
ICANN
ANNUAL GENERAL
66
MONTRÉAL
2–7 November 2019

IDN Program Update

ICANN66

6 November 2019

Overview of Session Presentations

- IDN Program Overview and Progress (Pitinan Kooarmornpatana)
- Update by Integration Panel (Asmus Freytag / Marc Blanchet)
- Community Updates
  - Latin GP (Mirjana Tasić)
  - Chinese GP (Wang Wei)
  - Japanese GP (Yoneya Yoshiro)
  - Korean GP (Kim Kyongsok)
  - Neo-Brahmi GP (Ajay Data)
  - Myanmar GP (May Oo)
- Q and A

IDN Program Overview and Progress

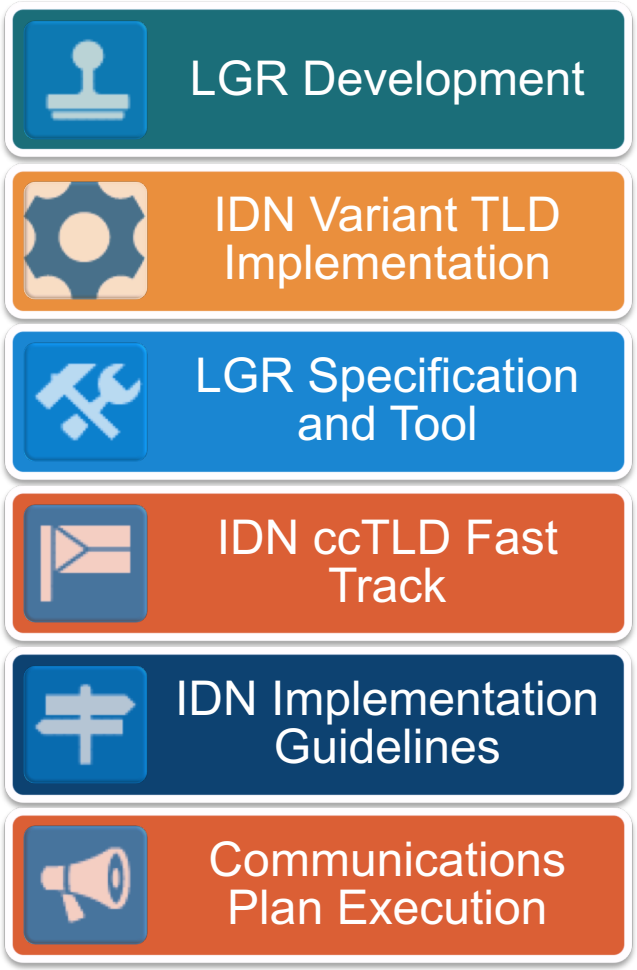
Pitinan Kooarmornpatana
Manager, IDN Programs

IDN Program Objectives

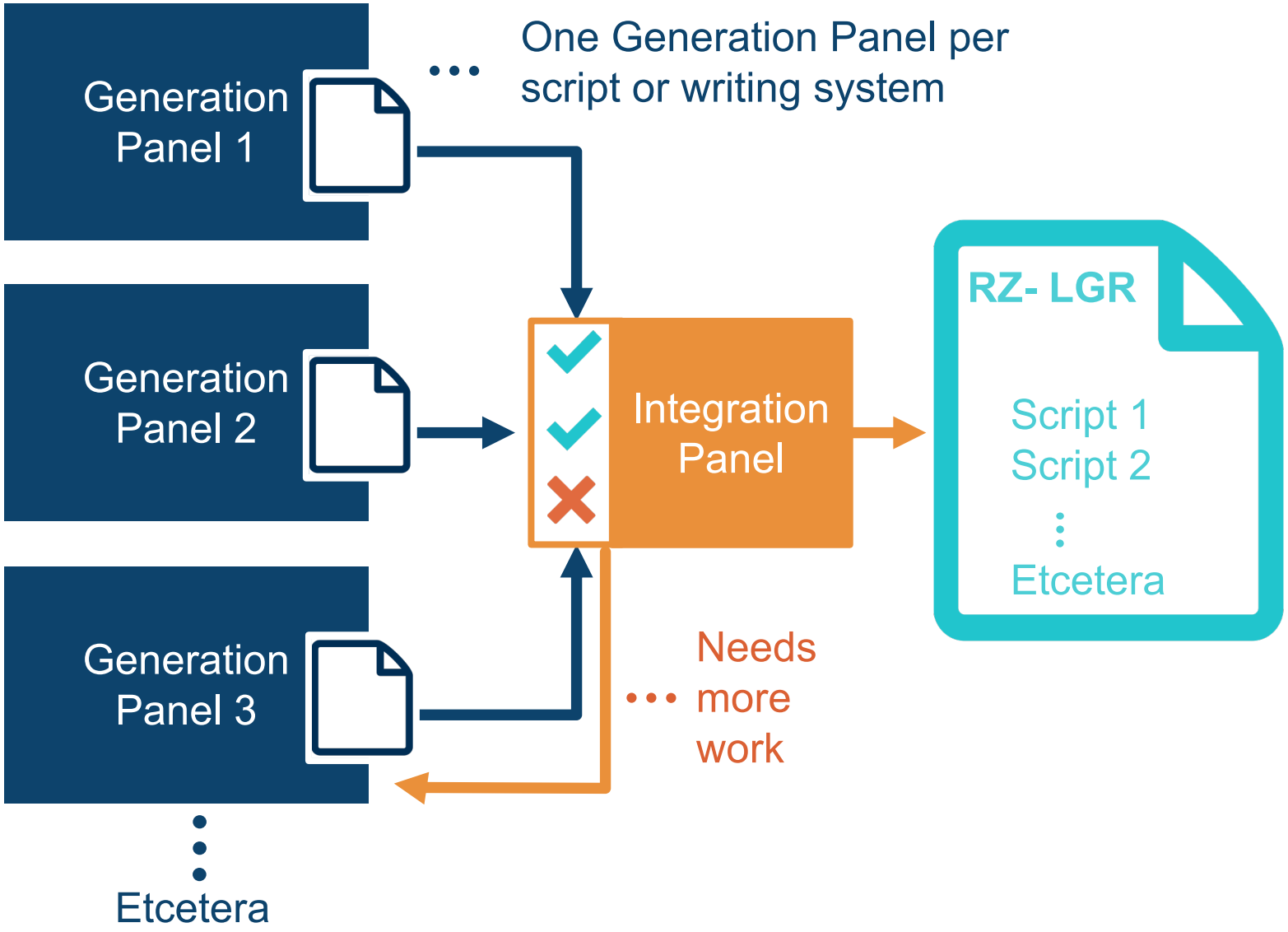
Enable deployment of domain names in the ***local languages and scripts*** of global communities in a ***secure and stable*** manner.

Overview of IDN Programs

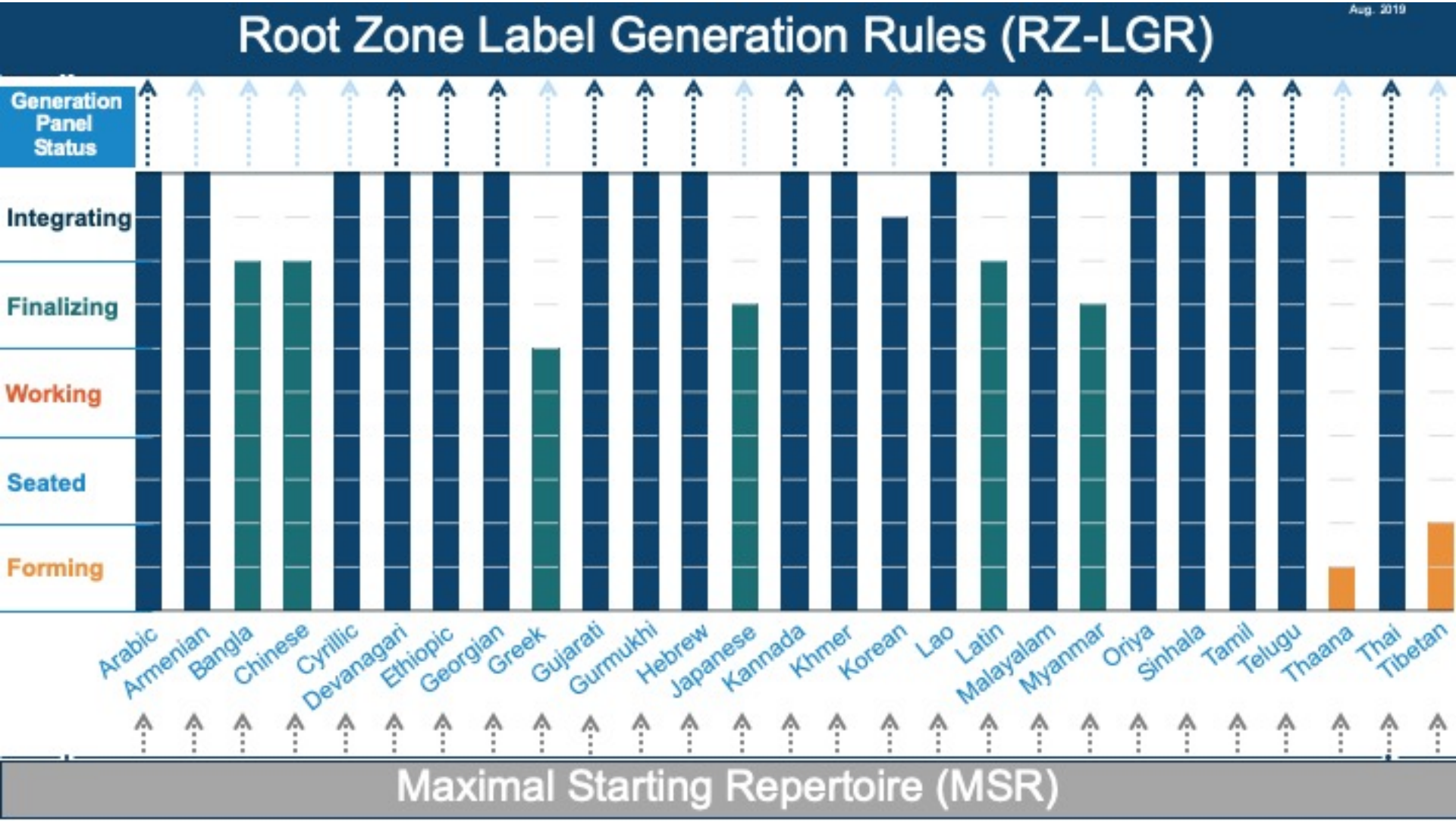
- IDNs at Top Level
  - IDN TLD Program
    - Root Zone Label Generation Rules (RZ-LGR)
    - IDN Variant TLD Implementation
    - LGR Toolset
  - IDN ccTLD Fast Track Process
- IDNs at Second Level for gTLDs
  - IDN Implementation Guidelines
- Community Outreach and Involvement



