I C A N N | **63**
ANNUAL GENERAL

BARCELONA
20–25 October 2018

IDN Program Update

IDN Program, ICANN

ICANN63
24 October 2018

Overview of Session Presentations

- ⊙ IDN Program Overview and Progress   - Sarmad Hussain

- ⊙ Update by Integration Panel   - Marc Blanchet

- ⊙ Community Updates
  - ○ Latin GP Update   - Mirjana Tasić
  - ○ Neo-Brahmi GP Update   - Ajay Data
  - ○ Sinhala GP Update   - Harsha Wijayawardhana
  - ○ Myanmar GP Update   - Thin Zar Phyo

- ⊙ Q/A

IDN Program Overview and Progress

Sarmad Hussain
Director, IDN Programs

IDN Program Objectives

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

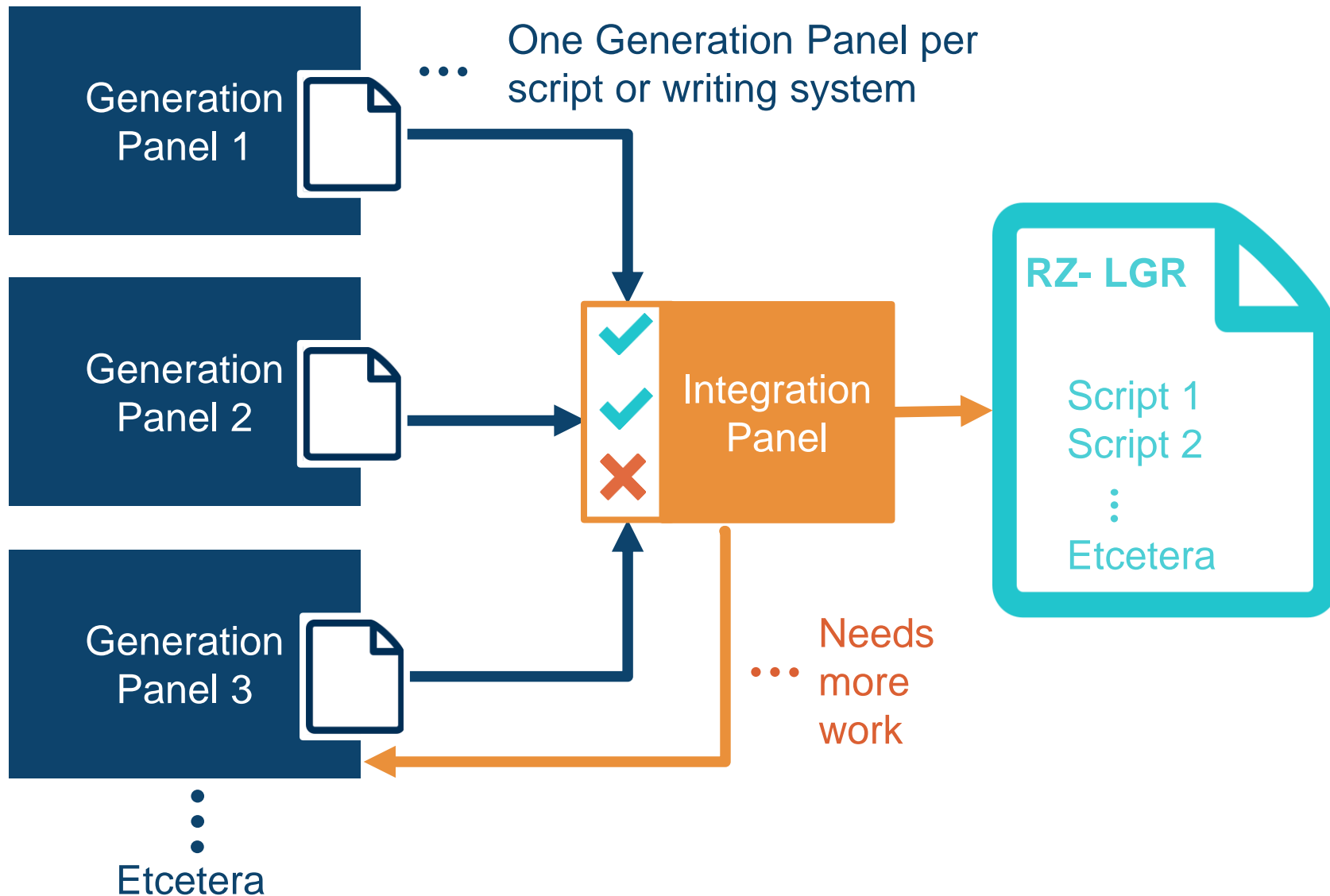| LGR Specification and Tool (P1) |
| LGR Development (P2.2) |
| IDN Variant TLD Implementation (P7) |
| IDN ccTLD Fast Track |
| Reference Second Level LGRs |
| IDN Implementation Guidelines |
| Communications Plan Execution |

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
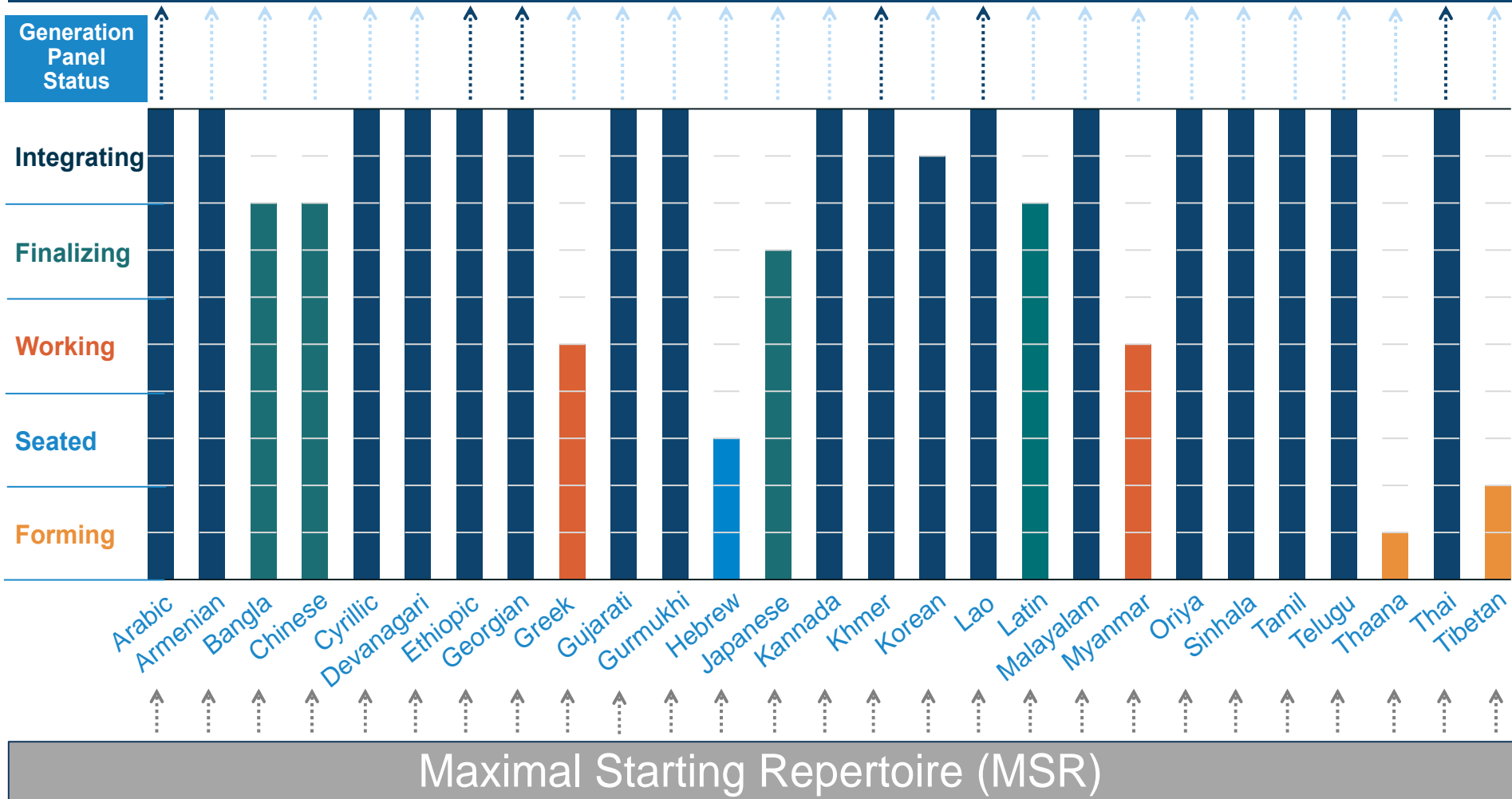 - ○ Determine variant labels of IDN TLDs

