GP - Michael Bauland,
William Jouris
 - Neo-Brahmi GP - Udaya Narayana Singh
 - Myanmar GP - Yin May Oo
- ⦿ Q/A

IDN Program Overview and Progress

Sarmad Hussain
Director, IDN Programs

Enable deployment of domain names
in the local languages and scripts
used by the communities globally
in a secure and stable manner

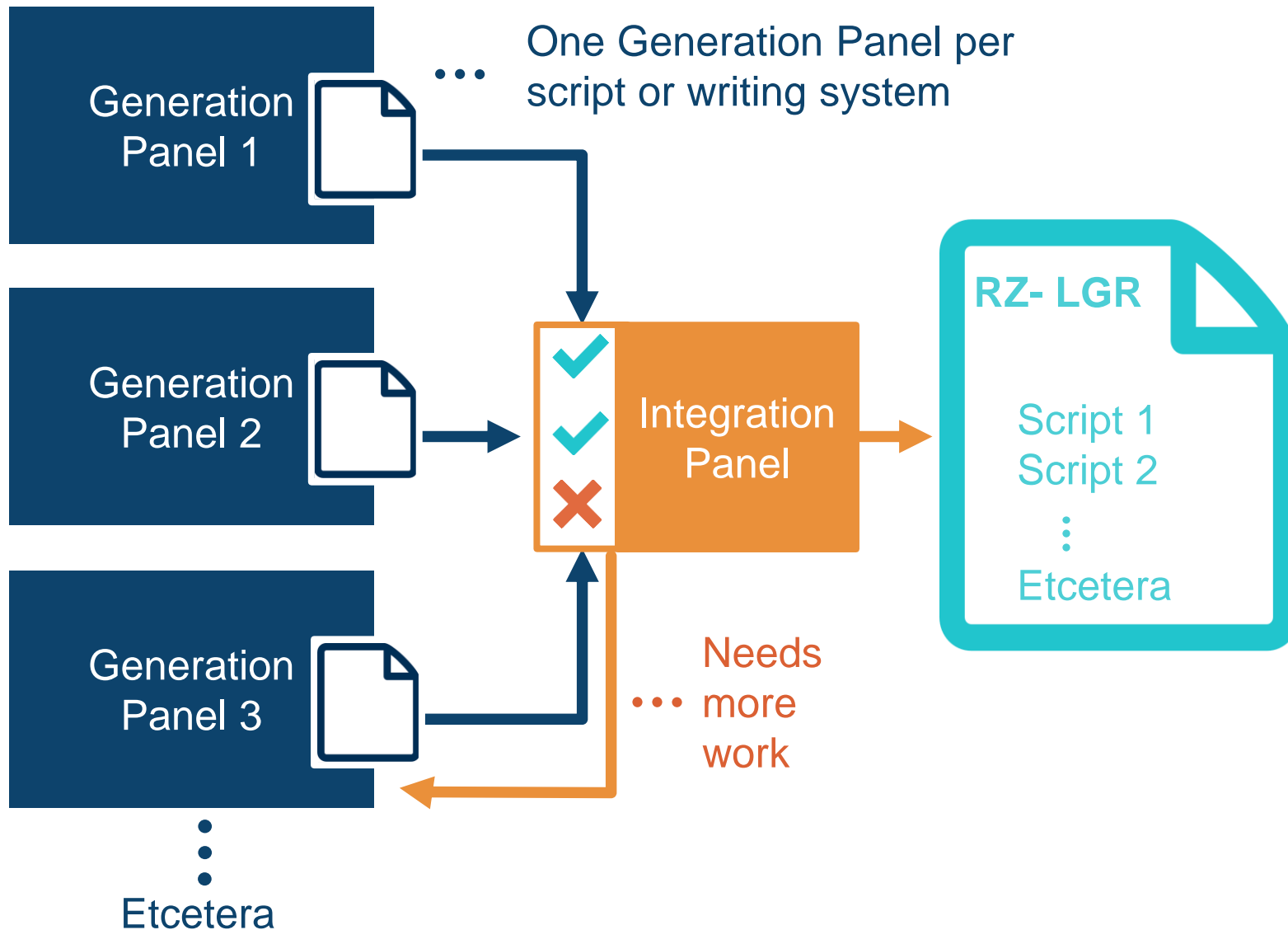
Overview of IDN Programs

- ◉ IDNs at Top Level
 - IDN TLD Program
 - Root Zone Label Generation Rules (RZ-LGR)
 - IDN Variant TLD Implementation
 - LGR Toolset
 - IDN ccTLD Fast Track Process
- ◉ IDNs at Second Level for gTLDs
 - IDN Implementation Guidelines
 - Reference Second Level LGRs
- ◉ Community Outreach and Involvement

Root Zone Label Generation Rules (RZ-LGR)

- ⊙ IDNA2008 expects registries at all levels, including the top-level, will reduce opportunities for confusion by, for example, restricting characters or using variant techniques
 - RZ-LGR basis for such mechanism for the Root Zone
- ⊙ RZ-LGR aims to:
 - Support IDN TLDs in scripts used by communities globally
 - Provide a secure and stable definition for valid IDN TLDs
 - Determine variant labels of IDN TLDs

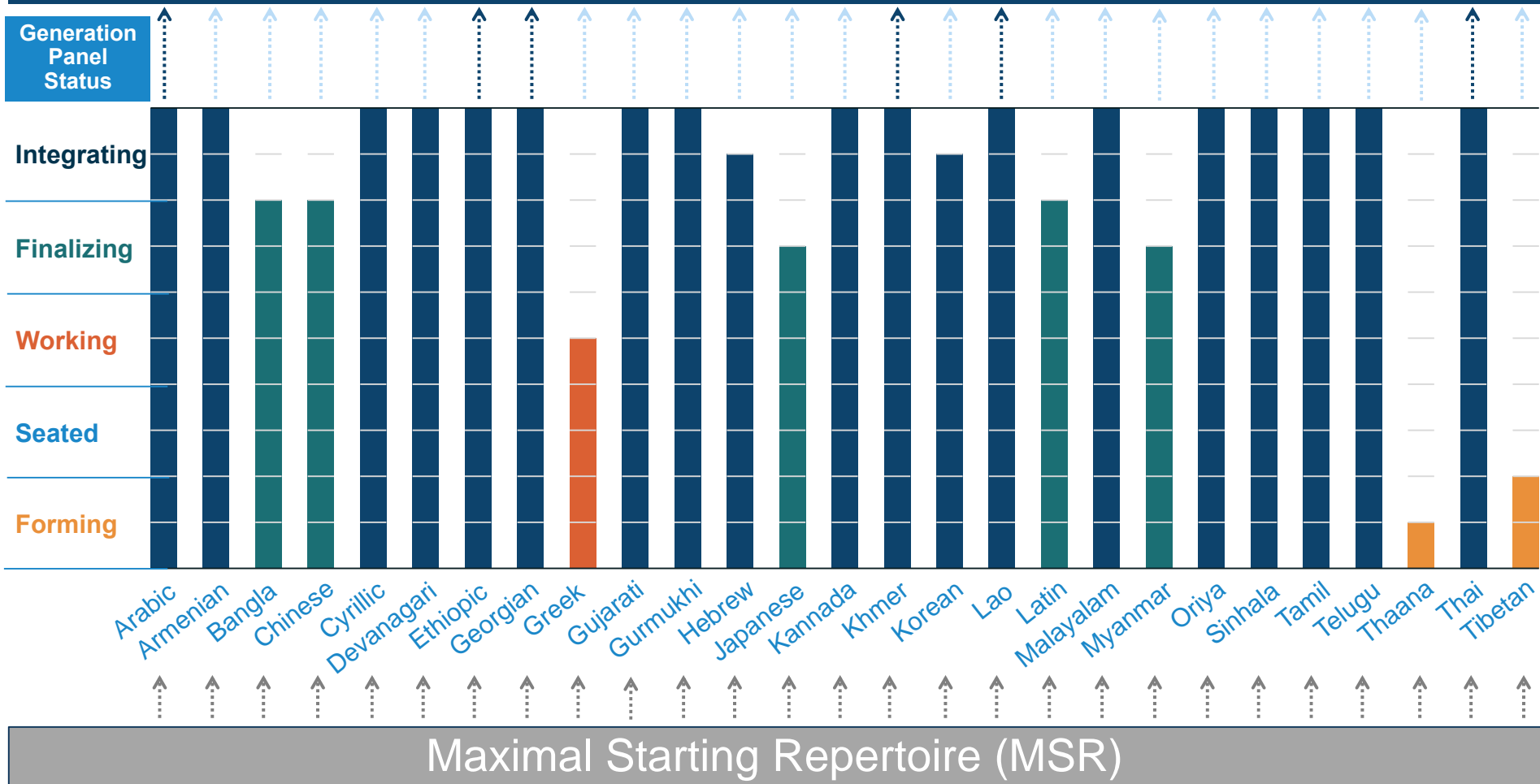
RZ-LGR Process



Status of Generation Panels (GPs)

Root Zone Label Generation Rules (RZ-LGR)

