



# IDN Program Update

22 February 2022

# Agenda

---

- ⦿ IDN Implementation Guidelines
- ⦿ Additional Scripts to Support in IDNs
- ⦿ RZ-LGR Status Update
  - Background
  - Myanmar Generation Panel (GP) Update
  - RZ-LGR-5 Update
- ⦿ Q&A

# IDN Implementation Guidelines

Pitinan Kooarmornpatana  
IDN Programs, Senior Manager

# Internationalized Domain Name (IDN) Implementation Guidelines

## Background

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

- Generic top-level domain (gTLD) registries and registrars offering IDNs are contractually bound.
  - New gTLD Base Registry Agreement, Specification 6 Section 1.4: *“Registry Operator shall comply with the ICANN IDN Guidelines”*
  - 2013 Registrar Accreditation Agreement, Additional Registrar Operation Specification Clause 3: *“Registrar shall also comply with the IDN Guidelines”*
- IDN country code top-level domains (ccTLDs) are expected to comply.
  - IDN ccTLD Fast Track Process: *“IDN domain names are to be registered in accordance with a publicly available registration policy that shall comply on an ongoing basis with relevant applicable standards to IDNs, such as the IDNA Protocol, and with the IDN guidelines”*

# Guidelines Update

---

Why are guidelines being updated now?

- ⦿ Since version 3 published in 2011, there has been additional analysis, data, and operational experience with IDNs.
  - Terminology for IDNs
  - Representation of the IDN tables (RFC 7940)
  - Design of the IDN tables (RFC 6912, RFC 8228)
  - Reference Second Level Label Generation Rules
  - Maximal Starting Repertoire (MSR) and Root-Zone Label Generation Rules (RZ-LGR)
  - User Experience Study and SSAC recommendations in SAC60 report
  - Proposals for Root Zone Label Generation Ruleset

# Topics Covered in Version 4

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



- Guidelines for the Implementation of Internationalized Domain Names Version 4.0 was published 10 May 2018.

# IDN Implementation Guidelines Timeline

---

- On 30 April 2019, the Generic Names Supporting Organization (GNSO) made a [request](#) to the ICANN Board to allow it to study the guidelines before implementation.
- On 20 May 2021, the GNSO Council initiated an Expedited Policy Development Process ([EPDP](#)) on IDNs to address additional issues related to IDNs not discussed in the SubPro report, including how to securely and stably define and manage variant gTLDs and how to update IDN guidelines in the future.
- On 18 August 2021, the GNSO [requested](#) that the adoption of the IDN Implementation Guidelines be delayed until the IDN EPDP is concluded due to some overlap.
- The ICANN Board [responded](#) on 20 October 2021 requesting the GNSO to analyze and identify the guidelines that overlap with the topics included in the IDN EPDP.
- On 21 January 2022, the GNSO [responded](#) supporting the Board's suggestion and identified the following guidelines that overlap with topics included in the IDN EPDP charter: 6a, 11, 12, 13, and 18.
- The ICANN Board is considering the GNSO response.

# Additional Scripts to Support in IDNs

Pitinan Kooarmornpatana  
IDN Programs, Senior Manager



# Background

---

- ⦿ [ICANN's mission](#), as part of its Bylaws, mandates:
  - ICANN's authority or ability to adopt or implement policies or procedures that take into account the use of domain names as natural-language identifiers.
- ⦿ Domain names are identifiers defined by the Internationalized Domain Names standard (IDNA2008), which is based on the Unicode standard.
- ⦿ IDNA2008 standard:
  - Uses an algorithmic mechanism for code points using their properties to determine inclusion.
  - Asks to put additional constraints to address to create secure and stable domain name labels.
  - Does not conduct a script-based review for inclusion.