RZ-LGR Process

Generation Panel 1

Generation Panel 2

Generation Panel 3

Etcetera

One Generation Panel per script or writing system

Integration Panel

Needs more work

RZ- LGR

Script 1
Script 2

Etcetera

Status of Generation Panels (GPs)

Root Zone Label Generation Rules (RZ-LGR)

October 2018

Generation Panel Status

	Integrating	Finalizing	Working	Seated	Forming

Arabic, Armenian, Bangla, Chinese, Cyrillic, Devanagari, Ethiopic, Georgian, Greek, Gujarati, Gurmukhi, Hebrew, Japanese, Kannada, Khmer, Korean, Lao, Latin, Malayalam, Myanmar, Oriya, Sinhala, Tamil, Telugu, Thaana, Thai, Tibetan

Maximal Starting Repertoire (MSR)

Understanding IDN Variant TLDs

◉ **Security**

◉ **Usability**

.epic
0065 0070 0069 0063

.epic
0435 0440 0456 0441

.澳门
6FB3 95E8

.澳門
6FB3 9580

.السعودية
0627 0644 0633 0639 0648 062F 064A 0629

.السعودية
0627 0644 0633 0639 0648 062F 06CC 06C3
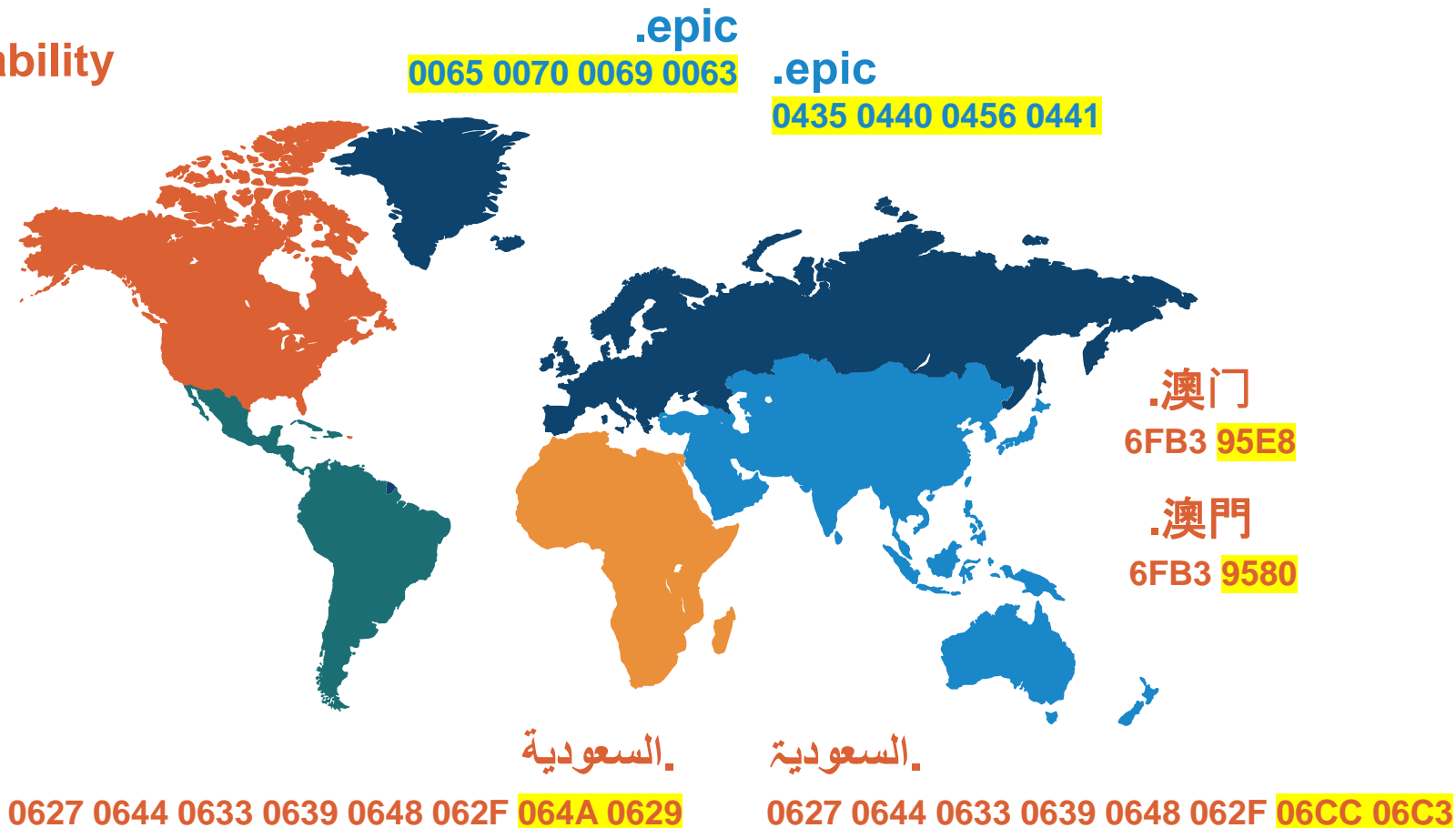
IDN Variant TLD Implementation