RZ-LGR Process



Status of Generation Panels (GPs)



Understanding IDN Variant TLDs

- Security
- Usability

Status of IDN Variant TLDs – Management Mechanism

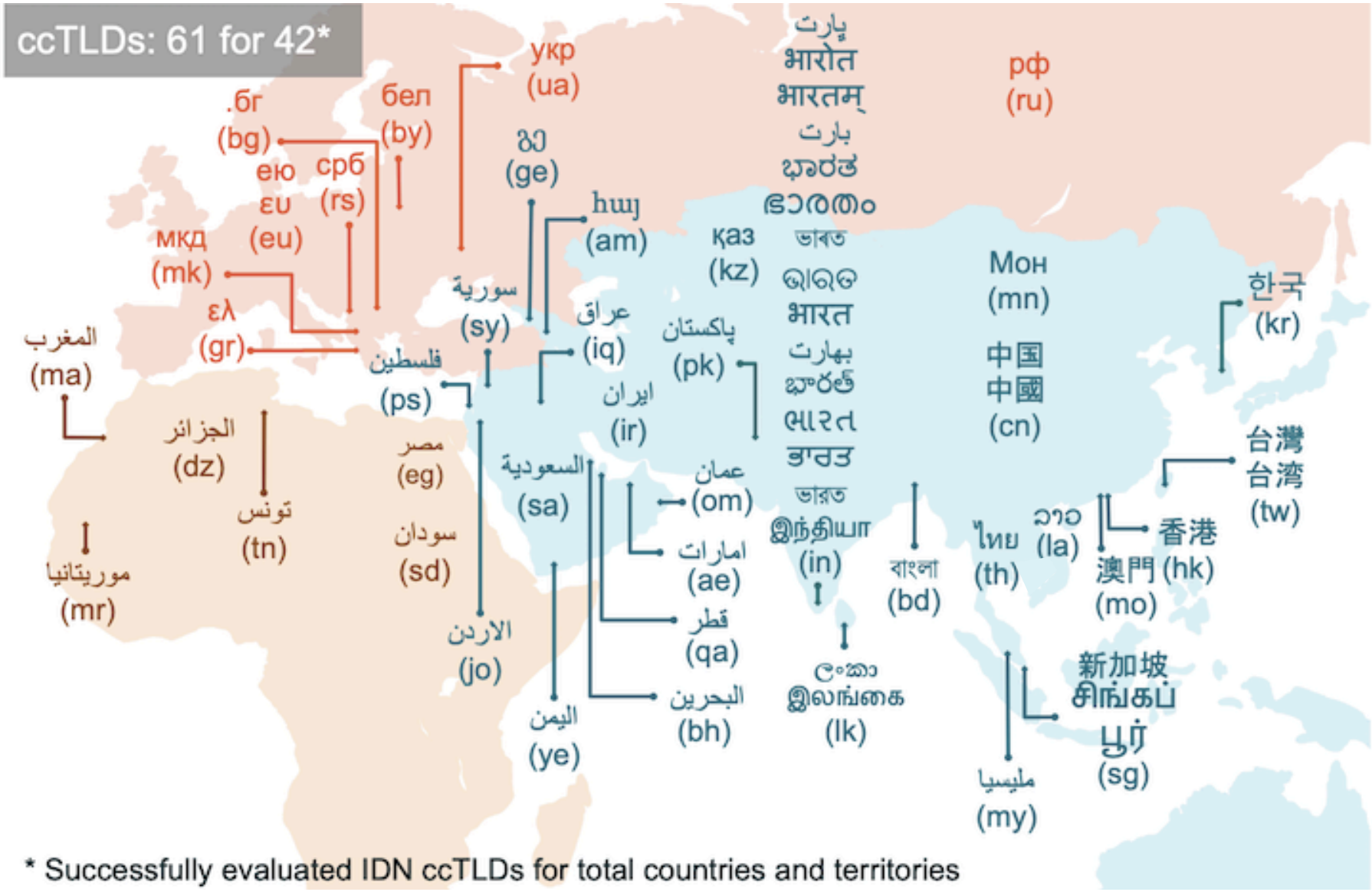
- Solution: ICANN org to work with the community to develop a feasible management mechanism.
  - Recommendations developed by ICANN org.
  - Recommendations presented to ICANN Board on 22 June 2018.
  - Recommendations released for Public Comment on 25 July 2018.
  - Updated version posted 25 January 2019.
  - ICANN Board approved 8 March 2019:

    "…the Board approves the Variant TLD Recommendations and requests that the ccNSO and GNSO take into account the Variant TLD Recommendations while developing their respective policies to define and manage the IDN variant TLDs for the current TLDs as well as for future TLD applications."

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 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 Implementation Guidelines

- IDN Guidelines developed to minimize the risk of cybersquatting and consumer confusion.
- IDN Guidelines apply to IDN registration policies at the second-level and practices under top-level domains (TLDs).
- Guidelines for the Implementation IDNs Version 4.0 finalized by IDN Guidelines Working Group and published on 10 May 2018.
  - Guidelines recommended a two-phase implementation:
    - Phase 1: Guidelines effective six months after board approval.
    - Phase 2: Guidelines 6a, 11, 12, 13, 18, and 19 effective 18 months after board approval.
- On 30 April 2019, GNSO made a request to the ICANN Board to allow it to study the guidelines before the implementation.

Communication and Outreach Efforts

- Direct outreach
  - UA/IDN workshop, 17-20 April 2019, Beijing, China.
  - Moscow registrar meeting, 6 June 2019, Moscow, Russia.
  - African Internet Summit Uganda, 15-17 June 2019, Kampala, Uganda.
  - Africa DNS Forum, 22-25 July 2019, Gaborone, Botswana.
  - Myanmar Script LGR Local Public Consultation, 10 August 2019, Yangon, Myanmar.
  - World Internet Conference, 18-22 October 2019, Wuzhen, China.
- Updates at ICANN meetings
- IDN web pages at icann.org/idn
- IDN community wiki pages
- IDN mailing lists: {lgr, ChineseGP, GreekGP, …}@icann.org

Thank you

Update by RZ-LGR Integration Panel

Asmus Freytag and Marc Blanchet
Members, Integration Panel

IP Activities Summary

- IP Activities Summary (since ICANN64, March 2019, Kobe, Japan)
  - Reviews
  - Root-Zone LGR-3
  - MSR
  - LGR-4 Plan
  - Future Work Model

IP Activities Summary: Reviews

- GP formation proposals
  - None
- Draft LGRs:
  - Bangla
  - Chinese
  - Greek
  - Myanmar
- Reviewed and integrated:
  - Hebrew
  - Neo-Brahmi:
    - Devanagari, Gurmukhi, Gujarati, Kannada, Malayalam, Oriya, Tamil, Telugu.
  - Sinhala

IP Activities: Root Zone LGR-3

- Publication process of Root Zone LGR-3
  - Added scripts: Devanagari, Gujarati, Gurmukhi, Hebrew, Kannada, Malayalam, Oriya, Sinhala, Tamil, and Telugu.
  - Various integration and verification tasks to create the final LGR.
  - Submitted for Public Comment
    - Period: 25 April – 4 June 2019
    - Confirmation of review from various GPs (Devanagari, Gujarati, Gurmukhi, Hebrew, Kannada, Oriya, Sinhala, Tamil)
    - One issue was submitted on Telugu regarding a rule restricting some Telugu words.
      - The GP agreed and submitted a change to the IP (by removing the rule) and sent it to Public Comment.
  - Final publication:
    - Includes the Telugu rule change.
    - Published on 10 July 2019.