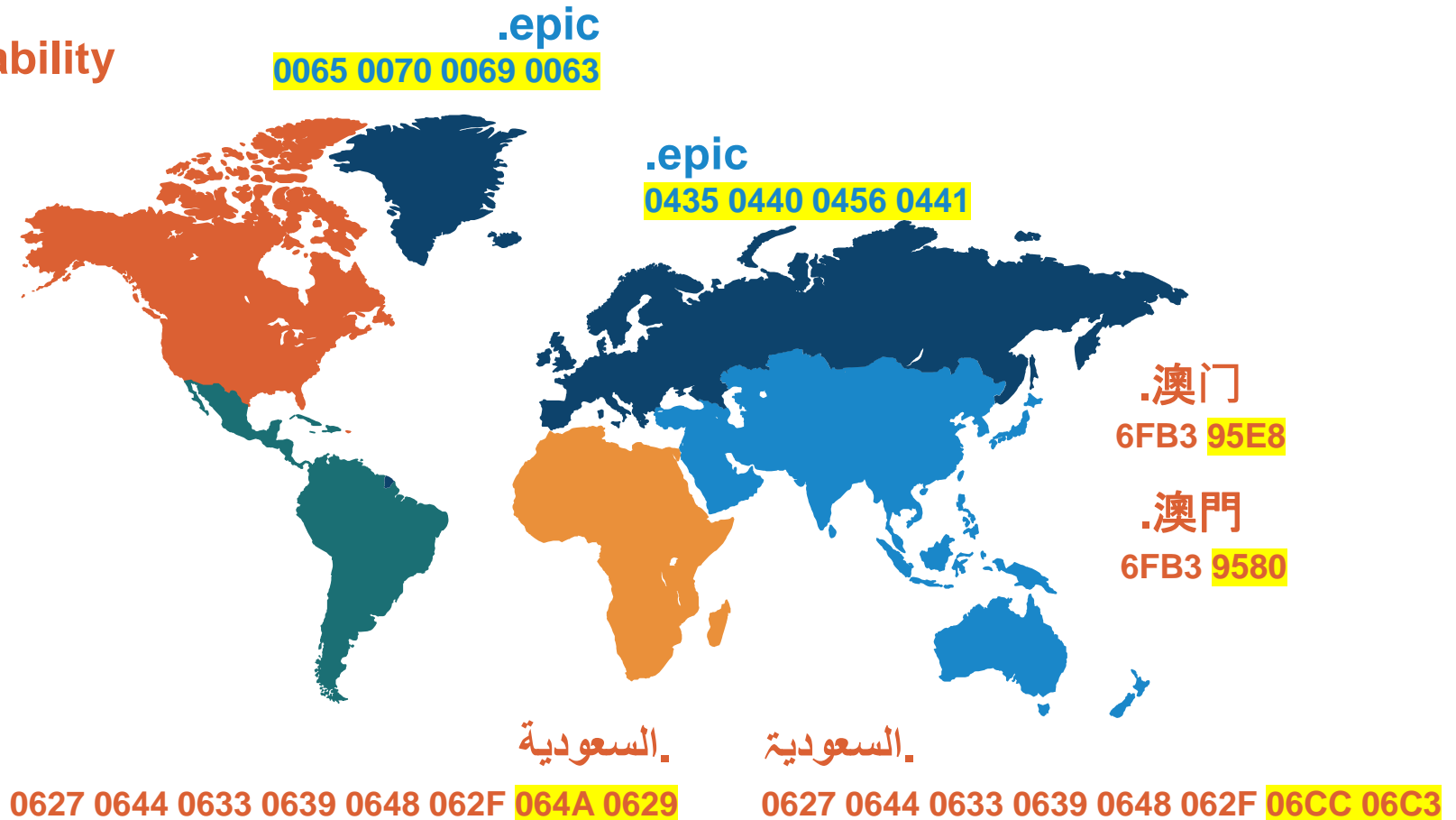
Feb. 2019



Understanding IDN Variant TLDs

⦿ Security

⦿ Usability

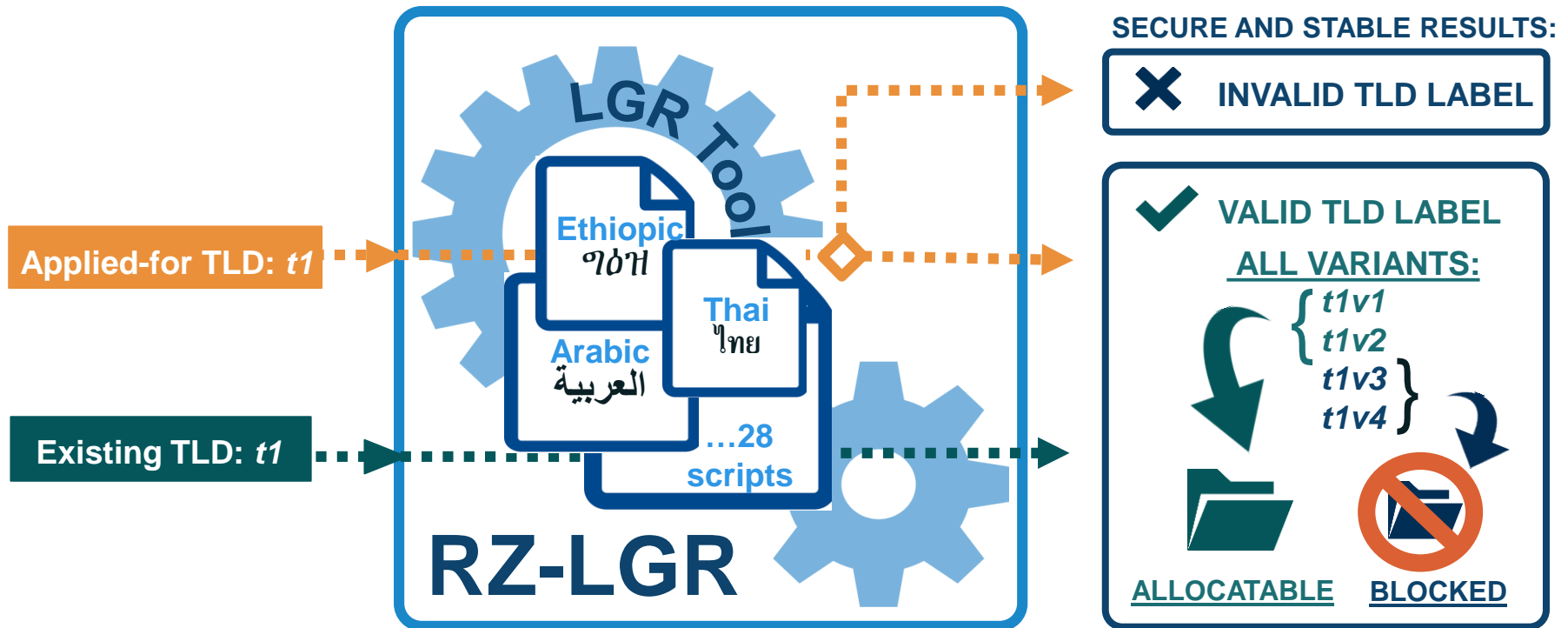


IDN Variant TLD Implementation

- ⦿ Determining variant labels is hard - interpretation of “same” varies across script
- ⦿ On 25 September 2010, the ICANN Board resolved:
 - “No variants of gTLDs will be delegated through the New gTLD Program until appropriate variant management solutions are developed.”
- ⦿ Undertook studies on [Arabic](#), [Chinese](#), [Cyrillic](#), [Devanagari](#), [Greek](#), and [Latin](#) scripts in 2011 to understand the variant phenomenon
- ⦿ Issues collated in the [Integrated Issues Report, IIR](#) (2012) - identified following gaps:
 1. No definition of IDN variant TLDs
 2. No IDN variant TLD management mechanism

Status of IDN Variant TLDs – Definition of Variants

- Gap 1: No definition of IDN variant TLDs
 - Solution: Define variant labels using Root Zone Language Generation Rules (RZ-LGR)
 - Next steps: RZ-LGR-Study Group initiated to review technical implementation



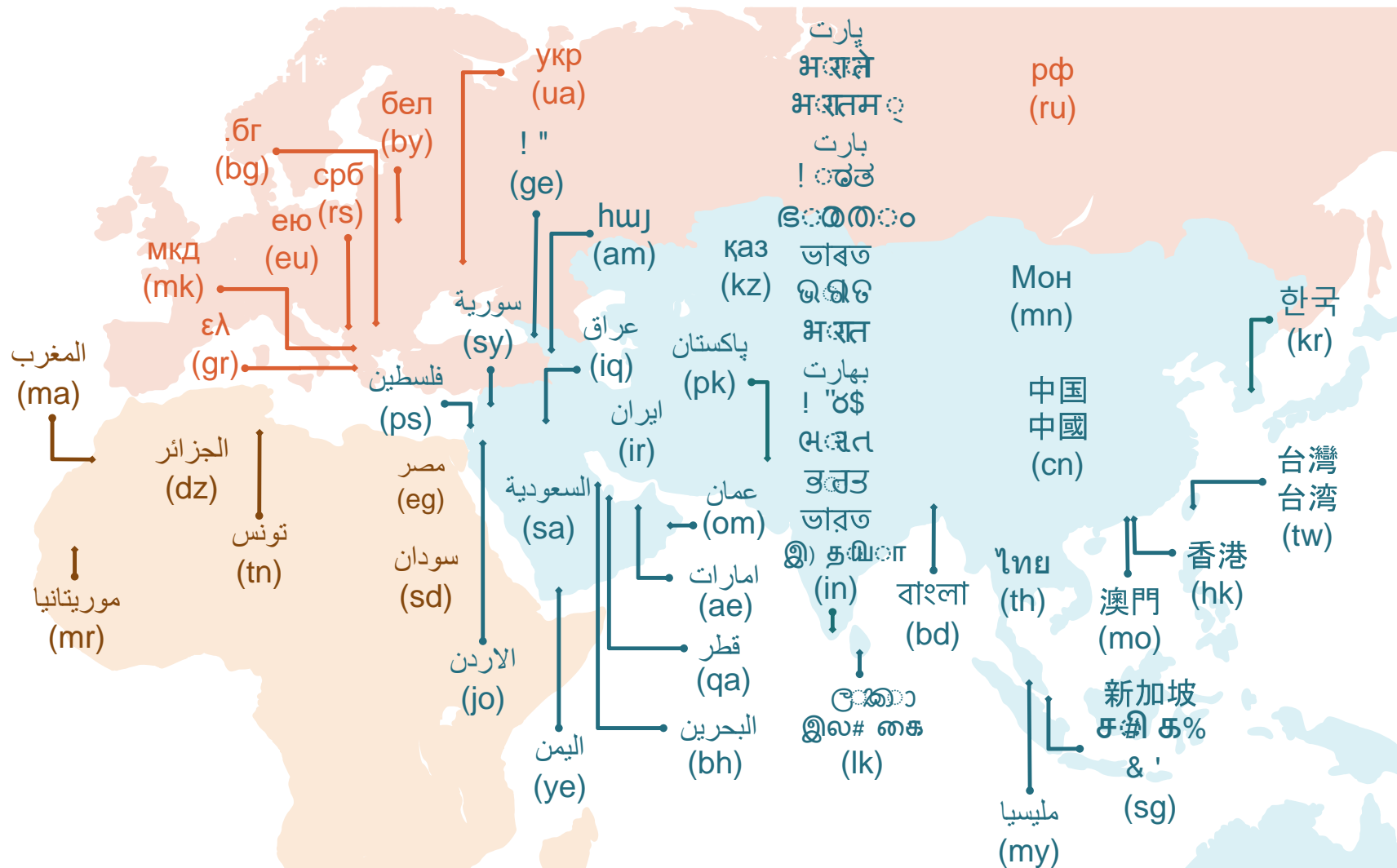
Status of IDN Variant TLDs – Management Mechanism