- 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
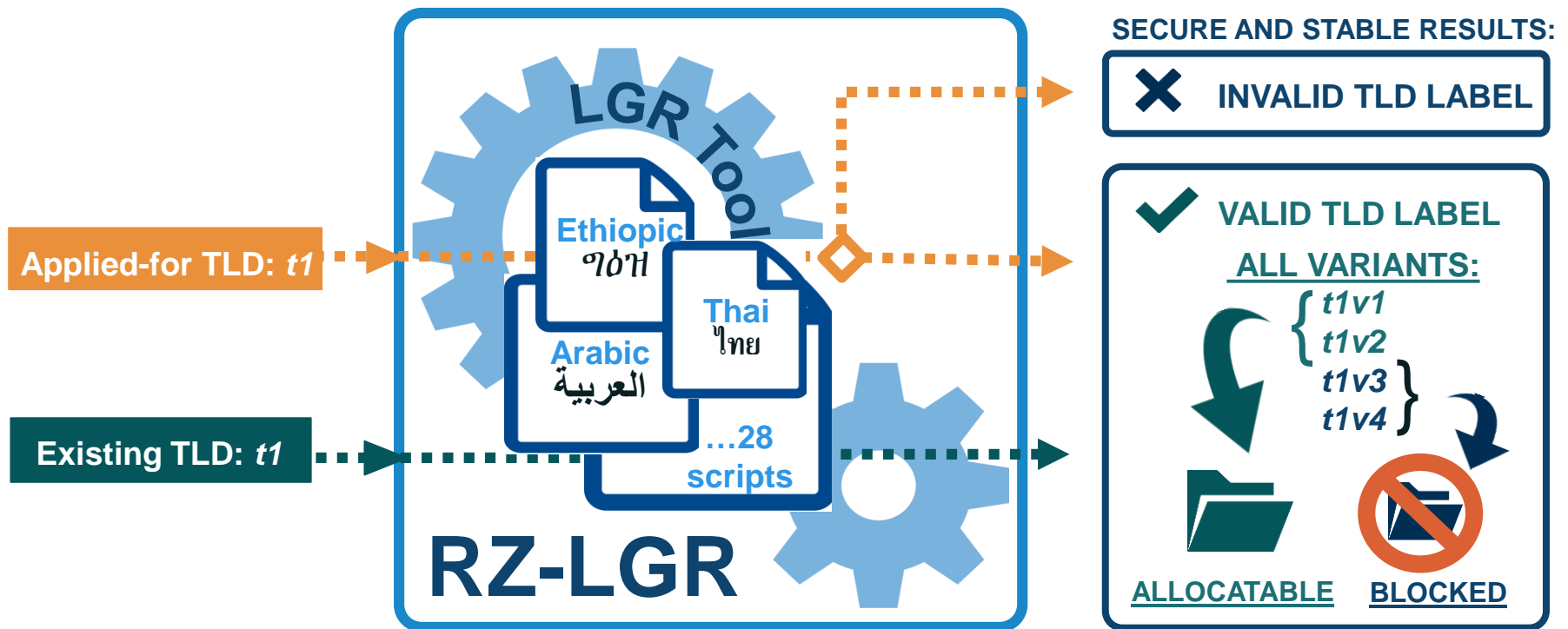
Status of IDN Variant TLDs – Management Mechanism

- 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

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** a 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
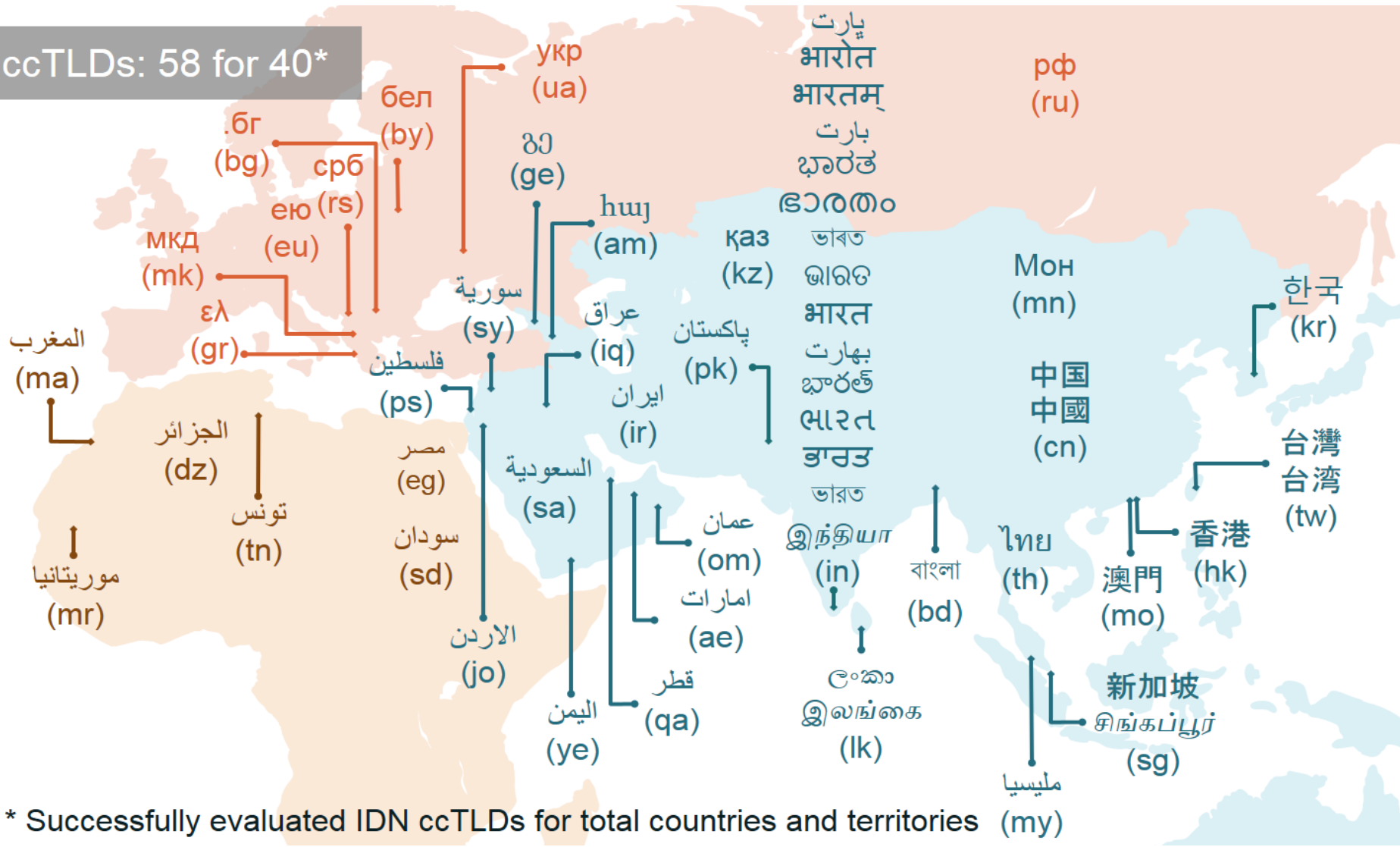
* Successfully evaluated IDN ccTLDs for total countries and territories

IDN ccTLD Fast Track Process

- ⊙ Launched in late 2009
  - ○ 58 IDN ccTLDs evaluated representing 40 countries/territories
  - ○ 56 IDN ccTLDs delegated representing 38 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 under guidance of 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

Status

Jul 2015	Oct 2015	ICANN 55 MARRAKECH 5 – 10 March 2016	ICANN 57 HYDERABAD	ICANN 58 COPENHAGEN	ICANN 60 ABU DHABI	ICANN 61 SAN JUAN	May 2018
Call for Community Experts	Working Group formed	Initial issues list presented	Interim draft presented	First draft for Public Comment presented	Final draft for Second Public Comment presented	Final Draft Presented to ICANN Board + Second Public Comment response presented	Final Proposed Ver. 4.0 published

Topics Covered and Next Steps

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

Transition (4)	Format of IDN Tables (2)	Consistency of IDN Tables and Practices (4)	IDN Variant Labels (3)