# Background

---

- ICANN has been working with the relevant script communities to develop secure and stable solutions to support the various scripts in IDNs.
- Using the RZ-LGR procedure, so far ICANN has focused on supporting 28 scripts by developing Root Zone Label Generation Rules (RZ-LGR).
  - The 28 scripts are shortlisted in the [Maximal Starting Repertoire \(MSR\)](#), having common and widespread use by the script communities.
  - The scripts include: Arabic, Armenian, Bengali, Cyrillic, Devanagari, Ethiopic, Georgian, Greek, Gujarati, Gurmukhi, Han, Hangul, Hebrew, Hiragana, Kannada, Katakana, Khmer, Lao, Latin, Malayalam, Myanmar, Oriya, Sinhala, Tamil, Telugu, Thaana, Thai, and Tibetan.
- Same 28 scripts are used to develop [reference LGRs for the second level](#), also used to evaluate the IDN tables for gTLDs.

# Question to Address

---

- ⦿ Unicode 14.0 has [159 scripts](#), including contemporary and historic scripts.
  - Unicode states that not all 159 scripts may be suitable for identifiers, like domain names.
- ⦿ Which additional scripts from these 159 could ICANN org support in domain names?
  - At the top level?
  - At the second and other levels?

# Script Categories by Unicode

---

- ⦿ Unicode standard discusses the use of these 159 scripts in identifiers in [Unicode Standard Annex #31: Unicode Identifier and Pattern Syntax](#).
- ⦿ Unicode categorizes 159 scripts into three sets:
  - Excluded scripts
  - Limited use scripts
  - Recommended scripts
- ⦿ **Excluded scripts (94 scripts)**
  - Scripts not in customary modern use, and thus may want to exclude from identifiers (include historic and obsolete scripts, scripts used mostly liturgically, and regional scripts used only in very small communities or with very limited current usage).
  - Some scripts also have unresolved architectural issues that make them currently unsuitable for identifiers.
  - Examples: Brahmi, Coptic, Egyptian Hieroglyphs, Gothic, Mongolian, Nabataean, Ugaritic, etc.

# Script Categories by Unicode

---

- ⦿ **Excluded scripts (34 scripts)**

- Modern scripts that are in more limited use. To avoid security issues, some implementations may wish to disallow the limited-use scripts in identifiers.
- Examples: Balinese, Canadian Aboriginal Syllabics, Cherokee, Javanese, Limbu, Nko, Syriac, Tifinagh, etc.

- ⦿ **Recommended scripts (29 scripts + common and inherited scripts)**

- Scripts with widespread modern customary use, or regional scripts in modern customary use by large communities.
- Include: Arabic, Armenian, Bengali, Cyrillic, Devanagari, Ethiopic, Georgian, Greek, Gujarati, Gurmukhi, Han, Hangeul, Hebrew, Hiragana, Kannada, Katakana, Khmer, Lao, Latin, Malayalam, Myanmar, Oriya, Sinhala, Tamil, Telugu, Thaana, Thai, and Tibetan.

# Evaluating Unicode Scripts for Use in IDNs

- ICANN engaged Unicode and IDN experts to review the UAX#31 in the context of IDNs.
- The experts analyzed the different scripts and script categories for their use in IDNs.
- They have finalized their analysis and recommendations in the report: [Evaluating Unicode Scripts for Use in IDNs](#).

Script Category in UAX#31	Root Zone	Second Level
Recommended scripts	✓	✓
Limited use scripts	X	✓
Excluded scripts	X	X

✓ = may qualify but requires analysis on case-to-case basis.

× = does not qualify.

# Evaluating Unicode Scripts for Use in IDNs

---

- ⦿ Why limited use scripts on case-to-case basis? (see the report for more detailed discussion)
  - A script can potentially present significant security or stability issues.
  - The script community is not available to consult to develop reference Label Generation Rules for the script.
  - There is very limited information available online about how the script is used.
  - ...
- ⦿ Without sufficient information on the use of the script and having active community support, it is challenging to develop reference Label Generation Rules for the script and therefore to support the script in IDNs safely.