IP Activities: MSR

- MSR-4 was published on 7 February 2019.
- Initial review of new Unicode additions since version 6.3.
  - About a dozen code points, mostly Arabic and Latin.
- No other real activity for MSR-5.
  - No requests yet to add any code points.
  - No requests for additional scripts.
- Over time, some codepoints were found by GP as not relevant (not used, historic, …), so MSR-5 might remove some of them (~<10)
- No strong need currently seen for MSR-5.

IP Activities Summary: LGR-4 Plan

- Expected LGRs to be reviewed in the next months, for final versions and integration for LGR-4
  - Bangla
  - Chinese
  - Greek
  - Japanese
  - Korean
  - Latin
  - Myanmar
- LGR-4 would also integrate the deferred scripts: Armenian and Cyrillic
- Current timeline assumes LGRs for LGR-4 will enter public comment no later than Q1/2020.

IP Activities: Future Work Model

- ICANN and IP intend to finalize the active LGRs and publish LGR-4 by end of FY2020.
- Therefore, FY2021 will see lower level of activities for IP.
- Currently discussing a work model of having IP inactive for a time and then become active at a fixed date to review the submitted LGRs (two times during the fiscal year).

Community Updates

- Latin Generation Panel
- Chinese Generation Panel
- Japanese Generation Panel
- Korean Generation Panel
- Neo-Brahmi Generation Panel
- Myanmar Generation Panel

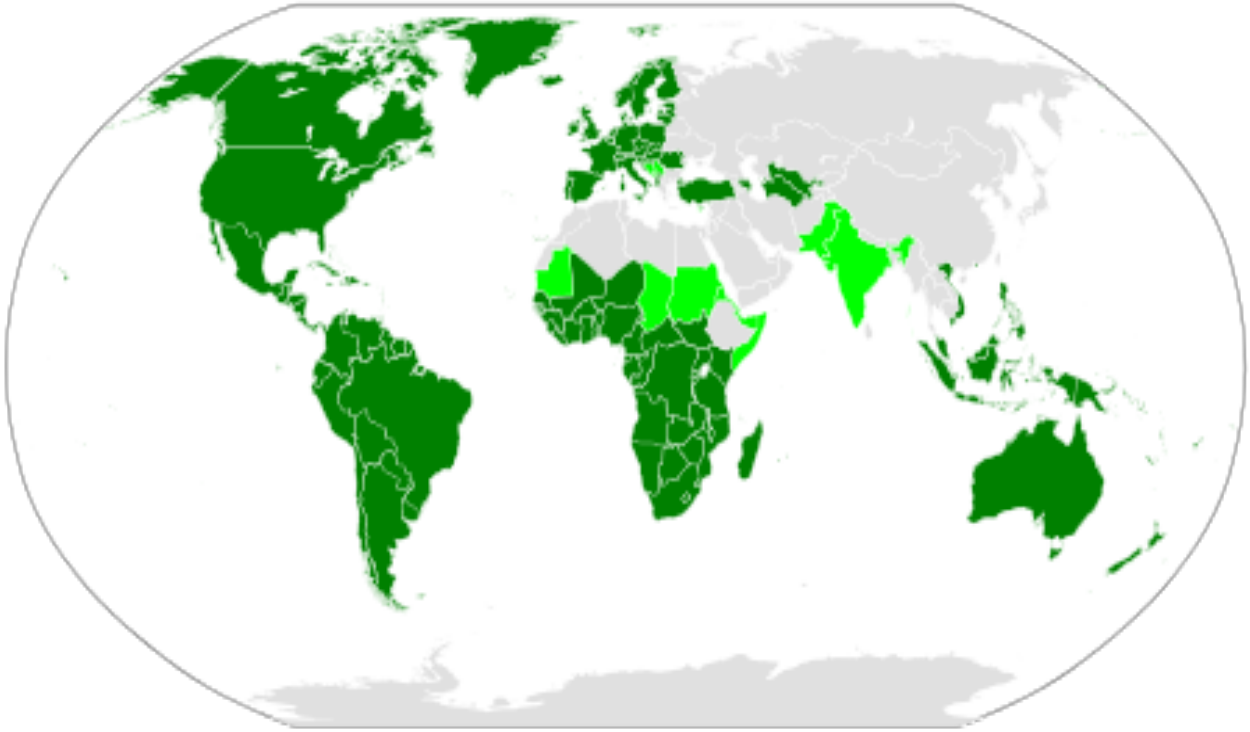
Update by Latin Generation Panel (GP)

Mirjana Tasić
Chair, Latin GP

Agenda Overview

1. Scope of work
2. Work Accomplished
3. Update since ICANN64
4. Project Timeline

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.

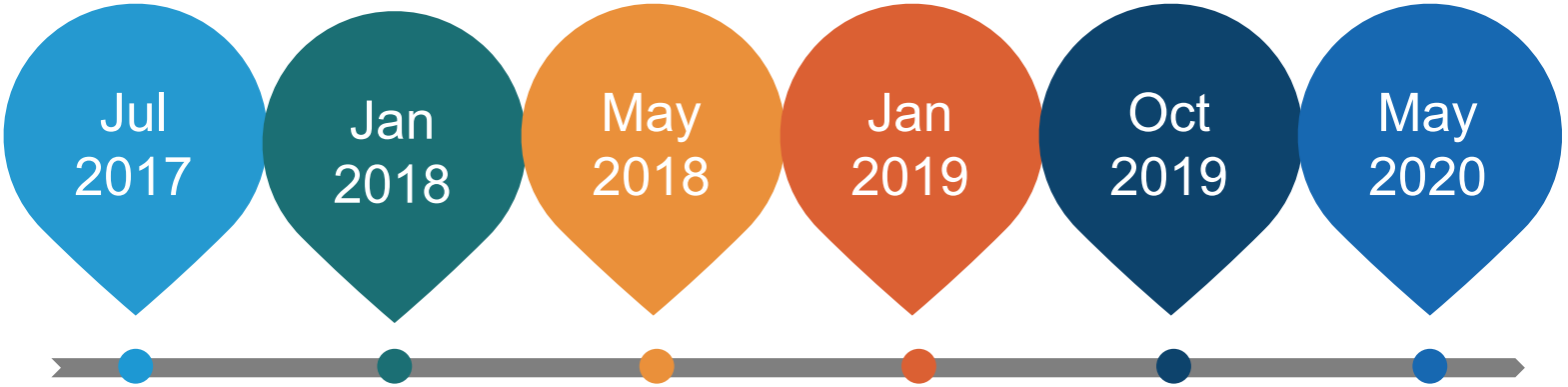
Latin GP – Work Accomplished

- Developing Repertoire
  - 181 of 181 EGIDS 1- 4 languages processed.
  - 29 EGIDS 5 languages processed.
  - 193 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.
  - 22 Code Point Sequences identified.
- Developing Variants
  - In-script variants still ongoing (80 percent finished).
  - Cross-script variants with Armenian script defined.
  - Cross-script variants with Cyrillic script defined.
  - Cross-script variants with Greek script defined.
  - Special HTML Link (underlining) analysis completed.
  - IDNA2003 compatibility analysis completed.
- Submitted the third version of proposal to the IP in May 2018.
- Submitted the fourth version of proposal to the IP in January 2019.
- Submitted the fifth version of proposal to the IP in October 2019.

Progress Since ICANN 64 - Variant Analysis

- In-script variant analysis
  - Visual variants
  - Non-visual variants
- Cross-script variant analysis
  - Greek script – with new input from Greek LGR
- Other considerations
  - Basic shapes (e.g., circle “o”, single line “l”, and crescent “c” or “ɔ”) within all scripts
  - Underlining analysis
  - IDNA2003 Compatibility

Latin GP – Project Timeline



Jul 2017	Jan 2018	May 2018	Jan 2019	Oct 2019	May 2020
Develop principles.	Develop code points repertoire and identify variants.	Submit proposal with code points repertoire.	Submit proposal adding cross-script variant analysis.	Submit proposal adding in-script variant analysis.	Finalize the LGR proposal and publish for Public Comment.