Similarity and Confusability of Labels (4)	Publishing IDN Registration Policy and Rules (1)	Terminology (1)	Additional Notes I - VII
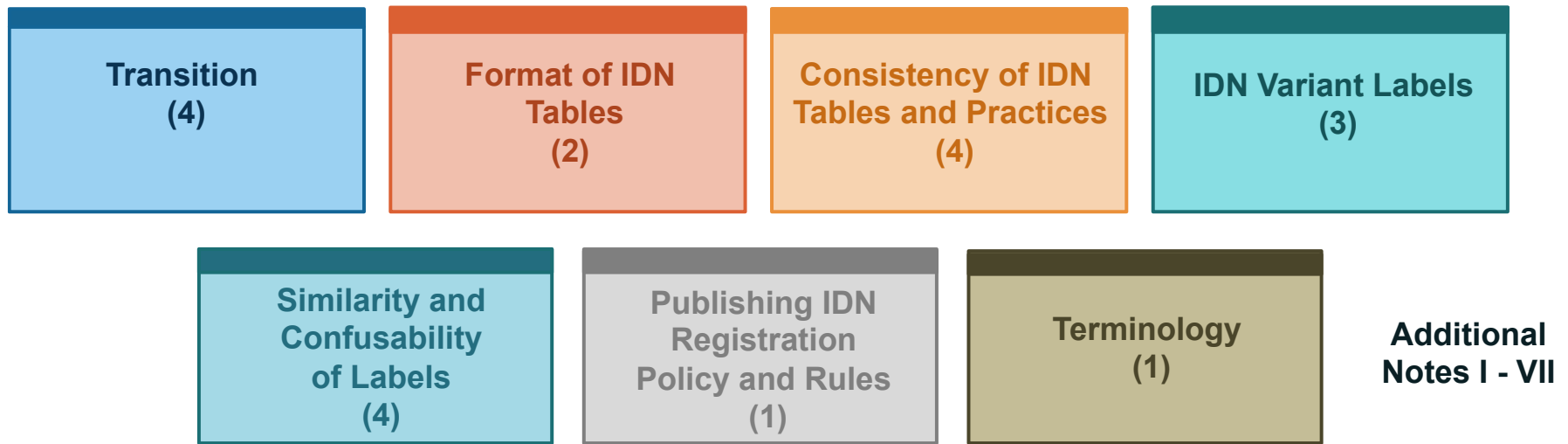
- Next steps: present IDN Guidelines 4.0 to ICANN Board for approval in Jan. 2019

Communication and Outreach Efforts

- ⊙ Direct outreach
  - ○ Neo-Brahmi GP F2F meeting, April 2017, India
  - ○ Myanmar GP kickoff meeting, June 2018, Myanmar
  - ○ Neo-Brahmi GP F2F meeting, July 2017, Bangladesh
  - ○ Africa DNS Forum, July 2018, Cotonou, Benin
  - ○ APTLD74, September, Uzbekistan
  - ○ Latin GP F2F meeting, October 2018, Belgium
  - ○ ICANN - WATRA - ITU Workshop, October 2018, Burkina Faso
- ⊙ 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 Integration Panel

Marc Blanchet
Member of Integration Panel

Agenda

- ⊙ Integration Panel Scope

- ⊙ IP Activities Summary (since ICANN 61, March 2018, San Juan)
  - ○ Reviews
  - ○ MSR
  - ○ Root-Zone LGR

Integration Panel Scope

- ⊙ "The Integration Panel is a panel of independent experts tasked with reviewing proposals presented by the Generation Panels and, if accepted, integrating them into a consistent set of Label Generation Rules for the Root Zone. The decisions by the integration Panel are required to be unanimous."
  - ○ https://community.icann.org/display/croscomlgrprocedure/Integration+Panel

- ⊙ "The integration Panel must take into account any public comments submitted in response to the posting of the generation panel's output."
  - ○ Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels, Version 2013-03-20b, March 20th 2013.

IP Activities Summary: Reviews

- ◉ GP formation proposals
  - ○ Myanmar

- ◉ Draft LGRs:
  - ○ Chinese
  - ○ Cyrillic[+]
  - ○ Japanese
  - ○ Latin
  - ○ Neo-Brahmi:
    - • Bengali, Devanagari*, Gurmukhi*, Gujarati*, Kannada*, Malayalam*, Oriya*, Tamil*, Telugu*
  - ○ Sinhala*

- ◉ * entered Public Comment since ICANN61
  [+] submission after Public Comments

- ⊙ LGR submitted after Public Comments
  - ○ Cyrillic
    - • Deferred for concurrent integration with Latin, Greek, Armenian

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

IP Activities: MSR

- ◉ MSR-3 Released on March 29th, 2018
  - ○ (3 CJK and 3 Latin codepoints added)

- ◉ MSR-4
  - ○ Currently not scheduled.
  - ○ Some additional Latin code points may be needed, but have not been requested

- ◉ Future MSR
  - ○ Currently not scheduled.
  - ○ Might include additional scripts when needed

IP Activities: Root Zone LGR-3

- ⊙ Integration Panel plans to produce a new Root Zone LGR (LGR-3)
  - ○ Target: ~~Q3 CY2018~~, Q1 CY2019
  - ○ Dependence: timely delivery of script LGRs (including public comments)
  - ○ Possible sets (only one set for LGR-3):
    - Neo-Brahmi scripts and Sinhala
    - CJK
    - Latin-Cyrillic-Greek-Armenian

Community Updates

- Latin GP Update
- Neo-Brahmi GP Update
- Sinhala GP Update
- Myanmar GP Update

Latin Generation Panel (Latin GP) Update

Mirjana Tasić
Latin GP Chair

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

- Summer 2016 - GP restarted with new call for volunteers. The GP seated on Monday, 15 May 2017 [Proposal for Formation of Latin Generation Panel](#)

- GP proposal for inclusion and exclusion principles were sent for an informal public review in September 2017

- During fall of 2017, GP has collected information from 209 languages

- GP proposed new code points for MSR-3

- GP submitted the code point repertoire to the Integration Panel in May 2018

- GP submitted the updated LGR proposal with the in-script variant analysis and the cross-script variant analysis in September 2018

- GP is proposing additional code points for next version of MSR.