- ⊙ Gap 2: No IDN variant TLD management mechanism
- ⊙ Solution: ICANN org to work with the community to develop a feasible mechanism
 - Recommendations developed by ICANN org
 - Recommendations presented to ICANN Board on 22 June 2018
 - Recommendations released for [public comment](#) on 25 July 2018
 - [Updated version](#) posted 25 January 2019

LGR Toolset

- ⦿ Label Generation Rulesets (LGRs) used to generate domain name labels, as specified in [RFC 7940](#)
- ⦿ LGR Toolset currently allows for the following:
 - **Create** single LGR or merge multiple LGRs
 - **View** LGR in XML form or user friendly HTML form
 - **Use** an LGR to validate a label and determine its variant labels
 - **Manage** LGRs, by comparing or combining them
 - **Review** impact of a new or a revised LGR on existing labels
- ⦿ Online deployment at: <https://lgrtool.icann.org/>
- ⦿ Open source package(s) released with BSD license at GitHub: [picu](#), [lgr-core](#), [lgr-django](#), [munidata](#)
- ⦿ [User guide](#) available for further details

IDN Country Code Top-Level Domains



* Successfully evaluated IDN ccTLDs for total countries and territories

IDN ccTLD Fast Track Process

- ⦿ Launched in late 2009:
 - 59 IDN ccTLDs evaluated representing 41 countries/territories
 - 57 IDN ccTLDs delegated representing 39 countries/territories
 - Requests cover 33 languages in 19 scripts

- ⦿ Currently under review:
 - [Public comment](#) in January 2015 raised issues with second similarity review process (EPSRP)
 - [Board resolution](#) in June 2015 to review EPSRP
 - ccNSO formed Working Group (WG) on EPSRP
 - [Public comment](#) in July 2016 on updated EPSRP guidelines
 - [Final report](#) published, incorporating feedback and discussion
 - ccNSO adopted the final report by WG on EPSRP
 - [Joint ccNSO SSAC Response](#) to ICANN Board
 - ICANN Board [approved](#) risk mitigation step in string similarity in October 2017
 - Finalizing mitigation step guidelines with feedback from ccNSO

IDN Implementation Guidelines

Background

- For second-level IDN registration policies and practices
- To minimize the risk of cybersquatting and consumer confusion

- gTLD – registries and registrars offering IDNs contractually bound
 - Required by most Registry Agreements
 - For example, new gTLD Registry Agreement: Specification 6 Section 1.4
 - Required by many Registrar Agreements
 - For example, 2013 Registrar Accreditation Agreement: Additional Registrar Operation Specification Clause 3
- IDN ccTLDs – “expected” by the Fast Track Process and the proposed IDN ccTLD Policy

Topics Covered and Next Steps

- Total of 7 topics and 19 guidelines with additional notes:



- IDN Guidelines 4.0 published 10 May 2018, currently undergoing analysis for implementation
- Anticipated for presentation to ICANN Board in May 2019

Communication and Outreach Efforts

- ⦿ Direct outreach
 - Neo-Brahmi (Malayalam) meeting, November 2018, India
 - Webinar on IDN, December 2018, Africa
 - Workshop Jawi Second Level LGR, December 2018, Malaysia
 - THNG8, December 2018, Thailand
 - Neo-Brahmi (Bangla) meeting, December 2018, India
 - Latin GP F2F meeting, February 2019, Belgium
 - Tibetan GP Workshop Meeting, February 2019, Bhutan
- ⦿ Updates at ICANN meetings
- ⦿ IDN web pages at icann.org/idn
- ⦿ [IDN community wiki pages](#)
- ⦿ IDN mailing lists: {lgr, ArmenianGP, ChineseGP, ...}@icann.org

Update by RZ-LGR Integration Panel

Wil Tan
Member, Integration Panel

Agenda

- ⦿ IP activities summary (since ICANN63, October 2018, Barcelona)
 - Reviews
 - MSR
 - Root-Zone LGR

IP Activities Summary: Reviews

- ⊙ GP formation proposals
 - Hebrew

- ⊙ Draft LGRs:
 - Hebrew
 - Myanmar
 - Chinese
 - Japanese
 - Latin
 - Neo-Brahmi:
 - Bengali, Devanagari⁺, Gurmukhi⁺, Gujarati⁺, Kannada⁺, Malayalam⁺, Oriya⁺, Tamil⁺, Telugu⁺
 - Sinhala⁺

+ updated drafts after public comments

IP Activities Summary: Reviews (cont.)

- ⊙ LGR drafts updated after public comments
 - Neo-Brahmi: Devanagari, Gurmukhi, Gujarati, Kannada, Malayalam, Oriya, Tamil, Telugu
 - Sinhala

- ⊙ LGR not yet updated after public comments
 - Korean
 - LGR has not yet been submitted for Integration Panel review
 - Awaiting next steps

IP Activities: MSR-4

- ⦿ MSR-4 was published on 7 February 2019
- ⦿ Additions to existing scripts
 - Latin - 3 code points
 - Myanmar - 12 code points
- ⦿ Future MSR
 - MSR-5 in the horizon
 - Repertoire update to track newer Unicode versions

IP Activities: Root Zone LGR-3

- ⦿ Integration Panel plans to publish RZ-LGR-3
 - Targeting early Q2 2019
 - Planned scripts:
 - Devanagari
 - Gurmukhi
 - Gujarati
 - Kannada
 - Malayalam
 - Oriya
 - Tamil
 - Telugu
 - Sinhala

IP Activities: Variants

- ⦿ IP has engaged in discussions and come up with guidance on some variant issues
- ⦿ Will be presented in the IDN RZ-LGR workshop in the afternoon