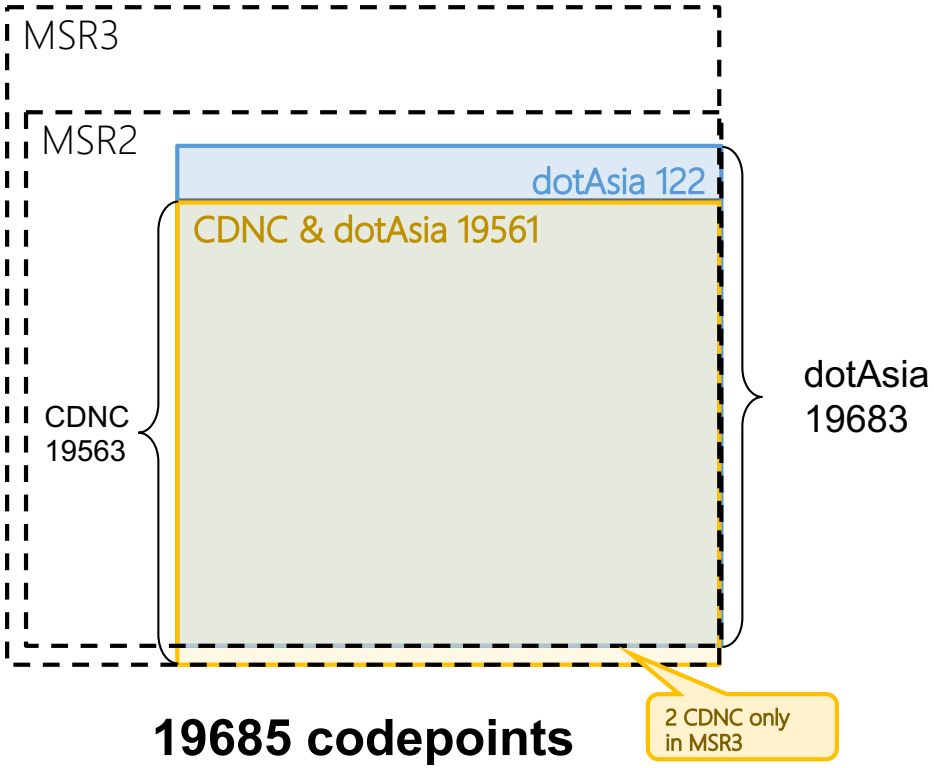
Update by Chinese Generation Panel (GP)

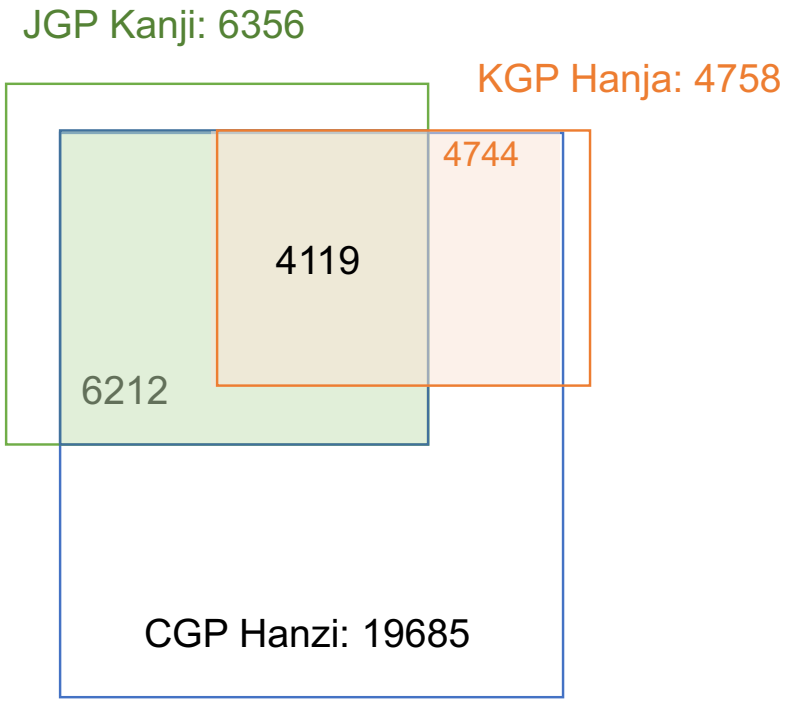
Wang Wei
Co-chair, Chinese GP

Agenda

1. Repertoire
2. Variant Mapping
3. Topic for Discussion

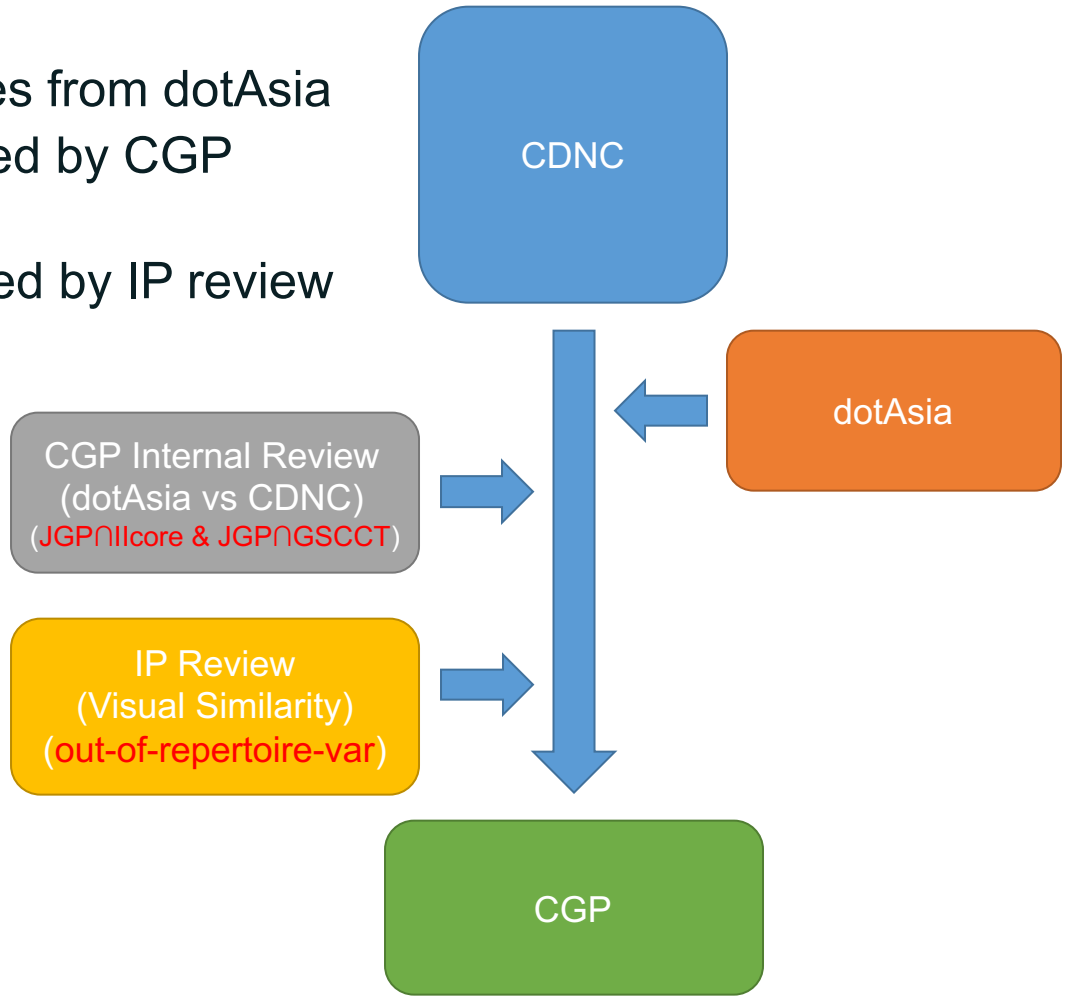
CGP repertoire = CDNC'2018 + dotAsia





19685 Variant mapping entries

- 19498 basic variant mapping entries from CDNC-2018
- 143 variant mapping entries from dotAsia
- 38 variant mappings revised by CGP internal review team
- 6 variant mappings changed by IP review



Multiple allocatable variant labels

Category	number	Original	Allocatable Simp	Allocatable Trad
Case 1	7	A	AB	A
Case 2	1	A	AB	C
Case 3	2	A	BC	A
Case 4	146	A	A	AB
Case 5	27	A	A	BC
Case 6	9	A	A	ABC
Case 7	2	A	A	ABCD

<action disp="allocatable" only-variants="simp r-simp both r-both **simp-1**" comment="all simplified label type 1" />
<action disp="allocatable" only-variants="simp r-simp both r-both **simp-2**" comment="all simplified label type 2" />

<action disp="allocatable" only-variants="trad r-trad both r-both **trad-1**" comment="all traditional label type 1"/>
<action disp="allocatable" only-variants="trad r-trad both r-both **trad-2**" comment="all traditional label type 2"/>

<action disp="blocked" any-variant="simp trad both **simp-1 simp-2 trad-1 trad-2**" comment="block any other mixed labels" />