- GP is currently reviewing IP feedback and finalizing the LGR proposal

Latin Script Geographic and Linguistic Spread
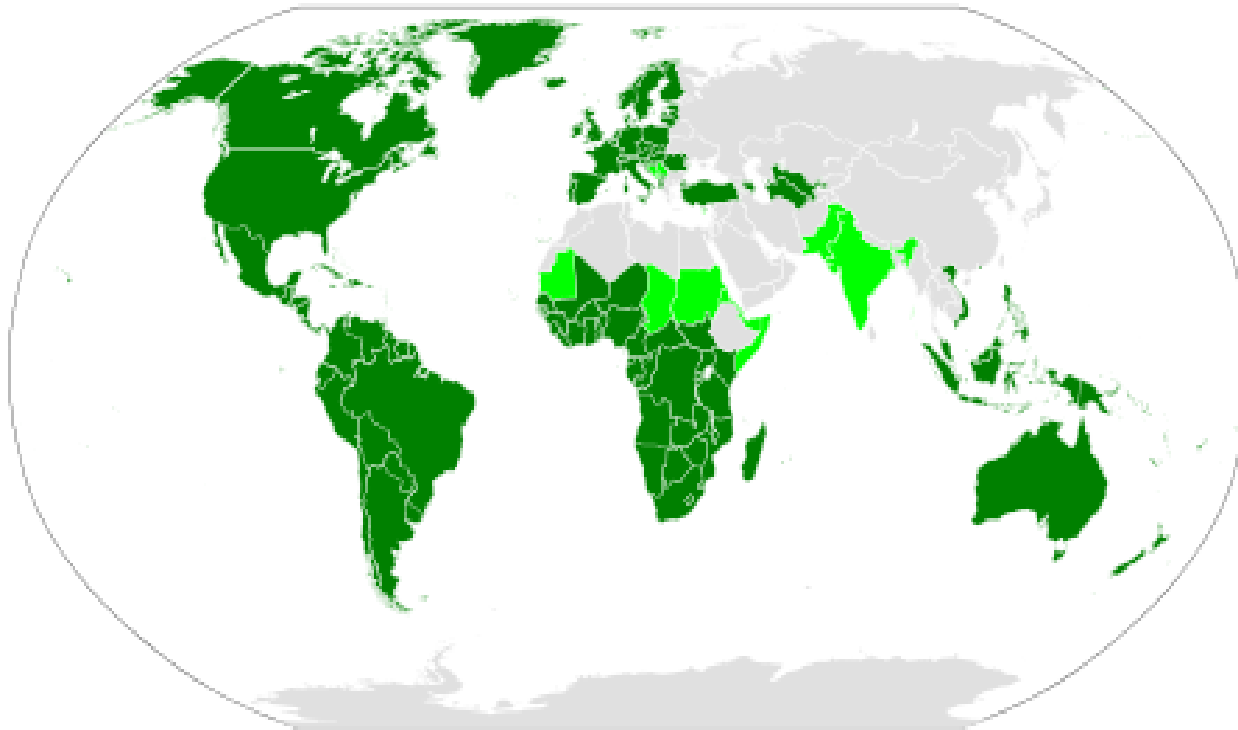
The dark green areas show the countries where the Latin script is the sole main script.
Light green shows countries where Latin co-exists with other scripts.
Grey areas - Latin-script alphabets are sometimes extensively used in areas
colored grey due to the use of unofficial second languages, such as French in Algeria
and English in Egypt, and to Latin transliteration of the official script, such as pinyin in
China or *rōmaji* in Japan.

Latin GP – Scope of Work for Code Point Analysis

- Maximal String Repertoire Version 3 (MSR-3)

  - 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

Latin GP – Members

◉ 14 members, 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
  - Not enough members to cover workload

- 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

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
  - 4 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 the MSR-3
  - ○ 3 non-MSR-3 code points are proposed for 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

- ◉ Submitted the second round proposal to the IP in September 2018

Latin GP – Project Timeline

Jul 2017	Jan 2018	May 2018	Sep 2018	Nov 2018	Dec 2018
Developing Principles	Developing code points repertoire and identifying variants	Review the code points repertoire with the IP	Review the variant rules with the IP	Finalize the Latin script LGR Proposal	Prepare Latin script LGR proposal for Public Comment
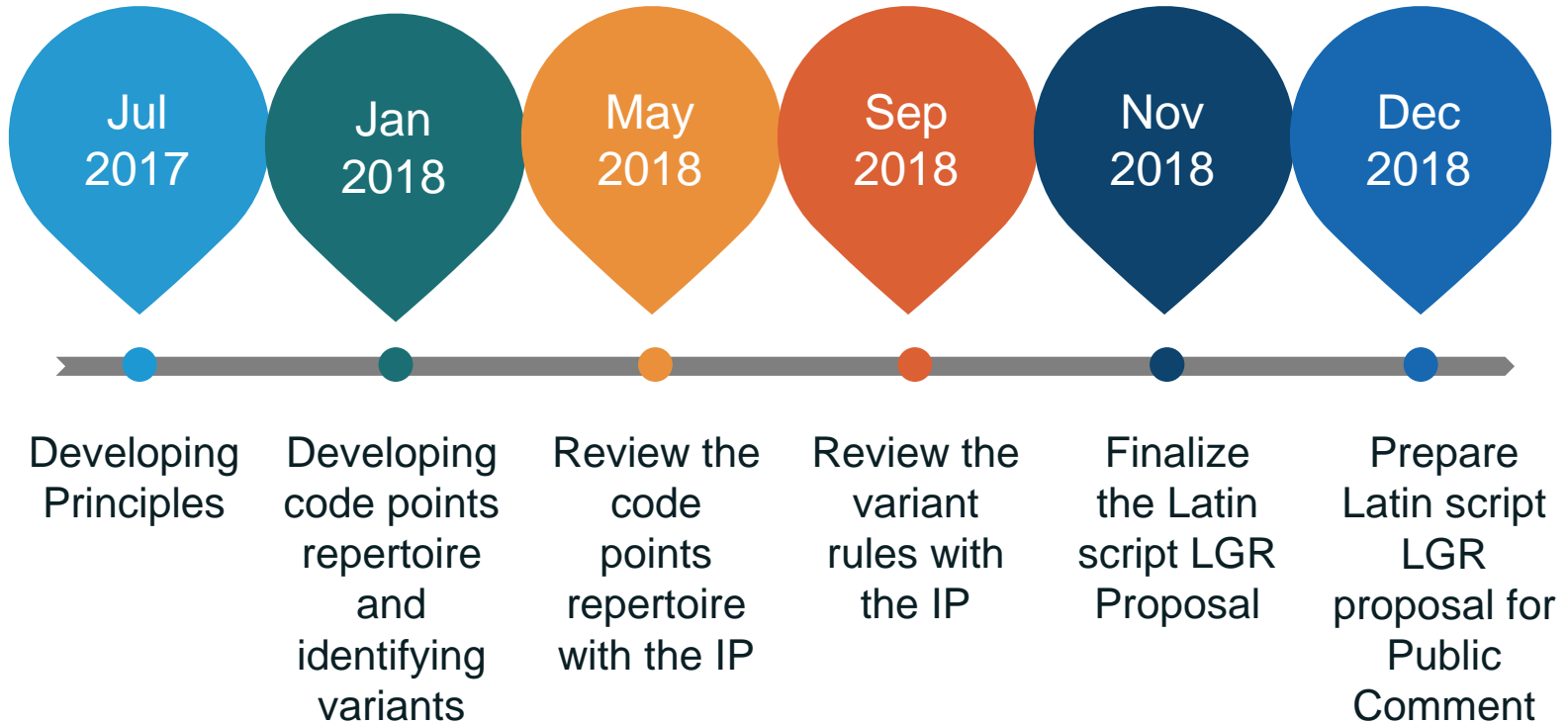
Questions?

Neo-Brahmi Generation Panel (NBGP) Update

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

Timeline

May 2010

.भारत

April 2018

J/A/S 2018

Egyp...