# Examples of Risks with Limited Used Scripts

U-Label	A-Label	Script	Code Points
APPLE	xn--g9de3mpaa	<a href="#">Cherokee</a>	U+13AA U+13E2 U+13E2 U+13DE U+13AC
SAP	xn--g9d9gza	<a href="#">Cherokee</a>	U+13DA U+13AA U+13E2
ceo	xn--qh9azay	<a href="#">Kayah Li</a>	U+A90D U+A915 U+A900
FACEBOOK	xn-- 3l8aoahl9fnwa	<a href="#">Lisu</a>	U+A4DD U+A4EE U+A4DA U+A4F0 U+A4D0 U+A4F3 U+A4F3 U+A4D7
VERISIGN	xn-- 9l8auaedo4dpb	<a href="#">Lisu</a>	U+A4E6 U+A4F0 U+A4E3 U+A4F2 U+A4E2 U+A4F2 U+A4D6 U+A4E0
NETFLIX	xn-- 7l8asahg4c0ar	<a href="#">Lisu</a>	U+A4E0 U+A4F0 U+A4D4 U+A4DD U+A4E1 U+A4F2 U+A4EB
ICANN	xn--em8amaa8cs	<a href="#">Lisu</a>	U+A4F2 U+A4DA U+A4EE U+A4E0 U+A4E0
new	xn--tgfg7f	<a href="#">Tai Le</a>	U+1952 U+1971 U+1955
BREAD	xn--2ne8jzc0dn	<a href="#">Canadian Syllabics</a>	U+15F7 U+1587 U+15F4 U+15C5 U+15DE
BIT	xn--sn8aqfuc	<a href="#">Vai</a>	U+A557 U+A56F U+A50B



# Examples of Risks with Limited Used Scripts

\*Recommended Scripts

U-Label	A-Label	Script	Code Points
၀၀၀၀	xn--9hf1bqr	<a href="#">New Tai Lue</a>	U+199E U+1985 U+19A2 U+1998
၀၀၀၀	xn--pid2bj1n	<a href="#">Myanmar</a> *	U+1040 U+1002 U+1015 U+1018

U-Label	A-Label	Script	Code Points
၀၀၀	xn--zifd6c	<a href="#">New Tai Lue</a>	U+19A0 U+199E U+19B1
၀၀၀	xn--dwcs4g	<a href="#">Malayalam</a> *	U+0D17 U+0D20 U+0D3E

U-Label	A-Label	Script	Code Points
ডঢব	xn--15bc4a	<a href="#">Bangla (Bengali)</a> *	U+09A1 U+09A2 U+09AC
ডঢব	xn--r98ajac	<a href="#">Syloti Nagri</a>	U+A812 U+A811 U+A80C

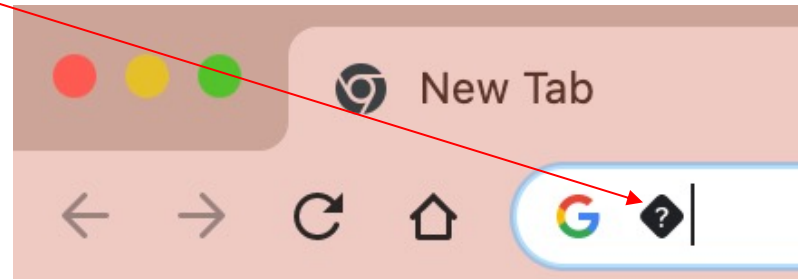
# Examples of Risks with Limited Used Scripts

## ⦿ Unstable Rendering Issues

### Newa script

	1140	1141	1142	1143	1144	1145	1146	1147
0	ञ्	ग	थ	व	ौ	०	ख	
	11400	11410	11420	11430	11440	11450	11460	
1	जा	घ	द	ष	ौ	१	ॐ	
	11401	11411	11421	11431	11441	11451	11461	
2	ॐ	ढ	ध	ष	्	२		
	11402	11412	11422	11432	11442	11452		
3	ॐ	झ	न	स	ं	३		
	11403	11413	11423	11433	11443	11453		
4	उ	व	र	ह	ं	४		
	11404	11414	11424	11434	11444	11454		
5	ऊ	ळ	प	ा	०	५		
	11405	11415	11425	11435	11445	11455		

Based on IDNA 2008, this code point is Protocol Valid (PVALID) as published on IANA [website](#).



U+11420 glyph does not render in the browser address bar (Chrome on MAC OS)

- ⦿ Some limited use scripts with rendering issues: Adlam, Chakma, Nyiakeng Puachue Hmong, Miao, Newa, Osage, Hanifi Rohingya, Tifinagh, and Wancho.