Community Updates

- Hebrew GP
- Latin GP
- Neo-Brahmi GP
- Myanmar GP

Update by Hebrew GP

Yoram Hacoen
Member, Hebrew GP

Agenda

1

Overview of the
Script and Language

2

Generation Panel
Membership

3

Summary of the
Progress

4

Current Work

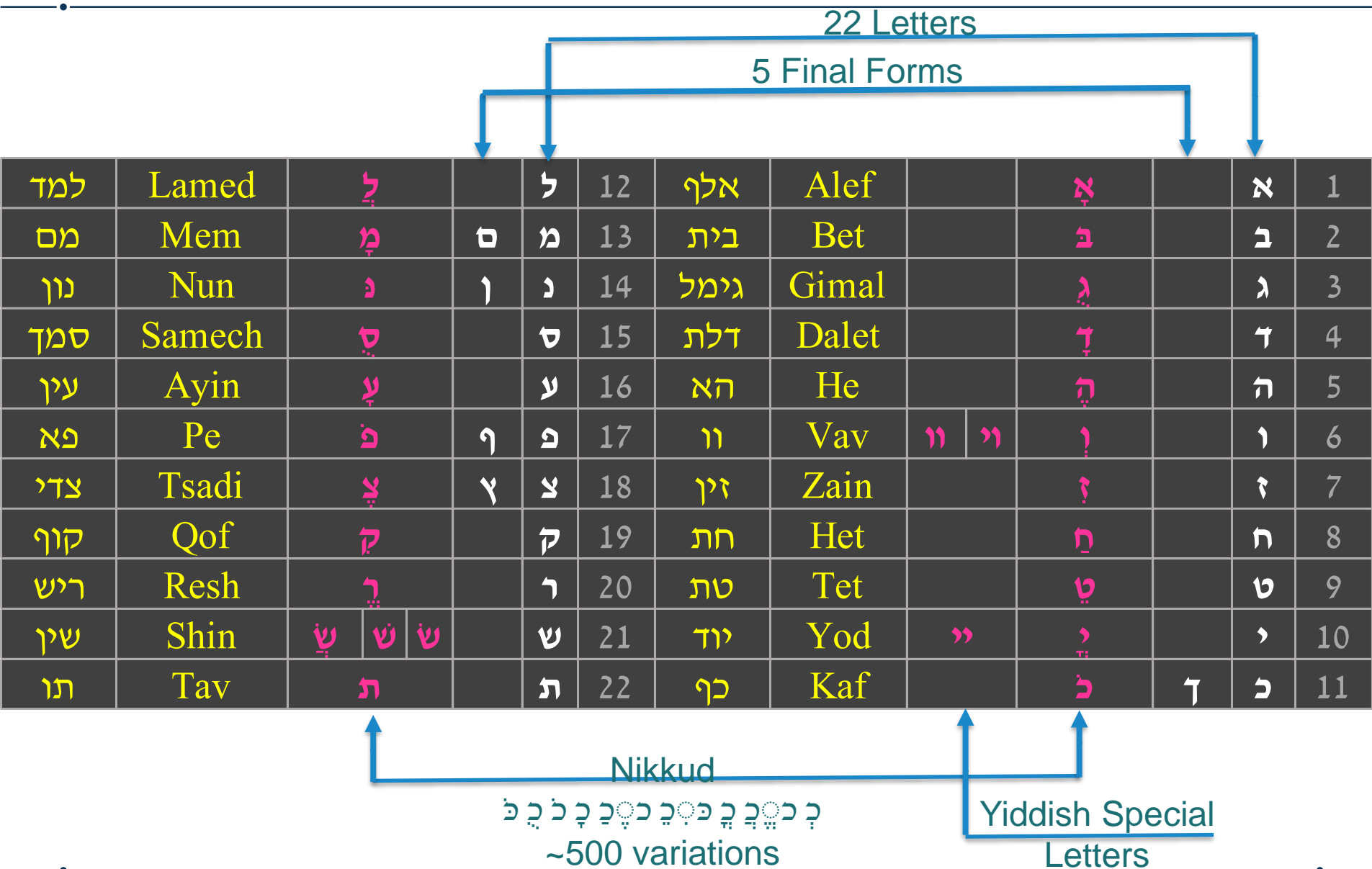
5

Plan and Next steps

Overview of the Hebrew Script and Language

- ISO 15924 Code: Hebr; ISO 15924 Key: 125; ISO 15924 English Name: Hebrew; Latin Transliteration of Native Script Name: Ivrit; Native Script Name: עברית; Minimal Starting Repertoire (MSR) Version: MSR-4
- Hebrew is one of the oldest used scripts (~3000 years), based on Phoenician script and written right-to-left. At the second half of the 1st millennium (6-10 CE), Jewish scribe scholars (“Masoretes”) formalized the script and added pointing (“Nikkud”) and accents (“Te’amim”)
- Used for Modern Hebrew (~8,400,000), Yiddish (~510,000), Ladino (~135,000)
- Modern Hebrew is not written with Te’amim, and rarely uses Nikkud (mainly used in texts for children and in poetry).
- In modern Hebrew, a few letters are also used as vowels, leading to spelling variations (out of the RZ-LGR scope)

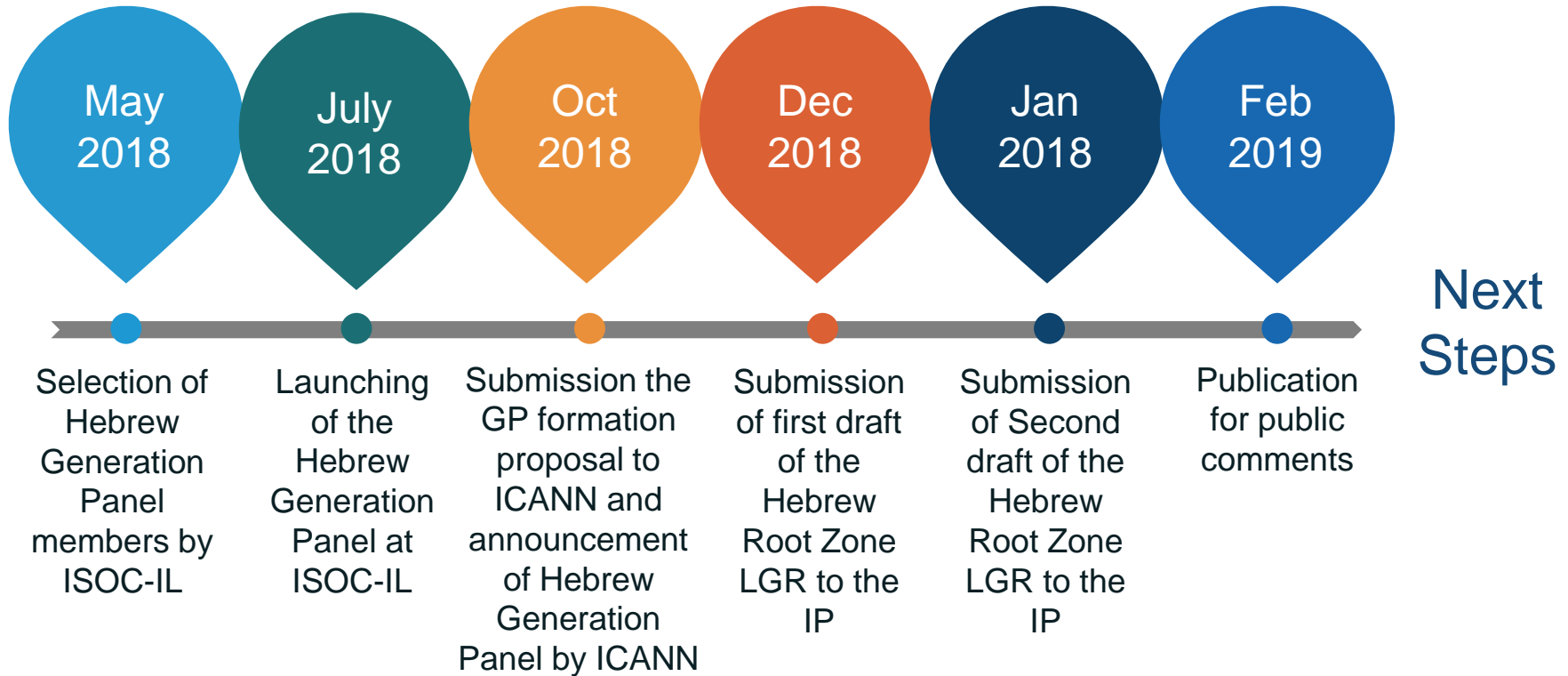
Hebrew Script



Generation Panel Membership

- ⦿ Mr. Doron Shikmoni, Founder, Israel Internet Association (ISOC-IL) and Forescout Technologies – Chair
- ⦿ Ms. Dorit Lerer, Deputy CEO, The Academy for the Hebrew Language - Member
- ⦿ Mr. Matitiah Allouche, Private Expert (Linguistics and Computers) - Member
- ⦿ Mr. Meir Keraushar, DNS Expert, Israel Internet Association (ISOC-IL) - Member
- ⦿ Mr. Yoram Hacoheh, CEO, Israel Internet Association (ISOC-IL) - Member

Summary of the Progress



To Summarize

From Beginning of work to final version it took 5 months and 4 meetings. A KISS (Keep It Simple and Safe) approach was used (easier to later add, harder to remove, code points)

RZ-LGR Repertoire

Hebrew section of Maximal Starting Repertoire [MSR] Version 4

Legend:
 Pink - IDNA2008 PVALID code points excluded from MSR-3
 Yellow - IDNA2008 PVALID code points
 White – Not IDNA2008 PVALID
 Orange line – RZ-LGR repertoire

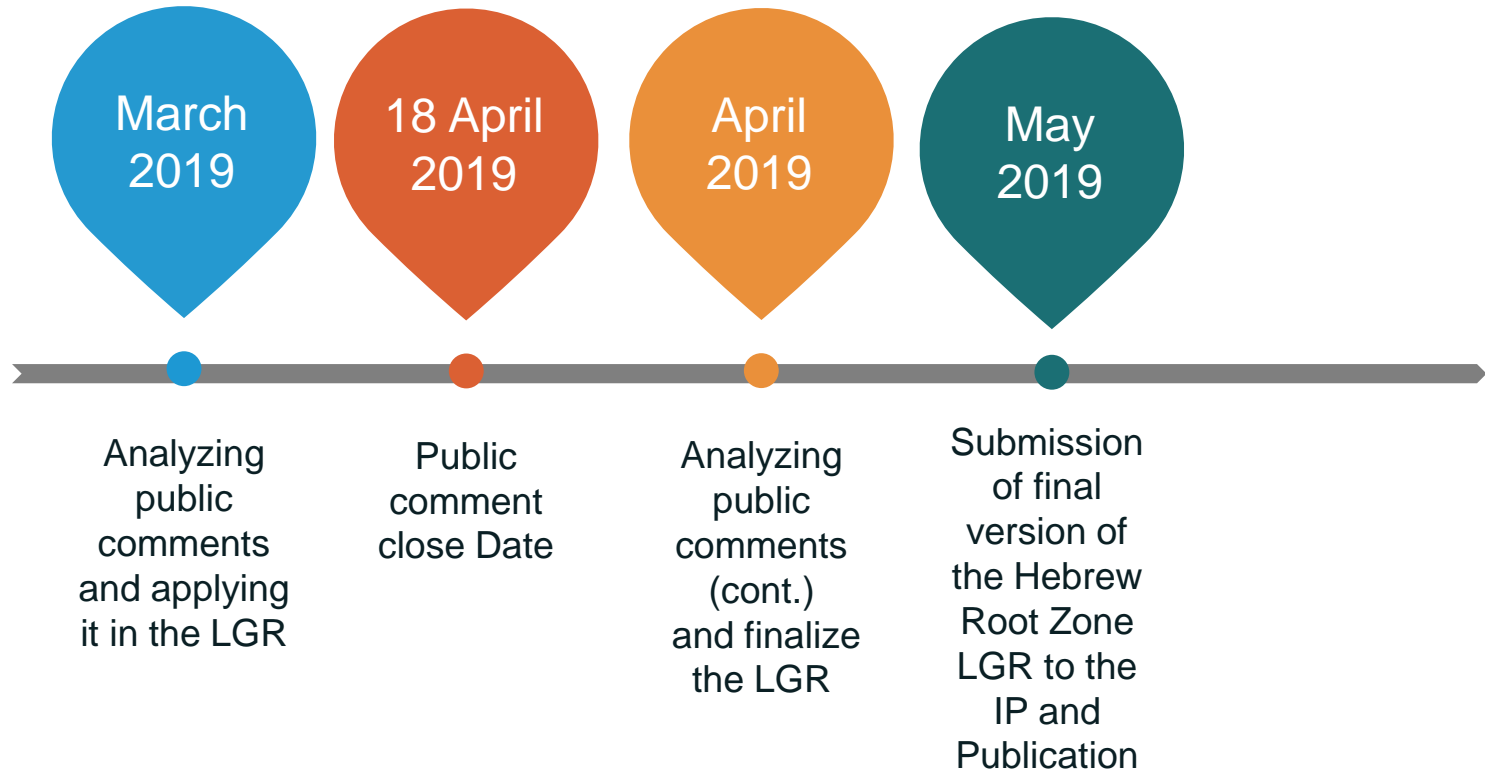
	059	05A	05B	05C	05D	05E	05F
0		◌◌◌ 05A0	◌◌◌ 05B0	◌◌◌ 05C0	◌◌◌ 05D0	◌◌◌ 05E0	◌◌◌ 05F0
1	◌◌◌ 0591	◌◌◌ 05A1	◌◌◌ 05B1	◌◌◌ 05C1	◌◌◌ 05D1	◌◌◌ 05E1	◌◌◌ 05F1
2	◌◌◌ 0592	◌◌◌ 05A2	◌◌◌ 05B2	◌◌◌ 05C2	◌◌◌ 05D2	◌◌◌ 05E2	◌◌◌ 05F2
3	◌◌◌ 0593	◌◌◌ 05A3	◌◌◌ 05B3	◌◌◌ 05C3	◌◌◌ 05D3	◌◌◌ 05E3	◌◌◌ 05F3
4	◌◌◌ 0594	◌◌◌ 05A4	◌◌◌ 05B4	◌◌◌ 05C4	◌◌◌ 05D4	◌◌◌ 05E4	◌◌◌ 05F4
5	◌◌◌ 0595	◌◌◌ 05A5	◌◌◌ 05B5	◌◌◌ 05C5	◌◌◌ 05D5	◌◌◌ 05E5	
6	◌◌◌ 0596	◌◌◌ 05A6	◌◌◌ 05B6	◌◌◌ 05C6	◌◌◌ 05D6	◌◌◌ 05E6	
7	◌◌◌ 0597	◌◌◌ 05A7	◌◌◌ 05B7	◌◌◌ 05C7	◌◌◌ 05D7	◌◌◌ 05E7	
8	◌◌◌ 0598	◌◌◌ 05A8	◌◌◌ 05B8		◌◌◌ 05D8	◌◌◌ 05E8	
9	◌◌◌ 0599	◌◌◌ 05A9	◌◌◌ 05B9		◌◌◌ 05D9	◌◌◌ 05E9	
A	◌◌◌ 059A	◌◌◌ 05AA	◌◌◌ 05BA		◌◌◌ 05DA		
B	◌◌◌ 059B	◌◌◌ 05AB	◌◌◌ 05BB		◌◌◌ 05DB		
C	◌◌◌ 059C	◌◌◌ 05AC	◌◌◌ 05BC		◌◌◌ 05DC		
D	◌◌◌ 059D	◌◌◌ 05AD	◌◌◌ 05BD		◌◌◌ 05DD		
E	◌◌◌ 059E	◌◌◌ 05AE	◌◌◌ 05BE		◌◌◌ 05DE		
F	◌◌◌ 059F	◌◌◌ 05AF	◌◌◌ 05BF		◌◌◌ 05DF		