Sau...iah)

United Arab Emirates: امارات (Emarat)

| IDN ccTLD | ccTLD .bharat | NBGP Seated | IP feedback and Draft **Devanagari** LGR is getting prepared | F2F meeting in Colombo | Draft of **Gurmukhi, Gujarati, Kannada, Telugu** LGR are gettering prepared | July Aug Sept, 8 scripts work completed. |
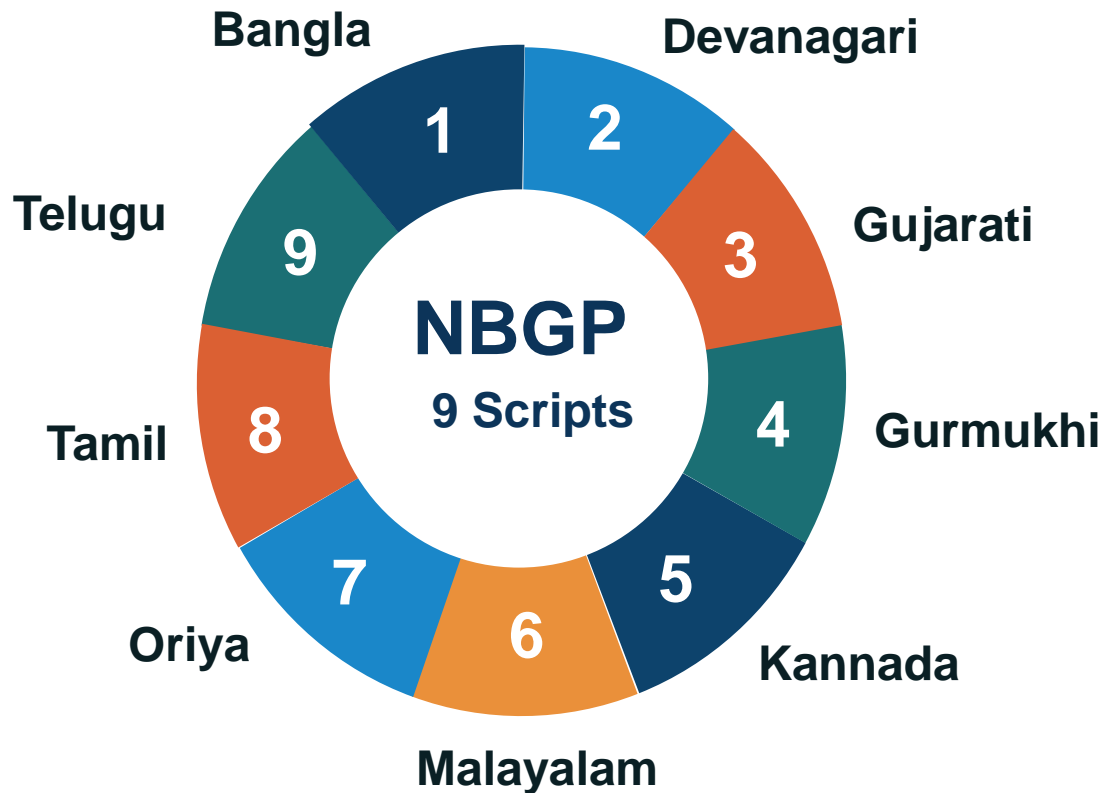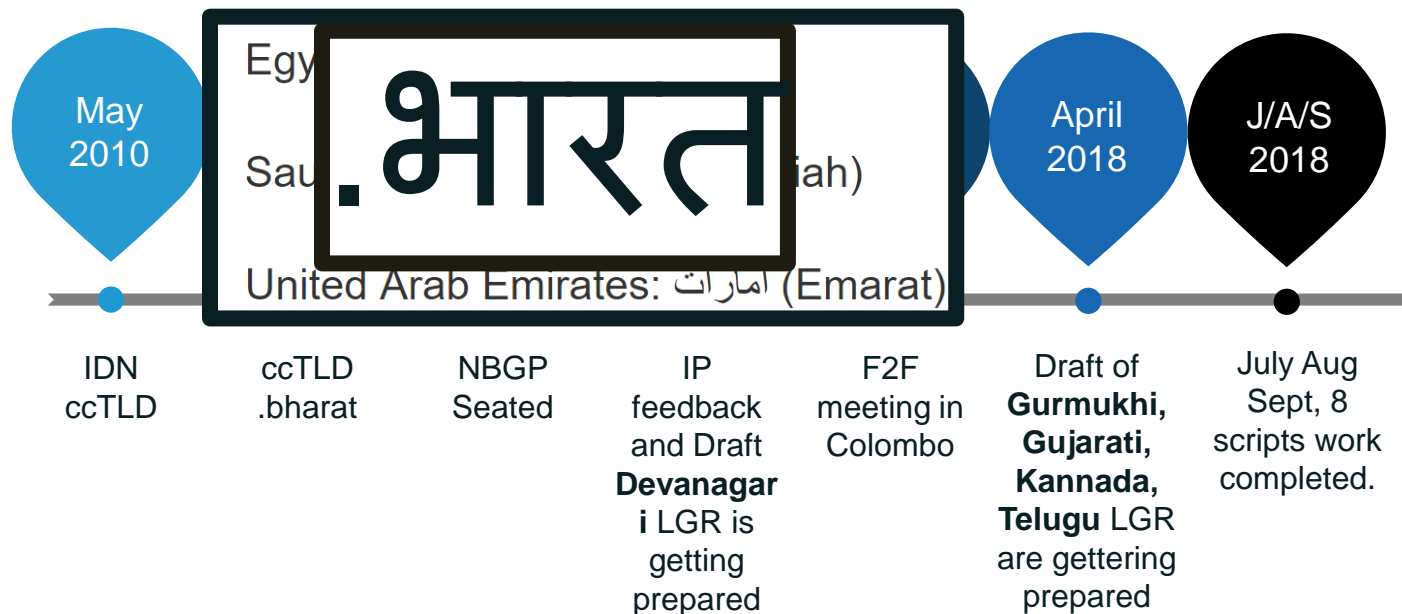
Nov/Dec 2018 - Bangla Script

Bangla script is close to completion as we are coordinating with Bangladesh too and hopefully soon the consensus will be made to go for public comment.

Public Comment Announcements

27Jul 2018	8 Aug 2018	25 Sept 2018	8 Oct 2018	Oct 2018
Devanagari, Gurumukhi and Gujarati	Kannada, Oriya and Telugu	Malayalam and Tamil	**First Two Public Comment Periods Closed**	Comments Analyzed by NBGP and ICANN
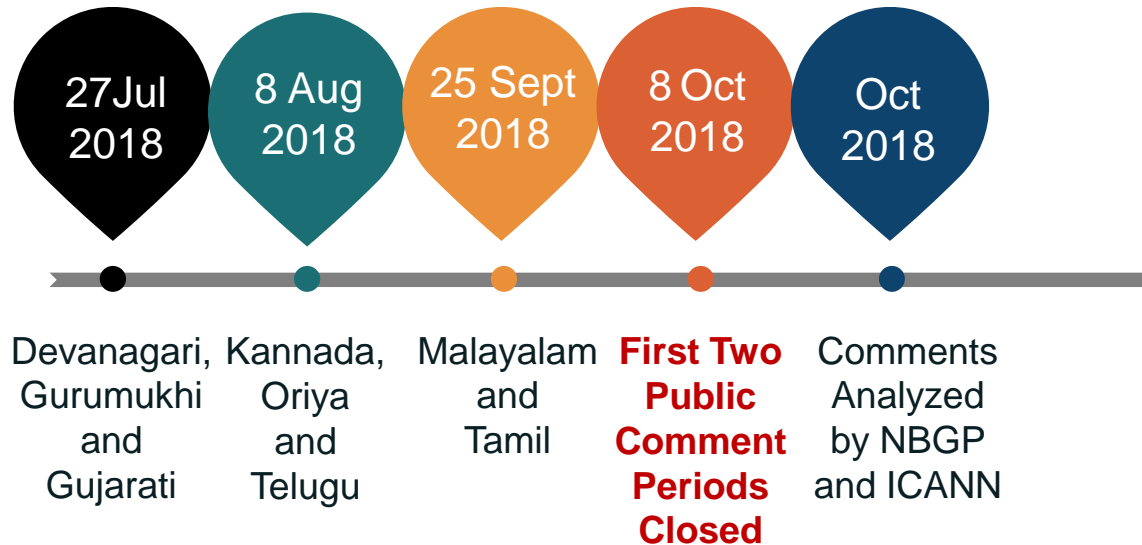
Summary