# Public Comment Summary

---

## Additional Unicode Scripts for Support in Internationalized Domain Names

- Public Comment: <https://www.icann.org/en/public-comment/proceeding/additional-unicode-scripts-for-support-in-internationalized-domain-names-05-01-2022>
  - Opened: 5 January 2022
  - Closed: 15 February 2022
  - Report Due: 8 March 2022

# Root Zone Label Generation Panel (RZ-LGR) Update

Pitinan Kooarmornpatana  
IDN Programs, Senior Manager

# A Brief History of RZ-LGR

---

- The ICANN community identified the need for variant top-level domains.
- The [Integrated Issues Report](#) identified the need to define variant TLDs as a prerequisite.
- The community identified RZ-LGR as the mechanism to define variant TLDs and specified the [LGR Procedure](#) to develop RZ-LGR which [approved](#) by the ICANN Board in 2013.
- In 2019, the ICANN Board [resolved](#) that the GNSO and ccNSO take into account the Recommendations for Managing the IDN Variant TLDs - which integrated the use of RZ-LGR - in their policy development processes.
- In 2020, the ICANN Board [resolved](#) that the GNSO and ccNSO take into account the Recommendations for the Technical Utilization of the Root Zone Label Generation in their policy development processes.
- In 2021, the GNSO published its [Report on New gTLD Subsequent Procedures](#) which incorporates the use of RZ-LGR for the next round.

# Summary of Generation Panel (GP) Work

Script	Start	End	Days	2014	2015	2016	2017	2018	2019	2020	2021	2022
Arabic	14-Feb-14	18-Nov-15	642									
Armenian	3-Feb-15	5-Nov-15	275									
Bangla	26-May-15	20-May-20	1821									
Chinese	24-Sep-14	26-May-20	2071									
Cyrillic	10-Dec-15	3-Apr-18	845									
Devanagari	26-May-15	22-Apr-19	1427									
Ethiopic	22-Dec-15	17-May-17	512									
Georgian	17-Jun-16	24-Nov-16	160									
Greek	31-Oct-16	15-Jul-21	1718									
Gujarati	26-May-15	6-Mar-19	1380									
Gurmukhi	26-May-15	22-Apr-19	1427									
Hebrew	15-Oct-18	24-Apr-19	191									
Japanese	17-Mar-15	30-Sep-21	2389									
Kannada	26-May-15	6-Mar-19	1380									
Khmer	17-Jun-15	15-Aug-16	425									
Korean	1-Feb-16	1-May-21	1916									
Lao	15-Sep-15	31-Jan-17	504									
Latin	15-May-17	23-Sep-21	1592									
Malayalam	26-May-15	26-Jun-20	1858									
Myanmar	28-Jun-18	ongoing	-									
Oriya	26-May-15	6-Mar-19	1380									
Sinhala	3-Jan-18	22-Apr-19	474									
Tamil	26-May-15	6-Mar-19	1380									
Telugu	26-May-15	7-Jun-19	1473									
Thaana	TBD											
Thai	6-Oct-15	25-May-17	597									
Tibetan	TBD											

# Myanmar Generation Panel Update

Yin May Oo  
Generation Panel Co-Chair

# Introduction



The word “Hello” written in listed languages:

Burmese: “မင်္ဂလာပါ” /mɪŋ-gə-la-ba/

Shan: “မိၤသုၣ်ခါး” /mi-sun-kha/

Rakhine: “သာလီၣ်စွပါ” /θa-li-zwə-ba/

Sgaw Karen: “နီၤလၢအဂ့ၤ” /ni:-ləə-ə-ye:/

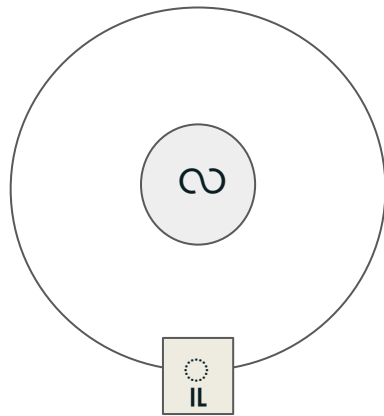
Mon: “မိၣ်ရအံ” /mɲeə-rə-aʊ/

Pa’O Karen: “မင်္ဂလာဒုးဒြး” /mɪŋ-gə-lə-dra-ɔː ]