	FB0	FB1	FB2	FB3	FB4
0	◌◌◌ FB00		◌◌◌ FB20	◌◌◌ FB30	◌◌◌ FB40
1	◌◌◌ FB01		◌◌◌ FB21	◌◌◌ FB31	◌◌◌ FB41
2	◌◌◌ FB02		◌◌◌ FB22	◌◌◌ FB32	
3	◌◌◌ FB03	◌◌◌ FB13	◌◌◌ FB23	◌◌◌ FB33	◌◌◌ FB43
4	◌◌◌ FB04	◌◌◌ FB14	◌◌◌ FB24	◌◌◌ FB34	◌◌◌ FB44
5	◌◌◌ FB05	◌◌◌ FB15	◌◌◌ FB25	◌◌◌ FB35	
6	◌◌◌ FB06	◌◌◌ FB16	◌◌◌ FB26	◌◌◌ FB36	◌◌◌ FB46
7	◌◌◌ FB07	◌◌◌ FB17	◌◌◌ FB27		◌◌◌ FB47
8		◌◌◌ FB18	◌◌◌ FB28	◌◌◌ FB38	◌◌◌ FB48
9		◌◌◌ FB19	◌◌◌ FB29	◌◌◌ FB39	◌◌◌ FB49
A		◌◌◌ FB1A	◌◌◌ FB2A	◌◌◌ FB3A	◌◌◌ FB4A
B		◌◌◌ FB1B	◌◌◌ FB2B	◌◌◌ FB3B	◌◌◌ FB4B
C		◌◌◌ FB1C	◌◌◌ FB2C	◌◌◌ FB3C	◌◌◌ FB4C
D		◌◌◌ FB1D	◌◌◌ FB2D		◌◌◌ FB4D
E		◌◌◌ FB1E	◌◌◌ FB2E	◌◌◌ FB3E	◌◌◌ FB4E
F		◌◌◌ FB1F	◌◌◌ FB2F		◌◌◌ FB4F

root-zone-label-generation-rule-set-for-the-hebrew-script.xml

```
<data>
<range ref="101" last-cp="05D9" first-cp="05D0"/>
<char comment="Final form" ref="101" cp="05DA">
  <var type="blocked" comment="Final form variant" cp="05DB"/>
</char>
<char comment="Nominal form" ref="101" cp="05DB">
  <var type="blocked" comment="Nominal form variant" cp="05DA"/>
</char>
<char ref="101" cp="05DC"/>
<char comment="Final form" ref="101" cp="05DD">
  <var type="blocked" comment="Final form variant" cp="05DE"/>
</char>
<char comment="Nominal form" ref="101" cp="05DE">
  <var type="blocked" comment="Nominal form variant" cp="05DD"/>
</char>
<char comment="Final form" ref="101" cp="05DF">
  <var type="blocked" comment="Final form variant" cp="05E0"/>
</char>
<char comment="Nominal form" ref="101" cp="05E0">
  <var type="blocked" comment="Nominal form variant" cp="05DF"/>
</char>
<range ref="101" last-cp="05E2" first-cp="05E1"/>
<char comment="Final form" ref="101" cp="05E3">
  <var type="blocked" comment="Final form variant" cp="05E4"/>
</char>
<char comment="Nominal form" ref="101" cp="05E4">
  <var type="blocked" comment="Nominal form variant" cp="05E3"/>
</char>
<char comment="Final form" ref="101" cp="05E5">
  <var type="blocked" comment="Final form variant" cp="05E6"/>
</char>
<char comment="Nominal form" ref="101" cp="05E6">
  <var type="blocked" comment="Nominal form variant" cp="05E5"/>
</char>
<range ref="101" last-cp="05EA" first-cp="05E7"/>
</data>
```

Next Steps



Update by Latin GP

Michael Bauland and William Jouris
Members, Latin GP

Agenda Overview

1

Short History

2

Scope of Work

3

Members

4

Organization of
Working Group

5

Work Accomplished

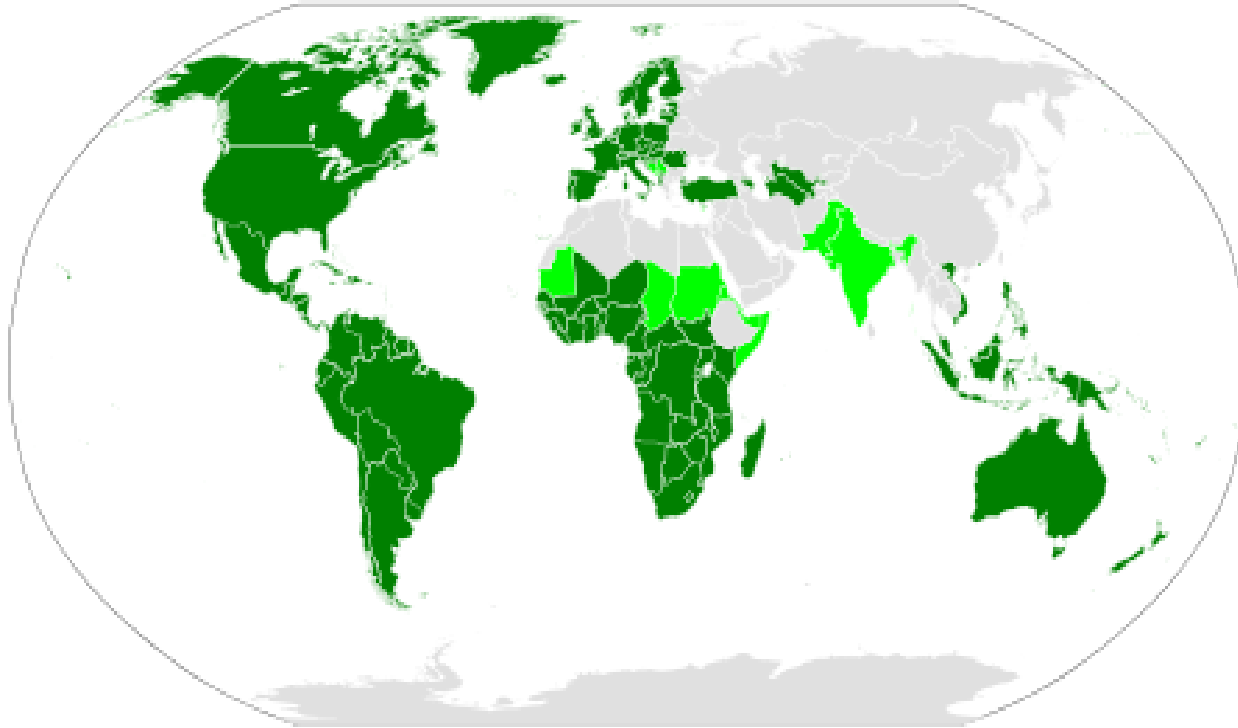
6

Project Timeline

Latin GP – Short History

- ⦿ June-August 2016 - GP restarted with new call for volunteers. The GP seated 15 May 2017 Proposal for Formation of Latin Generation Panel
- ⦿ September 2017 - GP proposal for inclusion and exclusion principles were sent for an informal public review in
- ⦿ During September-November 2017 - GP has collected information from 209 languages
- ⦿ GP proposed new code points for MSR-3 and MSR-4
- ⦿ May 2018 (for MSR-3) and October 2018 (for MSR-4) - GP submitted the code point repertoire to the Integration Panel
- ⦿ September 2018 & January 2019 - GP submitted the updated LGR proposal with the cross-script variant analysis and the initial in-script variant analysis
- ⦿ GP is currently reviewing IP feedback and finalizing the LGR proposal

Latin Script Geographic and Linguistic Spread



Dark green = Latin script is the sole main script.

