

I C A N N
ANNUAL GENERAL

63

BARCELONA

20-25 October 2018



IDN Root Zone LGR Workshop



Sarmad Hussain
Director, IDN Programs

ICANN63
24 October 2018

Overview of Session Presentations

- ⊙ RZ-LGR on Defining Variants - Asmus Freytag
- ⊙ Update on LGR Toolset - Audric Schiltknecht
- ⊙ Update by RZ-LGR Study Group - Dennis Tanaka
- ⊙ Community Updates
 - Chinese GP Update - Wei Wang
 - Korean GP Update - Dongman Lee
- ⊙ Q/A

Root Zone Label Generation Rules On Defining Variants

Asmus Freytag, Ph.D.
Member of Integration Panel

Agenda

- ⦿ Variants Scope
- ⦿ What Makes Code Points Substitutable?
- ⦿ Variant Disposition and Security
- ⦿ Extra Blocked Variants
- ⦿ Examples of Inherited Variants
- ⦿ Limits on Allocatable Variants
- ⦿ Integration: Cross-repertoire variants
- ⦿ Similarity vs. Variant
- ⦿ References

Variants Scope

- ⦿ “An IDN variant ...is an alternate code point (or sequence of code points) that could be substituted for a code point (or sequence of code points) in a candidate label to create a variant label that is considered the “same” in some measure by a given community of Internet users.”

— [Procedure] Section A.3.2

- ⦿ “**Contextual Safety Principle:** A code point in the Zone Repertoire or any of its Variants that present unacceptable risks of being used in malicious ways should not be permitted.”

— [Procedure] Section A.3.6

What Makes Code Points Substitutable?

- Same semantic
 - Chinese Simplified ↔ Traditional
- Same pronunciation (Ethiopic)
 - U+1200 ሀ HA ↔ U+1210 ሐ HHA
 - U+1200 ሀ HA ↔ U+1280 ጎ XA
 - U+1210 ሐ HHA ↔ U+1280 ጎ XA
- Same appearance (indistinguishable)
 - U+0B20 Ꭰ ORIYA TTHA
 - U+0D20 Ꭰ MALAYALAM TTHA
 - U+006F Ꭰ LATIN SMALL O
 - U+043E Ꭰ CYRILLIC SMALL O

Variant Disposition and Security

- ⊙ Disposition of a pair of labels that are variants
 - **Allocable:** both may be delegated, to same entity
 - **Blocked:** either may be delegated, never both

- ⊙ Effect on Security
 - Blocked variants:
 - prevent certain malicious registrations
 - the more blocked variants, the more secure.
 - Allocatable variants:
 - allow one entity to offer multiple equivalent labels where required; while preventing registrations by unrelated entities
 - must be restricted or lead to combinatorial explosion

Extra Blocked Variants

- ⦿ “[Integration may] produce labels that would violate the linguistic criteria for being considered true variants, and may also result in the generation of extra blocked variants that lead to the exclusion of other possibly useful labels. It is nevertheless appropriate in the root zone, where the **goal is not to maximize the number of possible labels** but to minimize the confusion possible in a shared environment supporting heterogeneous linguistic communities.”
 - [Procedure] Section B.4.1 (emphasis added)

- ⦿ In effect, the procedure
 - Permits additional blocked variants over and beyond the minimum
 - Does not recognize mere name-space restriction as a valid argument against variants
 - Focuses on making a shared zone safe for all users

Examples of Inherited Variants

◉ Arabic FEH/QAF

- U+0641 ف FEH ↔ U+0642 ق QAF (blocked)
- The two letters are distinct, but become variants due to transitivity because of a third character: 06A7 في
- 06A7 في is a semantic variant of U+0642 ق and a visual variant of U+0641 ف in middle and initial positions.

◉ Ethiopic homophones:

- Amharic language has many *phonetic* variants (homophones).
- These are not variants for other languages using Ethiopic.
- For security reasons, these are applied to all languages
- Despite the large number of homophones, only about 1% of words in other languages collide with other words in the same language
 - Still available, but on a first-come basis

Limits on Allocatable Variants

- ⦿ “The benefits of a strictly minimal variant set apply only to those variants for which the returned disposition would be “allocatable” .”
- ⦿ “...the output of this procedure should aim to maximize the number of blocked variants, and to minimize the number of allocatable variants.”