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

Cross-Script Variants for Devanagari and Gurmukhi

Code Point 2 + Glyph 2	Code Point 1 + Glyph 1	Type(blocked)
र (0930)	ਕ (0A15)	Appendix
ग (0917)	ਗ (0A17)	Blocked/ Norm
ढ (0922)	ਫ (0A2B)	blocked
उ (0909)	ਤ (0A24)	blocked
ट (091F)	ਟ (0A1F)	blocked
ठ (0920)	ਠ (0A20)	blocked
म (092E)	ਸ (0A38)	blocked
भ (092D)	ਮ (0A2E)	blocked
प्टि (092A 094D 091F 093F)	ਇ (0A07)	blocked
◌ं (0902)	◌ं (0A02)	blocked
ि◌ (093F)	ਿ◌ (0A3F)	blocked
◌ी (0940)	◌ੀ (0A40)	blocked
◌े (0947)	◌ੇ (0A47)	blocked
◌ै (0948)	◌ੈ (0A48)	blocked

TLD - Variant example for Devanagari

.ठग

TLD - Variant example for Gurmukhi
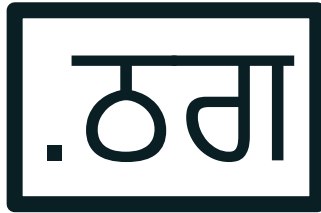
TLD - Variant example for Devanagari and Gurmukhi

Timeline

- ⊙ 8 Scripts – Completed

- ⊙ 1 Script – Close to completion

- ⊙ 18 months – From active work to finish

- ⊙ 60 Volunteers

Internationalized Domain Name: IDN

राजस्थान.भारत

मेल.डाटा.भारत

मेल.राहुलदेव.भारत

पत्रिका.भारत

मेल.माइक्रोसॉफ्ट.भारत

почта.рус

ડાટામેલ.ભારત

ਡਾਟਾਮੇਲ.ਭਾਰਤ

ডাটামেল.ভারত

ڈاٹامیل.بھارت

డేటామెయిల్.భారత్

डाटामेल.भारत

داد.امارات

இந.இந்தியா

电邮.在线

датамэйл.рф

датамэйл.рус

දත්තතැපැල.ලංකා

ดาต้าเมล.ไทย

डाटाइंफोसिस.भारत

एक्सजेनप्लस.भारत

Join & Contribute

Please join us on Twitter

@nbgp_community

Or

Visit out Wiki page for membership and current status
https://community.icann.org/display/croscomlgrprocedure/Neo-Brahmi+GP

Dr. Ajay Data
@ajaydata

Sinhala Generation Panel Update

Harsha Wijayawardhana
Sinhala GP Co-Chair

Agenda Overview

1

Scripts Covered and Where They are Used

2

Members of the GP

3

Code Points Repertoire

4

Variant Analysis

5

Whole Label Evaluation Rules

6

Next Steps

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

Sinhala GP – Members

- ⊙ Dr. Ruvan Weerasinghe
- ⊙ Mr. Harsha Wijayawardhana
- ⊙ Mr. Chamila Liyanage
- ⊙ Mr. Pathum Egodawatha
- ⊙ Mr. Viraj Welgama
- ⊙ Ms. Aruni Goonathilake
- ⊙ Mr. Chamara Dissanayake
- ⊙ Ms. Sagarika Wickramasekara
- ⊙ Prof. J.B.Dissanayake
- ⊙ Mr. Champika Wijayathunga
- ⊙ Mr. Rajeewa Abeygoonarathne
- ⊙ Rev. Mettavihari
- ⊙ Ms. Nimasha Dilshani

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	ඎaa	SINHALA LETTER IRUUYANNA	Usage unknown
2	0D8F	ඏ	SINHALA LETTER ILUYANNA	Usage unknown
3	0D90	ඐ	SINHALA LETTER ILUUYANNA	Usage unknown
4	0D9E	ඞ	SINHALA LETTER KANTAJA	Not in modern usage
5	0DA6	ඦ	SINHALA LETTER SANYAKA	Only used in the word 'ඉඦඃ' (this word is used to call dogs)
6	0DDF	ෟ	SINHALA VOWEL SIGN GAYANUKITTA	Usage unknown
7	0DF3	ෳ	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 ඝ (U+0D9D)
- බ (U+0DB6) and ඛ (U+0D9B)
- හ (U+0DC4) and භ (U+0DB7)
- ච (U+0DA0) and ව (U+0DC0)
- ඔ (U+0D94) and ඹ (U+0DB9)
- එ (U+0D91) and ඵ (U+0DB5)
- ඍa (U+0D8D) and සා (U+0DC3 U+0DD8)
- ඓ (U+0D93) and ඵෙ (U+0DB5 U+0DD9)
- ඒ (U+0D92) and ඵ් (U+0DB5 U+0DCA)
- ඕ (U+0D95) and ඹ් (U+0DB9 U+0DCA)

Cross-Script Variant Analysis

- ⊙ Sinhala GP concluded there is <u>no cross-script variant rules</u>
- ⊙ 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
ග (U+0D9C)	ഗ (U+0D17)
ශ (U+0DC1)	ശ (U+0D36)
◌ා (U+0DCF)	◌ാ (U+0D3E)

    ○ Sinhala and Myanmar

Sinhala	Myanmar
ග (U+0D9C)	တ (U+1010)
ශ (U+0DC1)	ဝ (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