Light green = Latin co-exists with other scripts.

Grey = Latin-script alphabets are sometimes extensively used due to the use of unofficial second languages, such as French in Algeria and English in Egypt, and to Latin transliteration of the official script, such as in China or in Japan.

Latin GP – Scope of Work for Code Point Analysis

- ⊙ Maximal String Repertoire Version 4 (MSR-4)
 - Subset of code points allowed in IDNA 2008
- ⊙ Unicode ranges
 - Controls and Basic Latin
 - Controls and Latin-1 Supplement
 - Latin Extended-A only lowercase
 - Latin Extended-B
 - IPA Extensions
 - Combining Diacritical Marks
 - Combining Diacritical Marks Supplement
 - Latin Extended Additional
 - Latin Extended-C
- ⊙ Non exhaustive list of 455 languages in scope
- ⊙ Non exhaustive list of EGIDS 1-5 languages contains 300 languages
- ⊙ Non exhaustive list of EGIDS 1-4 languages contains 181 languages

Latin GP – Scope of Work Variant Analysis

- ⊙ In-script variant analysis
- ⊙ Cross-script variant analysis
 - Armenian script
 - Cyrillic script
 - Greek script
 - Basic shapes (e.g., circle “o”, single line “l”, and crescent “c” or “o”) within all scripts

Latin GP – Members

- ⊙ 14 member, 3 observers
- ⊙ Language representatives
 - Africa
 - Asia
 - Australia and Oceania
 - Europe
 - North America
- ⊙ Diversity
 - Community Representatives
 - Linguistic Experts
 - Registry/Registrar Experts
 - Technical Community, DNS Experts
 - IDNA/Unicode Experts

Latin GP – Challenges and Solutions

⦿ Challenges

- Many languages
- Many code points to process
- Change of requirements

⦿ Solutions

- Process languages with EGIDS=1-4 first (180)
- Consider processing languages with EGIDS=5 (119)
 - 29 languages with at least 1 million users with sufficient reference are included
- Define simple procedure for developing Latin script repertoire
- Workload divided in two groups
 - Repertoire Working Group
 - Variant Working Group
- Extend planned working time (finish 2019 instead of 2018)

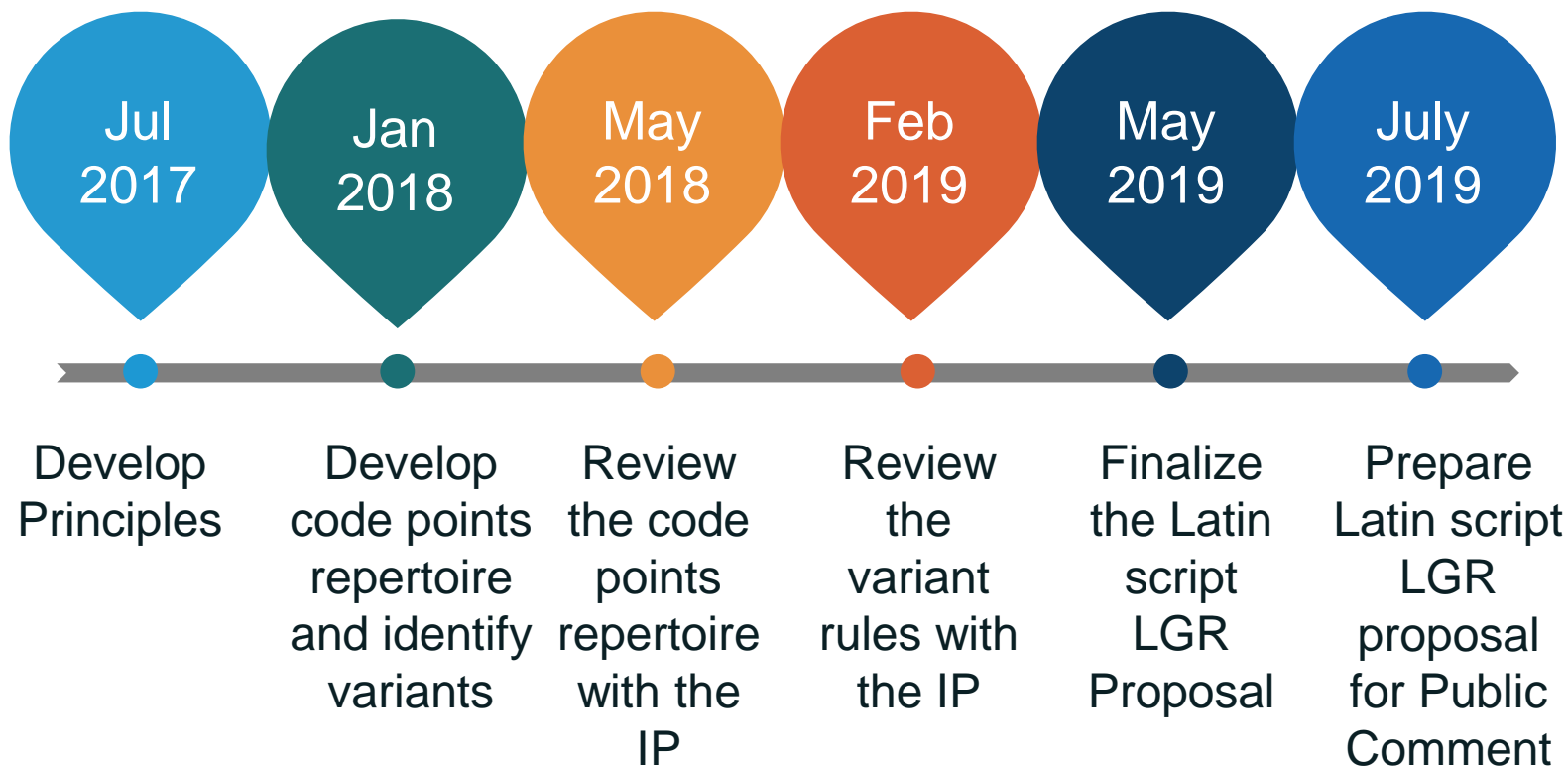
Latin GP – Organization of Working Groups

- ⊙ Repertoire Working Group
 - 10 members
 - Developing Principles for Inclusion and Exclusion of Code Points in Latin Script for the Root Zone LGR
 - Processing Languages to build the repertoire
- ⊙ Variant Working Group
 - 7 members
 - Developing Principles for Analysis of Variants in the Latin Script for the Root Zone LGR
 - Identifying variants with all Latin GP members

Latin GP – Work Accomplished

- ⦿ Developing Repertoire
 - 181 of 181 EGIDS 1- 4 languages processed
 - 29 EGIDS 5 languages processed
 - 195 of 279 MSR-2 code points attested
 - 3 non-MSR-2 code points are included in MSR-3
 - 3 non-MSR-3 code points are included in MSR-4
- ⦿ Developing Variants
 - In-script variants defined
 - Cross-script variants with Armenian script defined
 - Cross-script variants with Cyrillic script defined
 - Cross-script variants with Greek script defined
 - Considering special HTML Link scenario (underlining)
- ⦿ Submitted the third round proposal to the IP in January 2019