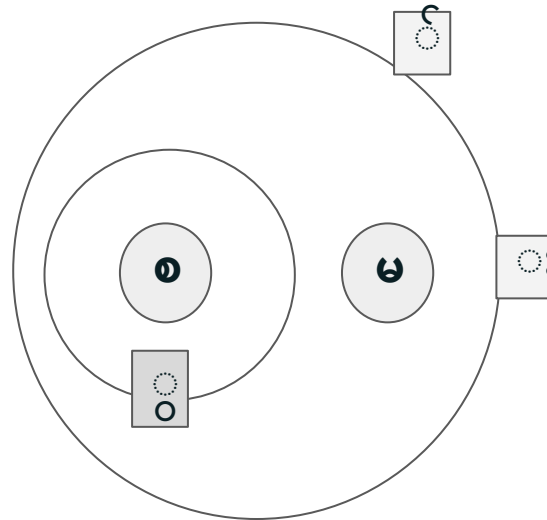
# Introduction

Example label - လူ့စွမ်း (meaning, human ability)



/ လူ / ̇ /

/ U+101C / U+1030 /



/ လူ့ / ̇ / စွမ်း / ̇ /

/ U+1005 / U+103D / U+1019 / U+103A / U+1038 /

# Repertoire

	100	101	102	103	104	105	106	107	108	109
0	က 1000	တ 1010	င 1020	ိ 1030	ဝ 1040	ရ 1050	ု 1060	ယ 1070	ဆ 1080	ဝ 1090
1	ခ 1001	ထ 1011	အ 1021	ေ 1031	သ 1041	ဖ 1051	ရှ 1061	ိ 1071	ဂ 1081	၁ 1091
2	ဂ 1002	ဒ 1012	က 1022	ဲ 1032	၂ 1042	ဇ 1052	ါ 1062	ိ 1072	ေ 1082	၇ 1092
3	ယ 1003	မ 1013	ဒိ 1023	ိ 1033	၃ 1043	ဗ 1053	ာ 1063	ိ 1073	ါ 1083	၂ 1093
4	င 1004	န 1014	ဗြ 1024	် 1034	၄ 1044	ဇ 1054	ါ 1064	ိ 1074	ေ 1084	၂ 1094
5	စ 1005	ပ 1015	ဇာ 1025	ိ 1035	၅ 1045	ဇ 1055	ဝ 1065	ဂ 1075	ိ 1085	၃ 1095
6	ဆ 1006	ဖ 1016	ဇာ 1026	ိ 1036	၆ 1046	ာ 1056	၄ 1066	ခ 1076	ိ 1086	၄ 1096
7	ဇ 1007	ပ 1017	ဇ 1027	ိ 1037	၇ 1047	ာ 1057	ာ 1067	ဂ 1077	ိ 1087	၂ 1097
8	ဈ 1008	ာ 1018	ဇ 1028	ိ 1038	၈ 1048	ာ 1058	ာ 1068	ဂ 1078	ိ 1088	၃ 1098
9	ဉ 1009	မ 1019	ဗြ 1029	ိ 1039	၉ 1049	ာ 1059	ာ 1069	ဂ 1079	ိ 1089	၃ 1099
A	ည 100A	ယ 101A	ဗြ 102A	ိ 103A	ါ 104A	၄ 105A	ာ 106A	ာ 107A	ိ 108A	ိ 109A
B	ဋ 100B	ရ 101B	ါ 102B	ါ 103B	ါ 104B	ါ 105B	ါ 106B	ါ 107B	ါ 108B	ါ 109B

	AA6	AA7
0	က AA60	= AA70
1	က AA61	က AA71
2	က AA62	က AA72
3	က AA63	က AA73
4	က AA64	က AA74
5	က AA65	က AA75
6	က AA66	က AA76
7	က AA67	က AA77
8	က AA68	က AA78
9	က AA69	က AA79
A	က AA6A	က AA7A
B	က AA6B	က AA7B

Color convention[1]:

Yellow background - All characters that are included in the [MSR]

Pinkish background - Characters that are excluded from the [MSR]

White - Characters that are disallowed by IDNA2008

Range:

1000 ~ 109F, AA60 ~ AA7F

# Variant Code Points

<b>Glyph</b>	<b>Code Point</b>	<b>Character Name</b>
◌	U+101D	MYANMAR LETTER WA
o	U+006F	LATIN SMALL LETTER O
ο	U+03BF	GREEK SMALL LETTER OMICRON
о	U+043E	CYRILLIC SMALLER LETTER O
օ	U+0585	ARMENIAN SMALL LETTER OH

# Variant Code Points

<b>Glyph</b>	<b>Code Point</b>	<b>Character Name</b>
ၵ	U+1002	MYANMAR LETTER GA
റ	U+0D31	MALAYALAM LETTER RRA
ი	U+10D8	GEORGIAN LETTER IN

# Whole Label Evaluation Rules

---

## Classification

C → Consonant (U+1000 ... U+)

IV → Independent Vowel 1023 ... 102A

DV → Dependent Vowel Sign 102B ... 1038, 1062, 1083, 1084, 1086

K → Killer or Asat: 103A \$ ̣

VIRAMA → 1039 “ ̣ ” (It is always in between two consonants and invisible, “ ̣ ” )

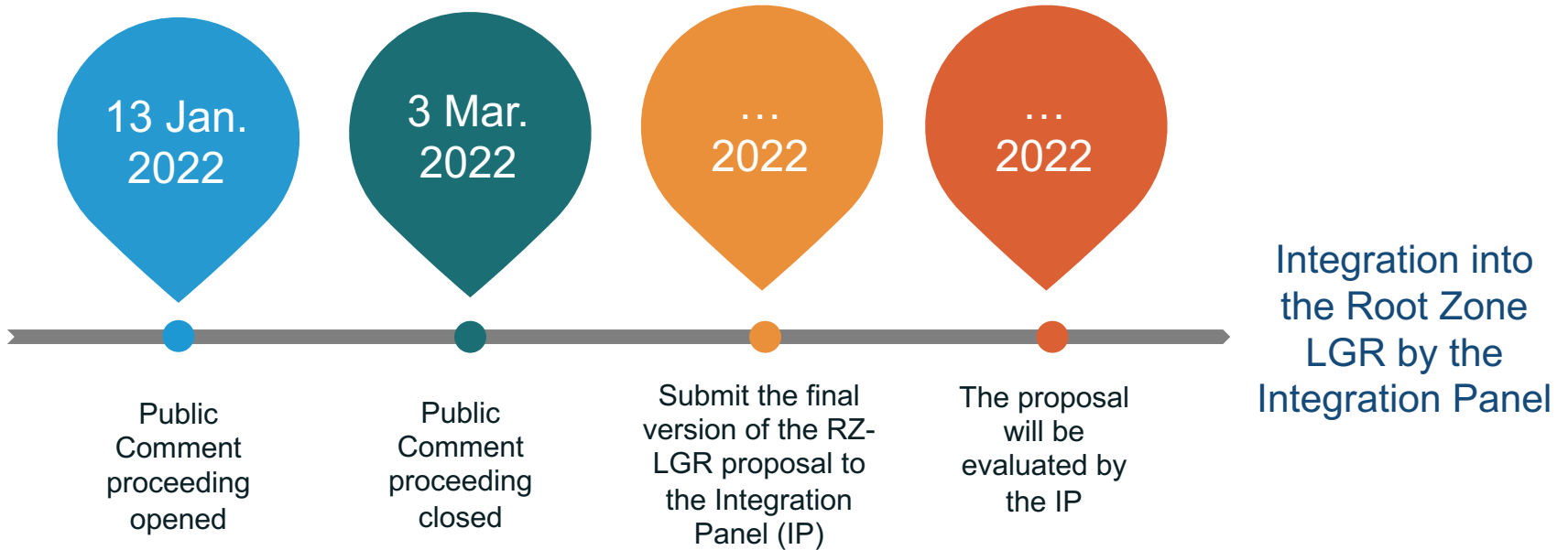
M → Dependent Consonant Sign (Medial)

...