Multiple allocatable variant labels

Proposed	Disposition	Description	Implemented
type: r-both-ms	Blocked	*r-both-ms indicates that for a given code point, its reflexive type is inherently r-both, but there is at least another 'simp' type (or other simplified types), and therefore it is preferred in a traditional context. Therefore, it is to be treated as a 'r-trad'.* The simp label contains "r-both-ms" char be BLOCKED	type: r-trad comment: r-both-ms
	Allocatable	The trad label contains "r-both-ms" char be ALLOCATABLEOriginal reflexive label contains "r-both-ms" be ALLOCATABLE	
type: r-both-mt	Blocked	*r-both-mt indicates that for a given code point, its reflexive type is inherently r-both, but there is at least another 'trad' type (or other traditional types), and therefore it is preferred in a simplified context. Therefore, it is to be treated as a 'r-simp'.* The trad label contains "r-both-mt" char be BLOCKED	type: r-simp comment: r-both-mt
	Allocatable	The simp label contains "r-both-mt" char be ALLOCATABLEOriginal reflexive label contains "r-both-mt" be ALLOCATABLE	
type: r-simp-m	Blocked	*r-simp-m indicates that for a given code point, its reflexive type is inherently r-simp, but there is at least another 'simp' type (or other simplified types), along with another 'trad' type and therefore it is never preferred in any variant labels. Therefore, it is to be treated as a 'r-neither'.* The simp label contains "r-simp-m" be BLOCKED	type: r-neither comment: r-simp-m
	Allocatable	The original reflexive label contains "r-both-m" be ALLOCATABLE	
type: trad-m	Blocked	*Allocatable trad is rare used, not in Modern Chinese Common Used Table in China mainland, nor Common used Chinese standard table in Taiwan. Set the allocatable trad as "trad-m" (muted) and treat it as a 'blocked'.* The trad label contains "trad-m" be BLOCKED	type: blocked comment: trad-m
type: simp-1	Allocatable	*Among the multiple allocatable simplified variants, set the allocatable simp with the smallest hex-code as "simp-1"* The simp label contains "simp-1" be ALLOCATABLE	type: simp-1
	Blocked	The simp label contains BOTH "simp-1" and "simp-2" be BLOCKED	
type: simp-2	Allocatable	*Among the multiple allocatable traditional variants, set the allocatable simp with the largest hex-code as "simp-2"* The simp label contains "simp-2" be ALLOCATABLE	type: simp-2
	Blocked	The simp label contains BOTH "simp-1" and "simp-2" be BLOCKED	
type: trad-1	Allocatable	*Among the multiple allocatable traditional variants, set the allocatable trad with the smallest hex-code as "trad-1"* The trad label contains "trad-1" be ALLOCATABLE	type: trad-1
	Blocked	The trad label contains BOTH "trad-1" and "trad-2" be BLOCKED	
type: trad-2	Allocatable	*Among the multiple allocatable traditional variants, set the allocatable trad with the largest hex-code as "trad-2"* The trad label contains "trad-2" be ALLOCATABLE	type: trad-2

Visual Similarity Variants

1） https://www.unicode.org/Public/security/11.0.0/confusables.txt

53E3	口	56D7	囗
571F	土	58EB	士
58AB	壿	58FF	壿
676E	柿	67FF	柿
8D7F	趐	8D86	趐
9E42	鶹	9E43	鶻

Three non-modern used pairs will be treated as visual identical variants
-- 676E柿 & 67FF柿、8D7F趐 & 8D86趐、58AB壿 & 58FF壿

Three pairs will be treated as unrelated singletons
-- 571F土 & 58EB士、9E42鶹 & 9E43鶻、53E3口 & 56D7 囗

2）Variant mapping relationship in dotAsia changed as proposed by IP

媼(5B0E)	媼(5B0E) >> 媼(5B14)	媼(5B0E) >> 媼(5B14)	媼(5B0E)媼(5B14)
媼(5B14)	媼(5B14)	媼(5B14)	媼(5B0E)媼(5B14)

Out-of-Repertoire Chars and Variants (TBD)

- In 2017, CGP reviewed 94 out of 144 J-only Kanji characters, because they are either included in IIcore or in GSCCT (General Standard Chinese Characters Table).
- (1) 39 characters have been set as singletons
  Disposition: to be excluded from CLGR
- (2) 55 characters have been set as variants to the existing CGP characters
  Disposition: to add the out-of-repertoire variants into CLGR
- For the rest 50 J-only Kanji characters not included in IIcore or GSCCT, CGP need further discussion whether to conduct the second round of out-of-repertoire variant review.

Unicode	Glyph	Srouce
3960	惰	K
4137	秬	K
4A12	霧	K
4E55	乕	J
4E8A	事	J
4EED	仭	J
5271	劔	J
5338	匸	J
5368	卼	K
56A2	囊	J
576E	垚	K
5841	塁	J

Update by Japanese Generation Panel (GP)

Yoneya Yoshiro
Member, Japanese GP

Evaluation of Visually Similar Characters for JGP – Preliminary study –

November 6, 2019 @ IDN Program Updates

Tatsuya Mori[1] and Yoshiro Yoneya[2]

1 Waseda University, 2 JPRS

Background and Our goal

- Background
  - Security threat caused by the IDN homograph attack has become vital
  - Homoglyphs exist even in a same script (in-script homoglyph)
  - RZ LGR Integration Panel requested several generation panels to investigate the in-script homoglyphs
- Our goal
  - Develop a generic scheme to evaluate in-script homoglyphs
  - Study whether Japanese in-script homographs (confusable characters) defined in the Unicode "confusables" are actually confusable to the users
  - Contribute to finalize JGP's proposal by sharing the results

In-script homoglyphs (Japanese)

Hiragana vs. Katakana – 3 pairs

- へ　　へ
- べ　　ベ
- ぺ　　ペ

Katakana vs. Chinese letter (Han) – 7 pairs

- ニ　　二
- ハ　　八
- カ　　力
- ト　　卜
- ロ　　口
- タ　　夕
- エ　　工

Questions to be addressed

We want to understand how the following factors affect the human perception to the in-script homoglyphs.

- Does the font family matter?
- Does a word matter?
- Does the size of letters matter?
- Does the type of user devices matter?
- Does the linguistic background/knowledge matter?

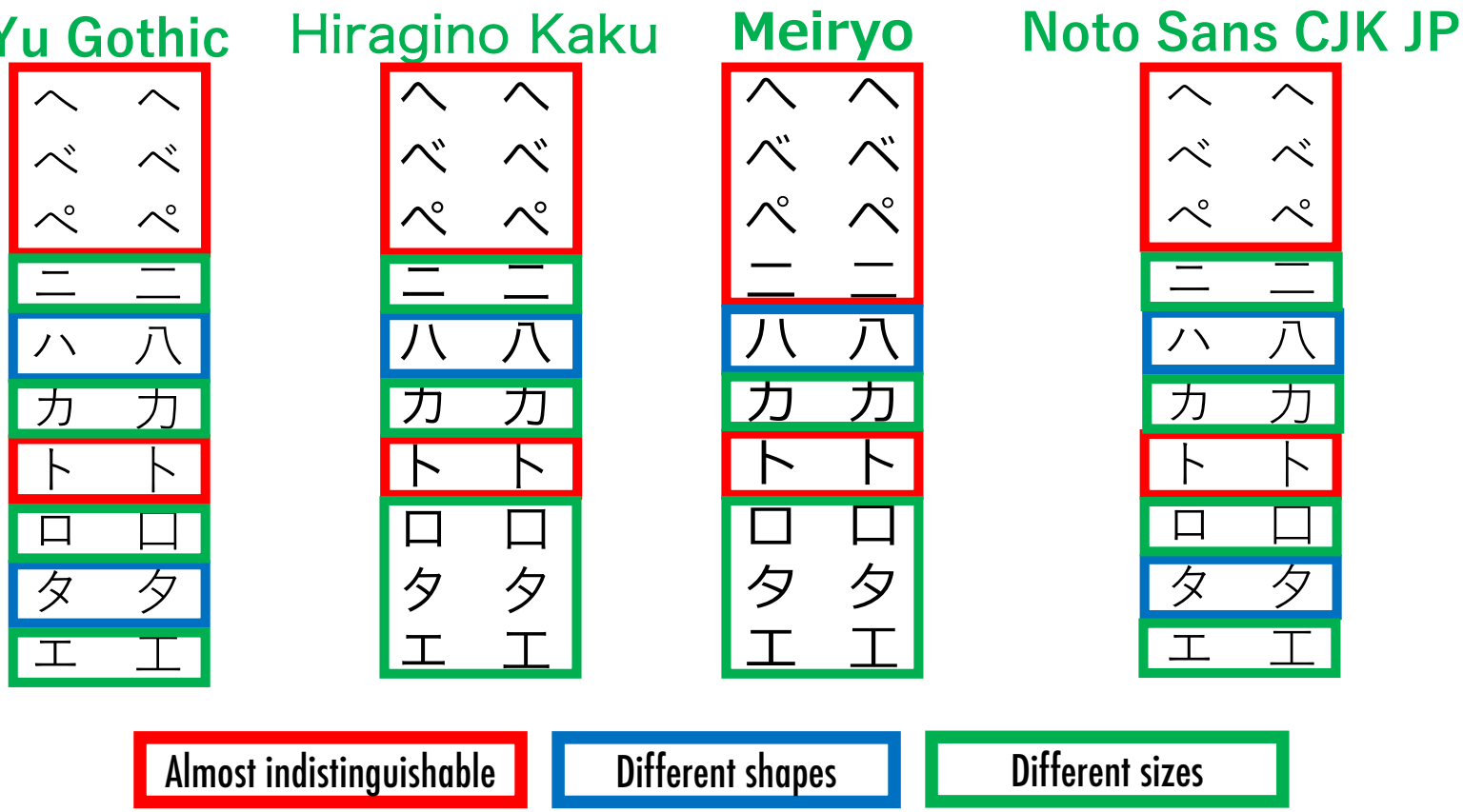
Questions to be addressed

We want to understand how the following factors affect the human perception to the in-script homoglyphs.