Latin GP – Project Timeline



Update on Neo-Brahmi GP

Udaya Narayana Singh
Co-chair, NBGP

Agenda

1

Introducing NBGP

2

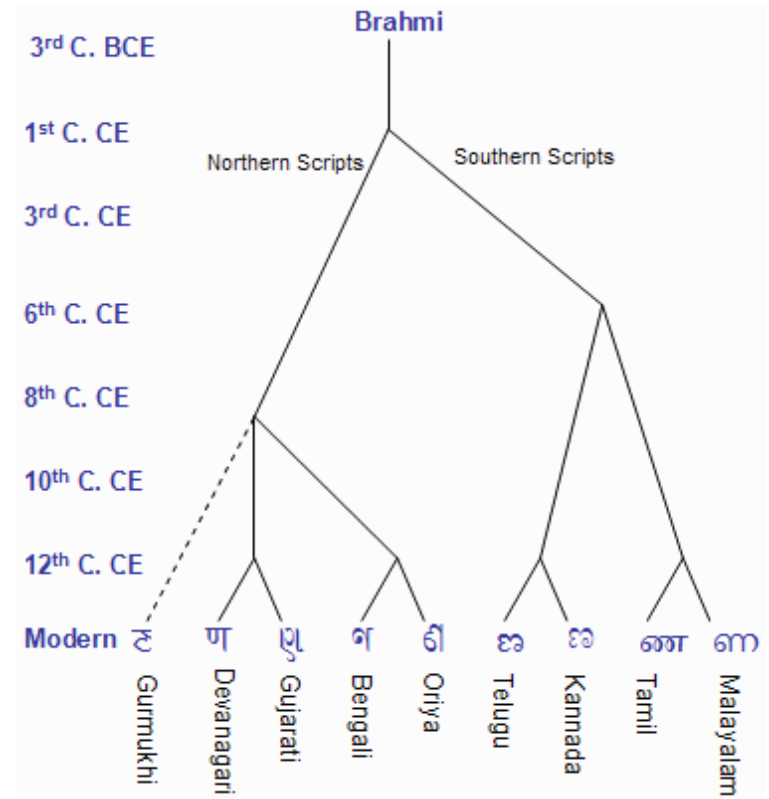
NBGP Members

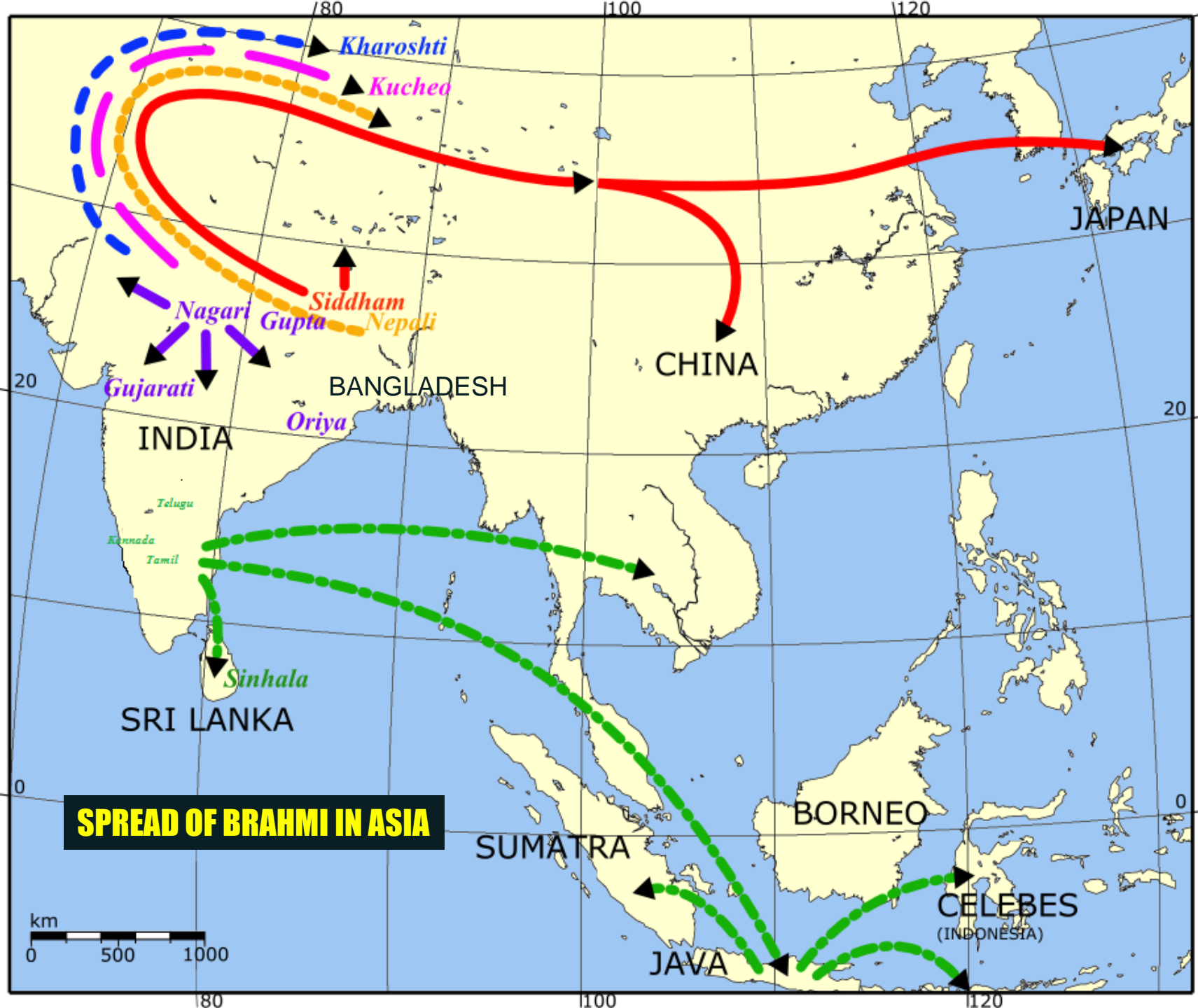
3

Efforts and
Progress Timeline

4

Future Plan of
Action





SPREAD OF BRAHMI IN ASIA

Introducing NBGP

1

Introduction

Generate LGR-proposals for Neo-Brahmi scripts spread all over South Asia, drawing expertise from linguists, printing and publishing specialists, authors & users – looking into their requirements. Also, ensure Global Acceptability of Neo-Brahmi Script based language IDN'S and variants.

2

Scope

Nine writing systems, each used by several languages - Bangla, Devanagari, Gujarati, Gurmukhi/Punjabi, Kannada, Malayalam, Odia/Oriya, Tamil, and Telugu. Among them, Devanagari alone is used by over eleven major & 100 other languages, and Bangla by three major languages.

3

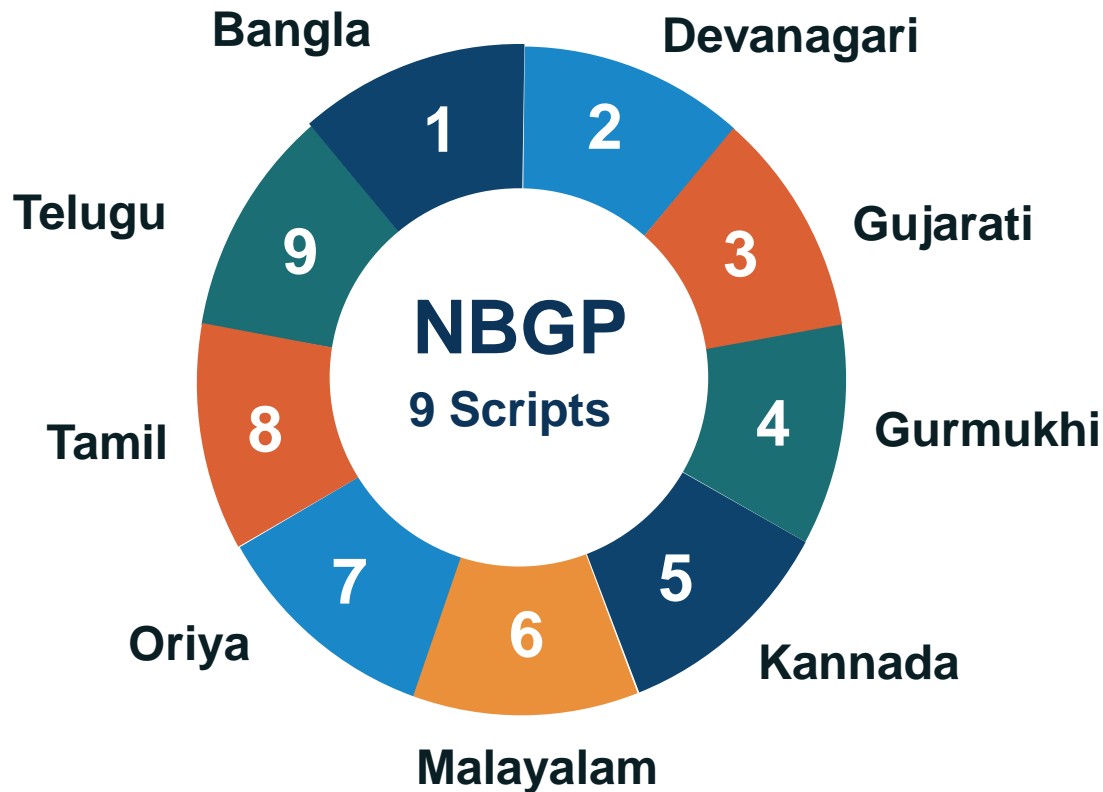
Geo Coverage

Bangladesh, India, Nepal, Singapore, and Sri Lanka as well the South Asian Diaspora spread over 70 countries

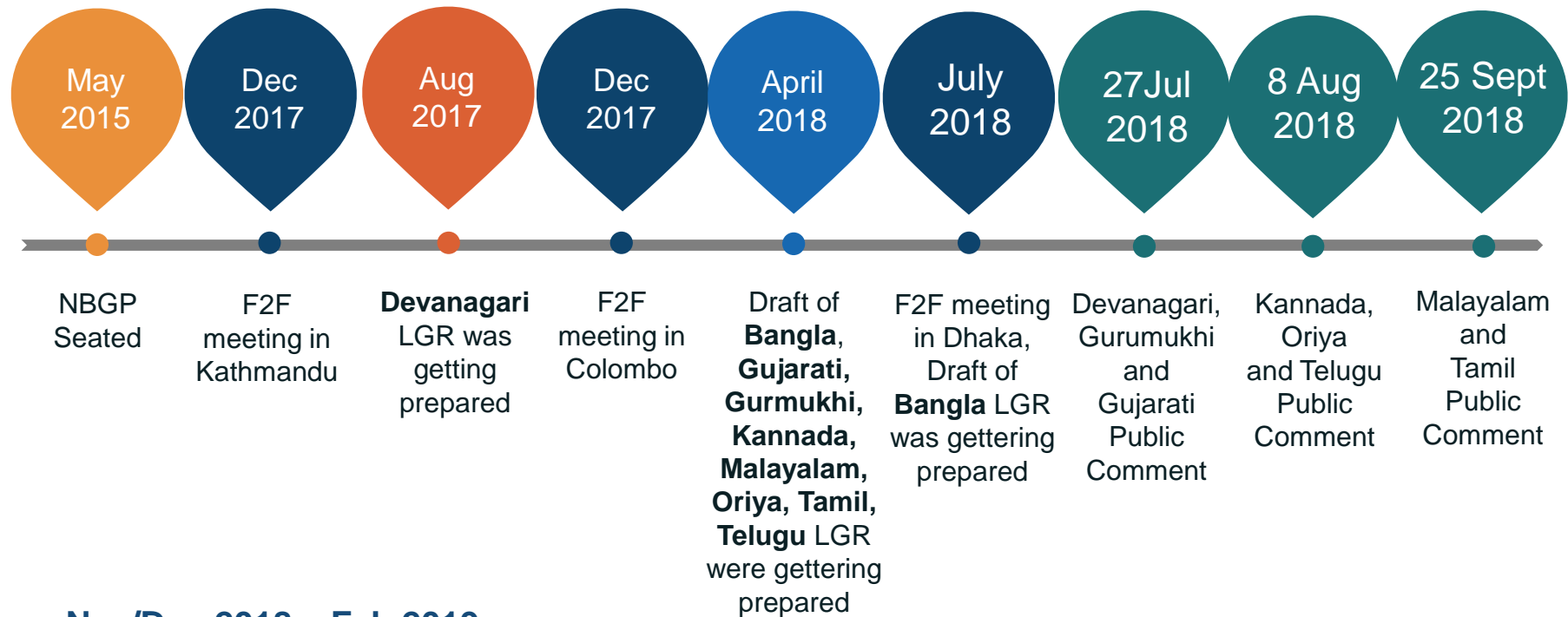
Members

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

Members: 68 members from Bangladesh, India, Nepal, Singapore, and Sri Lanka



Timeline



Nov/Dec 2018 – Feb 2019

Finalize the proposals based on feedback from public comment.