Jan 2018	Mar 2018	July 2018	Sep 2018	Oct 2018	Dec 2018
Form the Sinhala GP	Finalize the code points repertoire, develop variant rules and WLE rules	Submit the first draft LGR proposal to the IP	Submit the second draft LGR proposal to the IP	Released for public comment	Finalize and submit to ICANN
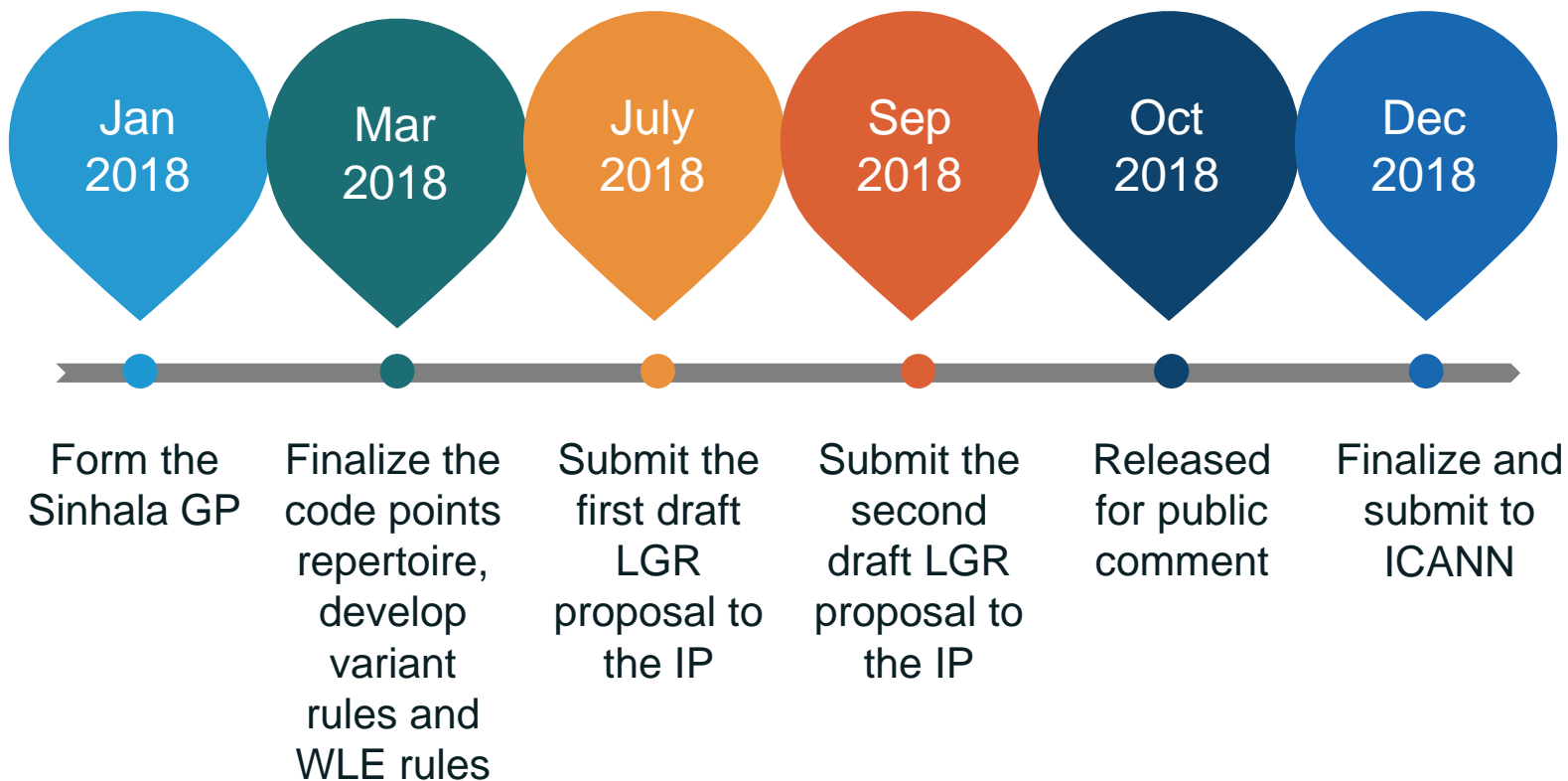
Myanmar Generation Panel Update

Thin Zar Phyo
Myanmar GP Chair

Agenda Overview

1

Scripts Covered and Where They are Used

2

Members of the GP

3

Code Points Repertoire

4

Variant Analysis

5

Whole Label Evaluation Rules

6

Next Steps

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	လိဂ်ႏတီး	3	3,295,000
Rakhine	[rki]	Myanmar	ရခိုင်	3	2,020,000
Karen, Sgaw	[ksw]	Myanmar, Thailand	စှီၢ	3	1,560,000
Mon	[mnw]	Myanmar, Thailand	မန်	5	851,000
Pa'O Karen	[blk]	Myanmar	ပအိုဝ်း	5	560,740

Myanmar GP – Members

- Dr. Myint Myint Than   Unicode & Policy Expert
- Dr. Khin Aye   Linguistics Expert
- Mr. Ngwe Tun   Unicode Expert
- Ms. Thin Zar Phyo   Software Development and NLP Expert
- Mr. Naing Win Oo   DNS & IDNA Expert
- Mr. Kaung Khant Zaw   Software Development and Unicode Expert
- Ms. Yin May Oo   Computational Linguistics Expert
- Mr. Ye Zarni Aung   Unicode and DNS Expert
- Mr. Min Paing Khant Oo   Unicode and DNS Expert
- Mr. Thura Soe   Software Development and NLP Expert
- Mr. Sai Zin Di Di Zone   Shan Language Expert

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	ဣ	u1000 + u1039 + u1000	ဣ
2	u1029	ဩ	u101E + u103C	ဩ
3	u102A	ဪ	U1029 + u1031+u102C+u103A	ဪ
4	u102A	ဪ	u101E+ u103C + u1031+ u 102C+ u103A	ဪ
5	u1061	ၡ	u101B + u103E	ၡ
6	u1009 + u103A	၉	u1025 + u103A	၉
7	u1009 + u1037 + u103A	၉	u1025 + u1037 + u103A	၉
8	u107E	ၾ	u107D + u103E	ၾ

In-Script Variant Analysis

⊙ Myanmar-Malayalam

No.	Glyph	Code Point	Myanmar Character Name	Glyph	Code Point	Malayalam Character Name
1	ဂ	U+1002	MYANMAR LETTER GA	ഋ	U+0D31	MALAYALAM LETTER RRA
2	ဝ	U+101D	MYANMAR LETTER WA	ഠ	U+0D20	MALAYALAM LETTER TTHA

⊙ Myanmar-Oriya

No.	Glyph	Code Point	Myanmar Character Name	Glyph	Code Point	Oriya Character Name
1	ဝ	U+101D	MYANMAR LETTER WA	ଠ	U+0B20	ORIYA LETTER TTHA
2	ေ◌	U+1031	MYANMAR VOWEL SIGN E	େ◌	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	ၪ	u1008	ഖ	U+0D31	The sequence U+1005 U+103B is invalid
2	�	u1009 + u102C	ഠ	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	ယ	U+101A	MYANMAR LETTER YA	ധ	U+0D27	MALAYALAM LETTER DHA
3	ကာ	U+1000, U+102C	MYANMAR LETTER KA , MYANMAR VOWEL SIGN AA	ന്ന	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	ဥ	U+1025	MYANMAR LETTER U	උ	U+0D8B	SINHALA LETTER UYANNA
2	ာ	U+102C	MYANMAR VOWEL SIGN AA	ා	U+0DCF	SINHALA VOWEL SIGN AELA-PILLA
3	ေ	U+1031	MYANMAR VOWEL SIGN E	ෙ	U+0DD9	SINHALA VOWEL SIGN KOMBUVA
4	ော	U+1031, U102C	Sequence: MYANMAR VOWEL SIGN AA, MYANMAR VOWEL SIGN E	ො	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 $့
T_LONG	→	1038 $း
K	→	Killer: 103A $်
VIRAMA	→	1039
M	→	Dependent Consonant Sign
C1	→	103F ဿ
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 ဥ can only be followed by 102F $ ဖ + (1036 $ or 1038 $ း)
6. 1026 ဦ can only be followed by 1038 $ း
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 1023ဣ] or 1037 $ ့
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

Jun 2018	July-Sep 2018	Oct 2018	Nov 2018	Dec 2018	Jan 2019
Form the Myanmar GP	Develop the code points repertoire, variant rules and WLE rules	Submit the first draft LGR proposal to the IP	Public workshop in Yangon	Released for public comment	Finalize and submit to ICANN

Engage with ICANN and IDN Program

Thank You and Questions

Visit us at **icann.org/idn**
Email: IDNProgram@icann.org

 @icann

facebook.com/icannorg

youtube.com/icannnews

flickr.com/icann

linkedin/company/icann

slideshare/icannpresentations

soundcloud/icann