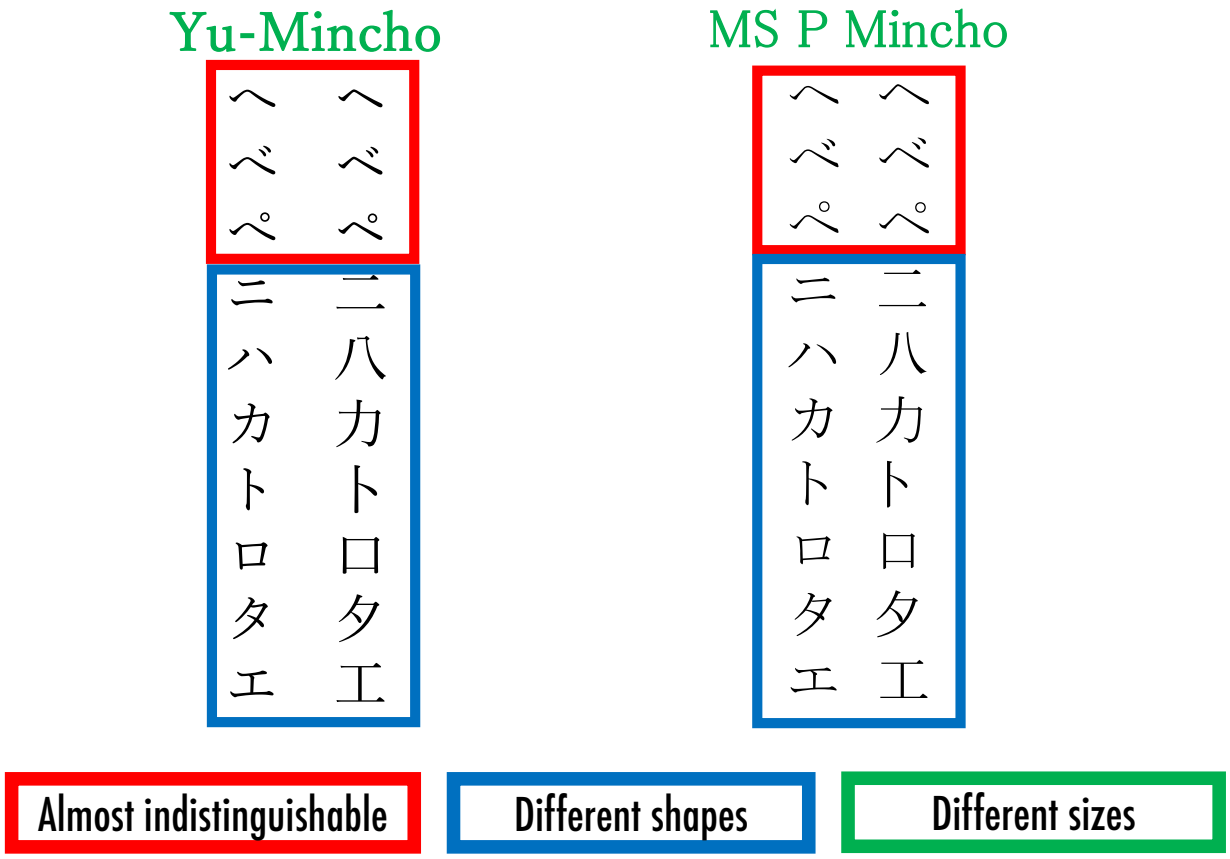
- Does the font family matter?
- Does a word matter?
- Does the size of letters matter?
- Does the type of user devices matter?
- Does the linguistic background/knowledge matter?

← Will report the first item today (preliminary study)

In-script homoglyphs with various fonts (Gothic fonts; i.e., Sans-serif)



In-script homoglyphs with various fonts (Mincho fonts; i.e., Serif)



Example words with the in-script homographs

- **Proper nouns**
  - 二馬力 (homograph) → 二馬力 (original)
- **General nouns (type of university)**
  - 工業大学 (homograph) → 工業大学 (original)
- **General nouns (type of vegetables)**
  - トマト (homograph) → トマト (original)
  - ヘチマ (homograph) → ヘチマ (original)

Experiment

- **Participants:**
  - lab students (undergrads/graduates)
  - Native Japanese speakers and few native Chinese speakers.
  - They have knowledge about homoglyph attacks (possible source of the bias)
    - “skeptical tech-savvy”
- **Task**
  - From the following 8 words, pick up words that you notice somewhat unnatural with respect to the composing characters.

1. 宮崎駿
2. 二馬力
3. 工業大学
4. 国際大学
5. キュウリ
6. トマト
7. ピーマン
8. ヘチマ

Data

Task#1

Meiryo

1. 宮崎駿
2. 二馬力
3. 工業大学
4. 国際大学
5. キュウリ
6. トマト
7. ピーマン
8. ヘチマ

Task#2

Yu-Mincho

1. 宮崎駿
2. 二馬力
3. 工業大学
4. 国際大学
5. キュウリ
6. トマト
7. ピーマン
8. ヘチマ

In-script homographs

Results

Task#1	#voters / 16
1. 宮崎駿	1
2. 二馬力	13
3. 工業大学	10
4. 国際大学	1
5. キュウリ	1
6. トマト	5
7. ピーマン	2
8. ヘチマ	5

Task#2	#voters / 16
1. 宮崎駿	0
2. ニ馬力	15
3. エ業大学	16
4. 国際大学	0
5. キュウリ	0
6. トマト	10
7. ピーマン	1
8. ヘチマ	2

#voters = number of participants who thought that the word was somewhat unnatural.
(notice the suspiciousness)

Key Findings

- **Font types affected the human perception**
  - Generally, Gothic fonts are more confusable than Mincho fonts.
  - Note: Majority of web browsers use the Gothic fonts as the default font (address bar etc.)

- **Human perception varied among the in-script homoglyphs**
  - The most confusable homoglyphs: {へ,ヘ} {ト , ト}
    - For the Gothic font, only 1/3 of participants noticed that these pairs were unnatural
  - The most distinguishable homoglyphs: {エ,工} {ニ,二}
    - Majority of participants noticed that these pairs were unnatural

Remaining tasks

- Work on the remaining questions:
  - Does a word matter?
  - Does the size of letters matter?
  - Does the type of user devices matter?
  - Does the linguistic background/knowledge matter?
- Change the set of words presented to the participants.
- Increase the number/diversity of participants.

- Summarize the results as a final report.

Update by Korean Generation Panel (GP)

Kim Kyongsok
Chair, Korean GP

Agenda

1. Overview of the Script and Language
2. Generation Panel Membership
3. Progress Summary
4. Current Work
5. Plan and Next Steps

Overview of the Script and Language

- K-LGR covers Korean script (= Hangul + Hanja).
- “Korean script” usually means “Hangeul” or “Hangul”. However, in the context of the Korean LGR (K-LGR), Korean script is a union of Hangul (한글) and Hanja (한자).
- Korean language has a long history, more than 2000 years.
- Hangul: invented in 1443.
- Hanja was used before Hangul was invented. Hanja is still used in Rep. of Korea.
- Korean language is mainly used in Republic of Korea (South Korea) and Democratic People’s Republic of Korea (North Korea).
  - Also used by Korean people living in China, USA, Japan, Europe, Brazil, Russia, Vietnam, and so on.

Korean Generation Panel (KGP) Membership

- Technical Experts: Kyongsok KIM (Chair), Dongman LEE
- Linguists: Jeongdo CHOI (Hangul), Sanghyun SHIN (Hanja), Sungduk CHO (Hanja)
- Policy Makers: Youngeum LEE, Youn Jung PARK
- Community: Eungjun JEON, Boknam YUN, Byeongil OH
- Registry: Jungmin LEE, Boyoung KIM
- Registration Agency: Seong-jin PARK, ChangKi JANG, Myungsoo LEE

Progress Summary (1): K-LGR v1.0 (2017.12.10.)

- K-LGR v1.0 (2017.12.10.): repertoire and variant groups
  - Hangul: repertoire – 11172 syllables, no variant groups
  - Hanja: repertoire – 4758 characters, 152 variant groups
  - Variant groups composed of Hangul syllables and Hanja chars: 5 (3 Hanja chars: out-of-repertoire variant)

- 4758 Hanja chars in K-LGR v1.0

Source of Hanja Character Set	# chars
1) KS X 1001 (268 comptb. chars excluded)	4620
2) IICORE - K column marked	4744
K-LGR v1.0 (2017.12.10.): Hanja List (Union of 1) and 2))	4758

Progress Summary (2): Public Comment

(January – March 2018)

- A summary of public comments made:
  - Including Hanja in K-LGR repertoire: positive.
  - Allowing Hangul-Hanja mixed label: several negative comments, some positive comments.
  - Hangul-Hanja variant group: CJK agreement needed.
  - Specific details need be corrected or modified.

- Examples of issues raised by Mr. Kyyuhong BYUN:
  - References; quotes; etc.
  - Many Hanja chars allowed for personal names not included in K-LGR.
  - Hangul Jamo not included in K-LGR (actually not in MSR-3).
  - More Hangul-Hanja variant groups need be included.

Progress Summary (3): Public Comment

Reviewed (April 2018 – August 2019)

- Requests by Mr. BYUN for specific details.
  - Reviewed and discussed.
  - Mostly accepted in principle and reflected in the tentative version of K-LGR.

- Hangul-only labels, Hanja-only labels, Hangul-Hanja mixed labels.
  - KGP reconfirmed that there was a general consensus to allow Hangul-only labels and Hanja-only labels;
  - However, KGP has not reached a conclusion as to whether to allow Hangul-Hanja mixed labels.