Complications



Malayalam



Tamil



Gujara



Devanagari



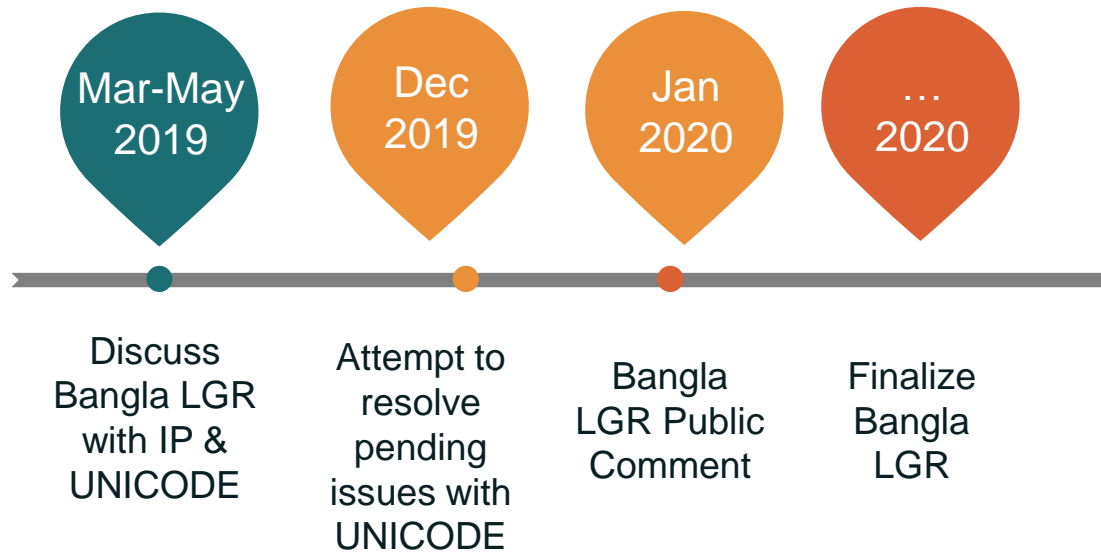
Oriya

- Each writing system in South Asia has numerous difficulties because of their graphics – protruding below, above, and on the aisles as well as through hanging letters. Our LGR-proposals for Neo-Brahmi must capture their characteristics.

Summary of Achievements

- ⦿ 8 Scripts – Completed
- ⦿ 1 Script – Close to completion
- ⦿ 18 months – From active work to finish
- ⦿ 65 Experts and users as volunteers

Future Plan of Action



Pending discussion

The mechanism to use the three Bangla characters in the Root Zone.

- ঞ YYA (U+09DF)
- ঠ RRA (U+09DC)
- ড RRHA (U+09DD).

Update on Myanmar GP

Yin May Oo
Co-Chair, Myanmar GP

Agenda

1

Overview of the
Script and Language

2

Generation Panel
Membership

3

Summary of the
Progress

4

Current Work

5

Plan and Next steps

Myanmar GP – Languages Using Myanmar Script

- Myanmar Script is mostly composed of full-circles, written from left to right, spelled phonetically including the tones with diacritics which could be added around the center character.
- Ethnic languages may use different characters or different diacritics
- 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

Generation Panel Membership

Advisor	Dr. Myint Myint Than	
	Dr. Khin Aye	Mr. Zaw Htut
	Mr. Ngwe Tun	Mr. Maung Sun
Chair	Ms. Thin Zar Phyo	
Co-Chair	Ms. Yin May Oo	
Member	Mr. Naing Win Oo	Mr. Sai Zin Di Di Zone
	Mr. Ye Zarni Aung	Ms. Nancy Aye
	Mr. San Lin Naing	Mr. Thura Soe
	Mr. Min Paing Khant Oo	Mr. Nai Saik Chan
	Mr. Kaung Khant Zaw	Mr. Khun Maung Maung

In-Script Variant Analysis [1/3]

- Myanmar GP defines the following are in-script variant code points due to the nearly identical glyph or the character's property of languages

Set#	Unicode Code Point	Glyph	Unicode Code Point	Glyph
1	U+1023	ဒ	U+1000 U+1039 U+1000	ꠊ
2	U+1029	ဩ	U+101E U+103C	ꠏ
3	U+102A	ဪ	U+1029 U+1031 U+102C U+103A	ꠐ
4	U+102A	ဪ	U+101E U+103C U+1031 U+102C U+103A	ꠐ
5	U+1061	၇	U+101B U+103E	ꠗ
6	U+107E	ဃ	U+107D U+103E	ꠛ
7*	U+1004	င	U+105A	ꠜ

Burmese င : Mon င
 / U+1004 / U+103A / : / U+105A / U+103A /

In-Script Variant Analysis [2/3]

- Myanmar GP defines the following are in-script variant code points due to the nearly identical glyph or the character's property of languages

Set#	Unicode Code Point	Glyph	Unicode Code Point	Glyph
8	U+1008	၈	U+105B	၈
9	U+1027	၉	U+1028	၉
10	U+1000	၀	U+1075	၀
11	U+1001	၁	U+1076	၁
12	U+1002	၂	U+1077	၂
13	U+1005	၅	U+1078	၅
14	U+1007	၇	U+1079	၇
15	U+100A	၁၀	U+107A	၁၀

In-Script Variant Analysis [3/3]

- Myanmar GP defines the following are in-script variant code points due to the nearly identical glyph or the character's property of languages

Set#	Unicode Code Point	Glyph	Unicode Code Point	Glyph
16	U+100F	ဏ	U+107C	ဧ
17	U+1014	န	U+107C	ဧ
18	U+1016	ဖ	U+107D	ဗ
19	U+1017	ဗ	U+107F	ဗ
20	U+1021	အ	U+1022	က
21	U+102E	ဝံ	U+1033	ဝံ
22	U+102B	ဝါ	U+1083	ဝါ
23	U+102C	ဝ	U+1083	ဝါ

Cross-Script Variant Analysis [1/2]

● 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	ଠ	U+0B20	ORIYA LETTER TTHA
2	၆၀	U+1031	MYANMAR VOWEL SIGN E	ଠୌ	U+0B47	ORIYA VOWEL SIGN E

Cross-Script Variant Analysis [2/2]

● Myanmar-Georgian

No.	Glyph	Code Point	Myanmar Character Name	Glyph	Code Point	Malayalam Character Name
1	ဂ	U+1002	MYANMAR LETTER GA	ინ	U+10D8	GEORGIAN LETTER IN
2	တ	U+1010	MYANMAR LETTER TA	თ	U+10D7	GEORGIAN LETTER TAN

- Georgian word “თთთთ” can be formed using Myanmar Consonants “တ” and “ဂ”

Confusable Code Point Analysis

- In-Script confusable code points

No.	Glyph	Code Point	Glyph	Code Point	Note
1	U+1008	ချ	U+1005 U+103B	ချ	The sequence U+1005 U+103B is invalid
2	U+1009 U+102C	ည	U+1025 U+102C	ည	The sequence U+1025 U+102C is invalid
3	U+105B	ရွှ	U+1007 U+103B U+103E	ရွှ	The sequence U+1007 U+103B U+103E is invalid
4	U+1070	ဃ	U+1003 U+103E	ဃ	The sequence U+1003 U+103E is invalid
5	U+1009 U+103A	ဠ	U+1025 U+103A	ဠ	The sequence U+1025 U+103A is invalid
6	U+1009 U+1037 U+103A	ဠ	U+1025 U+1037 U+103A	ဠ	The sequence U+1025 U+1037 U+103A is invalid

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 ऌ̣
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 [1/2]

1. [DVS or ANUSAVARA] must follow C or M
2. M must follow C
// M cannot be at the beginning of the string
3. (C+ K) or (C +1037+K) must follow C or [C+M] or [C+M+V]
4. C1 must follow C or [C+M] or [C+M+V] or 1023 or 1025
//C1 must not be at the beginning of the string
5. S16 can only follow these consonants (c , േ , ൈ , ൉ , ഴ , ഺ , ഻)
or S14 or S15
// this whole pattern cannot be at the beginning of the string
6. S17 must follow these consonants (c , േ , ൈ , ൉ , ഴ , ഺ , ഻)
// this whole pattern cannot be at the beginning of the string
7. S11 must follow C or M or V , and another C must follow S11
// S11 cannot be at the beginning of the string
8. Consonant + Virama must be only followed by Consonant
// this whole pattern cannot be at the beginning of the string
// Consonant+Virama pattern cannot be repeated
9. T_LONG must follow [LV or S12 or S13 or K]
10. T_SHORT must follow [LV or S12 or S13 or S14 or S15]

Whole Label Evaluation Rules [2/2]

11. U+103B must follow CS1
12. U+103C must follow CS2
13. U+103D must follow CS3
14. U+103E must follow CS4
15. S5 must follow CS5
16. S6 must follow CS6
17. S7 must follow CS7
18. S8 must follow CS8
19. S9 must follow CS
20. S10 must follow CS10

Engage with ICANN and IDN Program



Thank You and Questions

Visit us at icann.org/idn

Email: IDNProgram@icann.org



[@icann](https://twitter.com/icann)



facebook.com/icannorg



youtube.com/icannnews



flickr.com/icann



linkedin/company/icann



slideshare/icannpresentations



soundcloud/icann