# Whole Label Evaluation Rules

1. DV must follow C or M (example label - ဂ္ဂုစု, i.e., villages)
2. ANUSVARA must follow C or M (example label - အလံ, i.e., flag)
3. Rules of Medials combining with Consonants (example label - ရွှေပြည်, i.e., golden land)
  - Rules for Single Medial
  - Rules for Combined Medial
4. The Myanmar Great Saa (C1 or U+103F) must follow C or M or DV or 1023 or 1025 (example label - ဒဿန, i.e., philosophy)
5. (C+K) or (C2 + S16) or (C3+S17) must follow (example label - ရန်ကုန်, i.e., Yangon)
  - C or M or DV or OV
6. S11 must follow C or M or DV and another C must follow S11 (example label - မင်္ဂလာ)
7. VIRAMA must be in between two C (C+VIRAMA+C) (example label - ဥက္ကလာ)
  - C cannot be in between VIRAMAs to prevent virama-c-virama-c
  - The sequence containing a VIRAMA (U+1039) must not follow or precede a VIRAMA

# Current Step: Ongoing Public Comment Proceeding



- Link to the Public Comment Proceeding: <https://www.icann.org/en/public-comment/proceeding/proposal-for-myanmar-script-root-zone-label-generation-rules-13-01-2022>
- Everyone is encouraged to review the proposal and provide comments and suggestions.
  - Both minor and major comments are welcome.
  - All input will be considered by the Myanmar GP.

# Root Zone Label Generation Rules (RZ-LGR-5) Update

Marc Blanchet  
Integration Panel Member



# RZ-LGR-5: Scripts

---

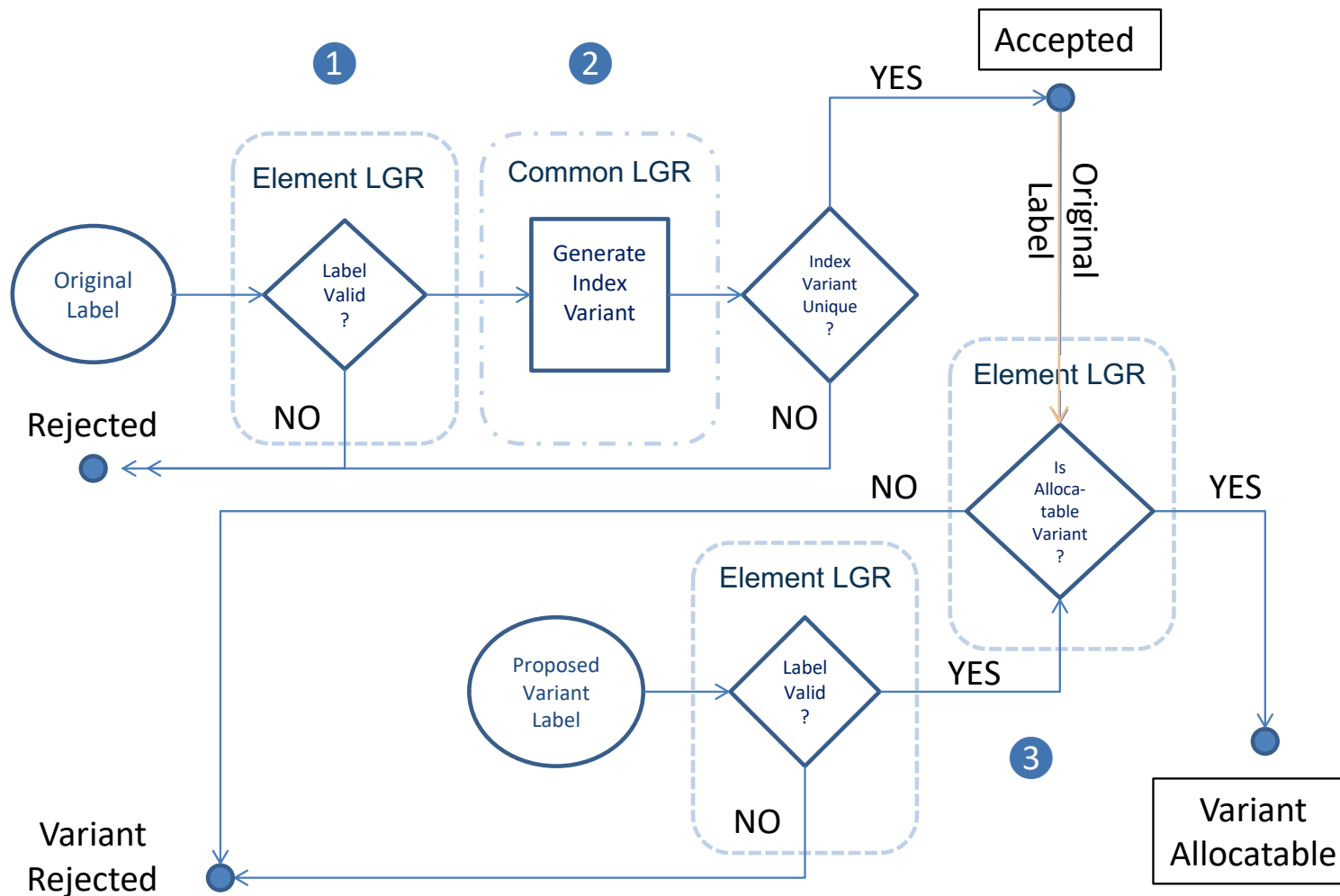
- ⦿ RZ-LGR-5: 5<sup>th</sup> version of the [Root Zone LGR](#)
- ⦿ Total 26 scripts
  - From RZ-LGR-4 (18 scripts):
    - Arabic, Bangla, Chinese(Han), Devanagari, Ethiopic, Georgian, Gujarati, Gurmukhi, Hebrew, Kannada, Khmer, Lao, Malayalam, Oriya, Sinhala, Tamil, Telugu, and Thai.
  - New scripts(8-2xHan=6):
    - Greek, Japanese(Hiragana, Katakana, Han), Korean(Hangul, Han), Latin, Myanmar.
  - Deferred scripts now being integrated into LGR5(2):
    - Armenian, Cyrillic
  - Two sets of related scripts are now fully integrated:
    - » Chinese, Japanese, Korean
    - » Latin, Greek, Cyrillic, Armenian