Current Work

- Waiting for the conclusion as to whether to include cross-script (visually identical or similar) variant groups.
  - Variant groups of Hangul syllables and Hanja characters.
  - Variant groups of Kana and Kanji characters.

- Unification of Hanja variant groups between Korea and China.
  - Whether to keep the variant groups unified between Korea and China.

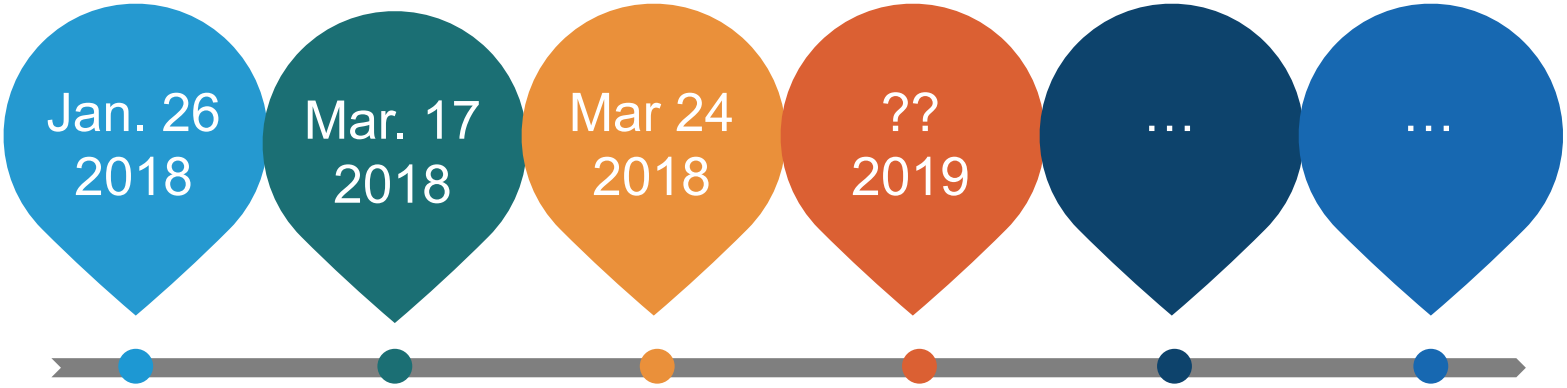
- Hangul-Hanja mixed labels
  - Will decide on a final conclusion.

- Revision of K-LGR 1.0
  - After the above issues are resolved, K-LGR will be revised and published.

Brief History of KGP Activities

- Dec. 2013: Korean GP (KGP) organized.
- May. 2015: K-LGR v0.1
- Feb. 2016: The Korean community “formally” forms Generation Panel for Developing the Root Zone Label Generation Rules (LGR).
- Dec. 2017: K-LGR v1.0
- January – March 2018: Public Comment for K-LGR v1.0
- March 2018 – August 2019: Public Comment for K-LGR v1.0 reviewed for possible reflection in the next version of K-LGR.
- 38 KGP meetings
- Several CJK coordination meetings during ICANN Public Meetings (49-64).
- Several CJK coordination meetings in Rep. of Korea, China and Taiwan.

Plan and Next Steps



Public Comment: Open Date	Public Comment: Close Date	Summary Report Due Date	Rev. K-LGR Proposal with any public comments reflected.	IP evaluates the final K-LGR Proposal.	K-LGR will be integrated into a subsequent version of the RZ-LGR

Update on Neo-Brahmi GP

Dr. Ajay Data
Co-chair, NBGP

Agenda

1. Introducing NBGP
2. NBGP Members
3. Efforts and Progress Timeline
4. Summary

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

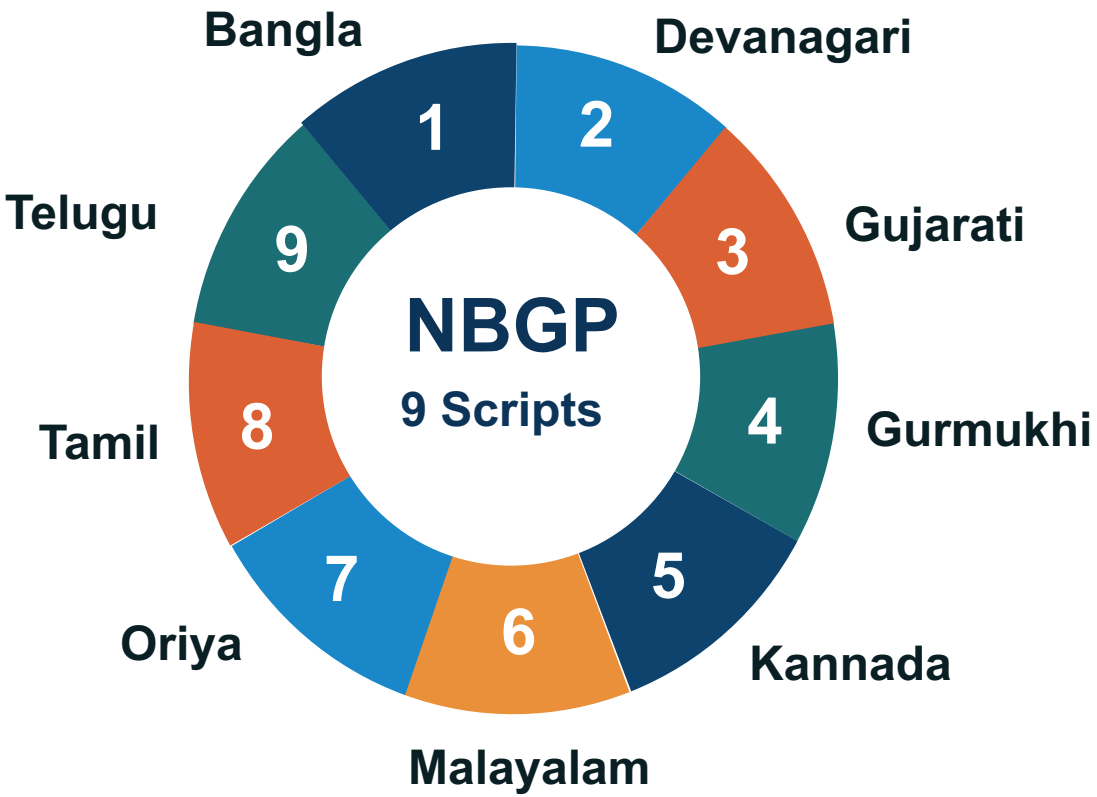
Geo Coverage

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

NBGP Seated

F2F meeting in Kathmandu

Devanagari LGR was getting prepared

F2F meeting in Colombo

Draft of **Bangla**, **Gujarati, Gurmukhi, Kannada, Malayalam, Oriya, Tamil, Telugu** LGR were gettering prepared

F2F meeting in Dhaka, Draft of **Bangla** LGR was gettering prepared

Devanagari, Gurumukhi and Gujarati Public Comment

8 out of 9 scripts delegated into root zone

Bangla is being finalised.

Summary of Achievements

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

Update by Myanmar Generation Panel

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

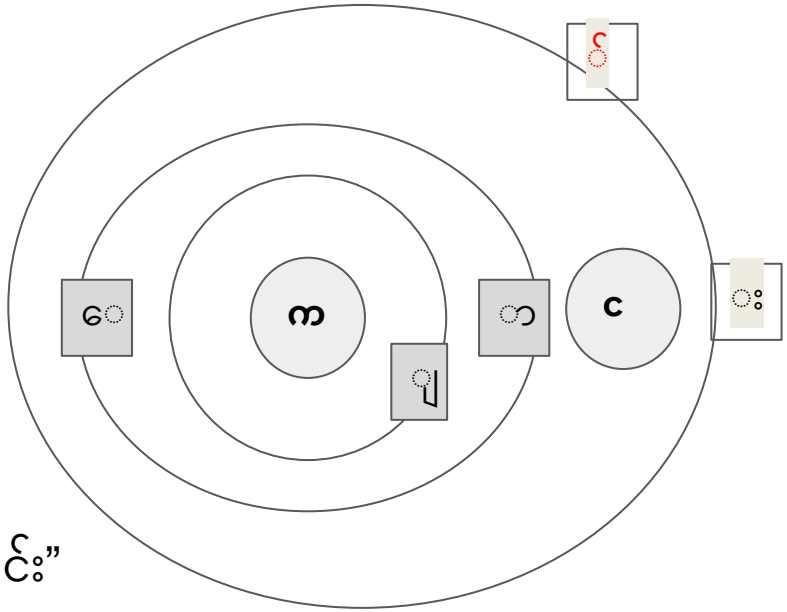
- Myanmar script have been evolved from the Brahmi script which has flourished in the Indian subcontinent between 5th Century B.C and 3rd Century A.D.

- 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 /mja-ma/	1	42,906,490
Shan	[shn]	Myanmar, China, Thailand	လိၵ်းတဵး /lik˦˥.taj˧/	3	3,295,000
Rakhine	[rki]	Myanmar	ရခိုင် /rə.khi/	3	2,020,000
Karen, Sgaw	[ksw]	Myanmar, Thailand	စ်ၤ /sɣɔʔ/	3	1,560,000
Mon	[mnw]	Myanmar, Thailand	မန် /mun/	5	851,000
Pa'O Karen	[blk]	Myanmar	ပအိုဝ်း /pə.ʔəʊ/	5	560,740

Nature of Myanmar Script [1/2]

- Myanmar Script is written from left to right and requires no spaces between words
- 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.