# RZ-LGR-5: Cross-Script Variants

---

- ⦿ Cross-script variants only complete in Common LGR.
- ⦿ Common LGR is used for collision testing among valid labels.
- ⦿ Requirements for variants processing:
  - Any variant uniquely defined by an Element LGR must be listed so it can be integrated.
  - Any in-script (in-repertoire) variants are always listed in the affected LGR.
  - ASCII variants are always listed (by policy).
  - Elements LGRs from the set of L/G/C/A scripts will list all variants within the set, but only Latin LGR will list variants outside that set.
  - Previously deferred scripts (Cyrillic, Armenian) have variants imposed by integration added.
  - A few other Element LGRs (Malayalam) have selected variants added that were imposed by new scripts in RZ-LGR-5.

# Steps for Processing a Label Using the Root Zone LGR



# Next Steps

---

- ⦿ RZ-LGR-5, consisting of 26 scripts, is expected to be published for Public Comments on 24 March.
  - Please review and comment!
  
- ⦿ Future:
  - Tentative target for finalizing RZ-LGR-5: Q2 of CY 2022.
  - Possible next version of the Root Zone LGR:
    - Tibetan and Thaana Generation Panels to be formed.
    - RZ-LGR can be updated after proposals have been submitted by those GPs and reviewed by IP.

⦿ Questions?

# Engage with ICANN and IDN Program



**Thank You and Questions**

Visit us at [icann.org/idn](http://icann.org/idn)

Email: [IDNProgram@icann.org](mailto:IDNProgram@icann.org)



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



[facebook.com/icannorg](https://facebook.com/icannorg)



[youtube.com/icannnews](https://youtube.com/icannnews)



[flickr.com/icann](https://flickr.com/icann)



[linkedin/company/icann](https://linkedin/company/icann)



[slideshare/icannpresentations](https://slideshare/icannpresentations)



[soundcloud/icann](https://soundcloud/icann)