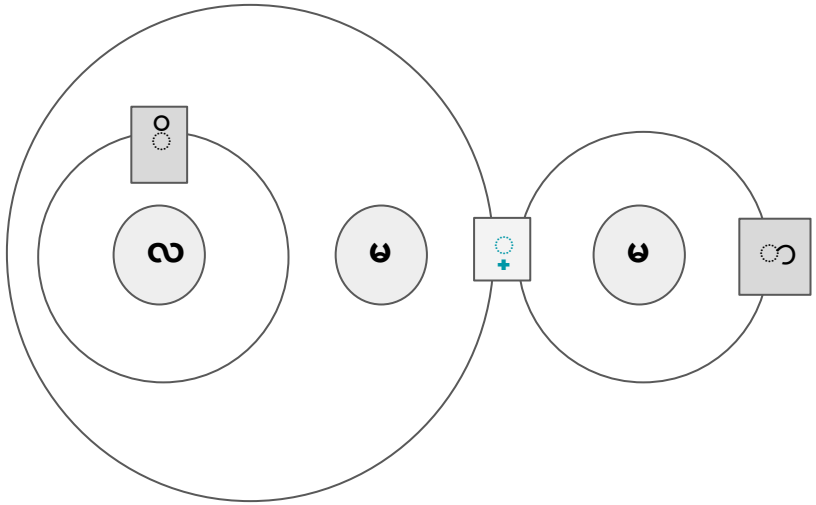
- Example word : “ေကျာင်း”
- Meaning : “School”

/ **က** / ျ / ေ / ာ / **c** / ် / း /

/ **U+1000** / U+103B / U+1031 / U+102C / **U+1004** / U+103A / U+1038 /

Nature of Myanmar Script [2/2]

- Myanmar Script composes with more than one phonetic syllables
  - Forming Orthographic syllable using Virama (invisible devowelizer)



- Example word : “လိမ္မာ”
- Meaning : “clever”

/ လ / ိ / မ / ◌် / မ / ာ /

/ **U+101C** / U+102D / **U+1019** / **U+1039** / **U+1019** / U+102C /

Cross-Script Variant Analysis

- Myanmar Characters vs Malayalam/ Oriya/ Georgian Characters

No.	Glyph	Code Point	Glyph	Code Point	Cross-Script Character Name
1	ဂ	U+1002	ഗ / ဂ	U+0D31 / U+10D8	MALAYALAM LETTER RRA/ GEORGIAN LETTER IN
2	ေ◌	U+1031	େ◌	U+0B47	ORIYA VOWEL SIGN E
3	ဏ	U+1010	ဏ	U+10D7	GEORGIAN LETTER TAN

- Myanmar Letter Wa [U+101D] vs Other

No.	Glyph	Code Point	Character Name
1	ഠ	U+0D20	MALAYALAM LETTER TTHA
2	ଠ	U+0B20	ORIYA LETTER TTHA
3	o	U+006F	LATIN SMALL LETTER O
4	ο	U+03BF	GREEK SMALL LETTER OMICRON
5	о	U+043E	CYRILLIC SMALL LETTER O
6	օ	U+0585	ARMENIAN SMALL LETTER OH

In-Script Variant Analysis [Examples 1/2]

- 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	Disposition
1	U+1023 MYANMAR LETTER I		U+1000 U+1039 U+1000 SV1		Block
2	U+101F U+103A Myanmar Letter Ha Asat		U+1015 U+102C U+103A (U+1015 ST1)		Block
3	U+1061 Myanmar Letter Sgaw Karen SHA		U+101B U+103E		Block

Example Rules

5. (C+K) or (C2 + S16) or (C3+S17) must follow C or M or DV or OV

15. U+102C Cannot follow three consonants U+1002, U+1015, U+101D
Extension of Rule 15: sequence U+1015 U+1039 U+1015 U+102C "  " is defined

In-Script Variant Analysis [Examples 2/2]

- 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	Disposition
18	U+102E Myanmar Vowel Sign II	◌ဲ	U+1033 Myanmar Vowel Sign Mon II	◌ဳ	Allocatable
19	U+102B Myanmar Vowel Sign TALL AA	◌ါ	U+1083 Myanmar Vowel Sign Shan AA	◌ၢ	Allocatable
20	U+102C Myanmar Vowel Sign AA	◌ာ	U+1083 Myanmar Vowel Sign Shan AA	◌ၢ	Allocatable
21	U+1004 U+103A Myanmar Letter NGA Asat	င်	U+1004 U+103A U+1039 Myanmar Sign Kinsi	◌္	Allocatable
22	U+105A U+103A Myanmar Letter Mon NGA Asat	ၚ်	U+105A U+103A U+1039 Mon Sign Kinsi	◌္	Allocatable

Myanmar Vowel Sign II "◌ဲ" : Mon Vowel Sign II "◌ဳ"

Burmese "ဝံ" : Mon "ဝံ"

/U+101D **U+102E** U+1000 **U+102E** / : / U+101D **U+1033** U+1000 **U+1033** /

Whole Label Evaluation Rules [1/2]

- **Classifications (Added classes)**

C1	→ 103F သ သ (GREAT SA, theoretical combination of two Myanmar Letter Sa)
C2	→ င , ဉ , ည , ဏ , န , မ , ယ (1004, 1009, 100A, 100F, 1014, 1019, 101A)
C3	→ င , ဉ , ည , ဏ , န , မ , ဝ , ဟ, ဠ, ဆ (1004, 1009, 100A, 100F, 1014, 1019, 101D, 1075, 107A, 107C)
LV	→ Long Vowel: 102B, 102C, 102E, 1030, 1031, 1032, 1036
SV	→ Short Vowel: 102D, 102F
Sh_Tone	→ 1087, 1088, 1089, 108A
Sh_Vowel	→ 102D, 102E, 102F, 1030, 1031, 1083, 1084, 1086
Pao_Tone	→ 108F, AA7B
Sgaw_Tone	→ 1064
OV	→ Other Various Sign (Section 3.3.3) (1035, 1062, 1085)

Whole Label Evaluation Rules [2/2]

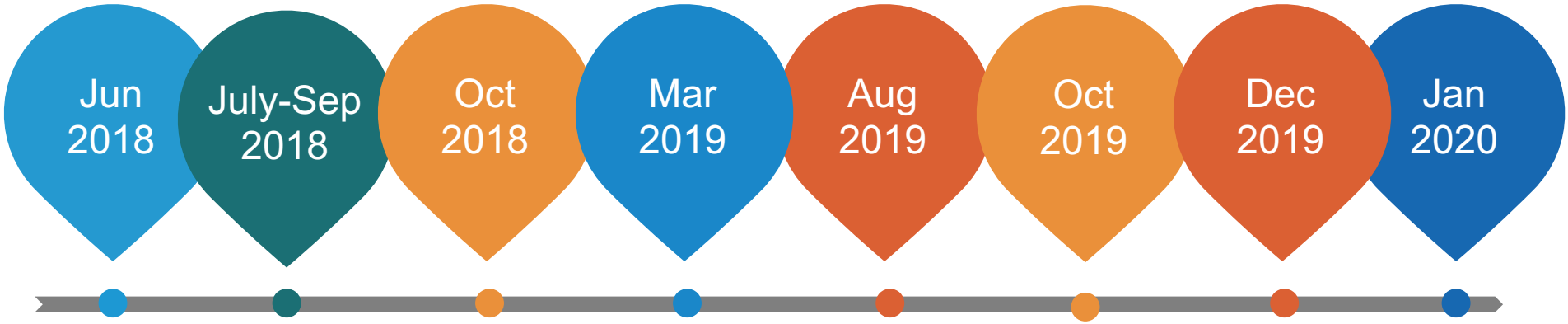
Whole Label Evaluation Rules (Added rules)

6. S11 must follow C or M or DV and another C must follow S11
7. VIRAMA must be in between two C (C+VIRAMA+C)

- C cannot be in between Vs to prevent v-c-v-c

8. T_LONG must follow C or M or LV or S12
9. T_SHORT must follow C or LV or S12
10. Sh_Tone must follow Sh_Vowel or (C+K) or S_Sh2 or S_Sh5
11. Pa'O Rules
    - Pao_Tone must follow DV or K
    - S_Pao must follow a Pa'O Consonant or Pa'O Medial
12. ST4 must follow U+1031, U+1032, S12, S14, S15
13. ST1, ST2, ST3 or Sgaw_Tone must follow C or M or DV
14. S_Mon4 must follow C or M or DV
    and another C must follow S_Mon4
15. U+102C Cannot follow three consonants U+1002, U+1015, U+101D

Timeline and Next Step



Jun 2018	July-Sep 2018	Oct 2018	Mar 2019	Aug 2019	Oct 2019	Dec 2019	Jan 2020
Form the Myanmar GP	Develop the code points repertoire, variant rules and WLE rules	Submit the first draft LGR proposal to the IP	Submit the Second draft LGR proposal to the IP	Public workshop in Yangon	Submit the Third draft LGR proposal to the IP	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