— [Procedure] Section A.3.3

- ⦿ The procedure treats the two types of variants differently:
 - Allocatable variants are to be the minimum set
 - Blocked variants are to be maximized

Integration: Cross-repertoire Variants

- ⦿ “...the integration panel creates additional variant rules to make the entire set ***transitive and symmetric***. The disposition for any such additional variants are implicit [and] always of type “blocked”. For those implicit variant rules that fall entirely *within a named repertoire*, the corresponding generation panel proposal will be rejected. (The generation panel would then reissue the proposal with the required rules added, but with explicit dispositions as chosen by the generation panel). ”

— [Procedure] B.4.1 (emphasis added).

- ⦿ Each LGR:
 - defines variants based on linguistic requirements
 - inherits any applicable cross-repertoire variants from other LGRs
 - must define disposition for any inherited *in-repertoire* variants.
 - may optionally define matching cross-script variants
see [Out-of Repertoire-Variants]

Examples of Cross-repertoire Variants

- Cyrillic LGR defines

0443 y ↔ 0079 y

- Latin LGR defines

0079 y ↔ 0443 y (matches Cyrillic)
0079 y ↔ **04AF** **γ** (new)

- Transitivity: *implicit* in-script variant for Cyrillic

0443 y ↔ **04AF** **γ** (Cyrillic GP must define matching mapping and assign variant type)

- Note: never possible to cause in-repertoire variants in ASCII

Similarity vs. Variant

- ⦿ “...the process may not be able to replace case-by-case analysis altogether: there will still be a role for additional types of review, such as for String Similarity, and which are not included in the LGR process. Instead the LGR process is **designed to clear the table of all the straightforward, non-subjective cases**, mainly by returning a “blocked” disposition” — [Procedure] Section A.3.3

- ⦿ Limit considerations of appearance to cases that
 - are unambiguous
 - have overriding security concerns
 - exhibit true exchangeability (homoglyph)

Examples

- ⦿ The “circle”

Code	Glyph	Name
-------------	--------------	-------------

006F	o	LATIN SMALL LETTER O
03BF	ο	GREEK SMALL LETTER OMICRON
043E	о	CYRILLIC SMALLER LETTER O
0585	օ	ARMENIAN SMALL LETTER OH
0B20	ᱠ	ORIYA LETTER TTHA
0D20	ഠ	MALAYALAM LETTER TTHA
101D	ဝ	MYANMAR LETTER WA

- ⦿ simple glyph (few clues as to script membership)
- ⦿ part of many scripts
- ⦿ can be used to spoof delegated ASCII-TLD “.ooo”

CJK Visual Examples

Script	Glyph	Meaning	Script	Glyph
Han	今日	Chinese: "today"	Hangul + Han	스님
Han	人丛	Chinese: "crowd"	Han + Hangul	人스
Han	占卜	Chinese: "fortune telling"	Han+Kana	占卜
Katakana	卜口	Japanese: "fatty part of tuna"	Han	卜口
Han	樽	U+58AB vs. U+58FF	Han	樽
Kana	一	Length Mark	Han	一

References

- ⦿ [IIR]: “A Study of Issues Related to the Management of IDN Variant TLDs,” <https://www.icann.org/en/system/files/files/idn-vip-integrated-issues-final-clean-20feb12-en.pdf>
- ⦿ [ISO 15924] “Codes for the representation of names of scripts”, ISO 15924, <https://www.unicode.org/iso15924/iso15924-codes.html>
- ⦿ [Unicode 10-Han] In Unicode 10.0 Core Specification, Chapter 18 “East Asia”, sub-section 18.1 Han (page 682):
<https://www.unicode.org/versions/Unicode10.0.0/ch18.pdf>
- ⦿ [ISO10646] “Universal Multiple-Octet Coded Character Set”, ISO/IEC 10646:2017 (5th Edition),
http://standards.iso.org/ittf/PubliclyAvailableStandards/c069119_ISO_IEC_10646_2017.zip

References (cont'd)

- ⊙ [Annex S] In [ISO10646], pages 2681-2691, Annex S “Procedure for the unification and arrangement of CJK ideographs.”
- ⊙ [Procedure] ICANN, “Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels, “
<https://www.icann.org/en/system/files/files/draft-lgr-procedure-20mar13-en.pdf>
- ⊙ [UTS 39-confusable] “Unicode Security Mechanism”,
<http://www.unicode.org/reports/tr39/> , list of confusable
<https://unicode.org/Public/security/latest/confusables.txt> , others
related data files available in: <https://unicode.org/Public/security/latest/>
- ⊙ [Out-of-Repertoire-Variants] Integration Panel, “Background:Out of Repertoire Variants in Root-Zone LGR and Proposals”, 2017-09-25
<https://www.icann.org/en/system/files/files/root-zone-lgr-repertoire-variants-25sep17-en.pdf>

Update on LGR Toolset

Audric Schiltknecht, Julien Bernard, Marc Blanchet
Viagénie

Contents

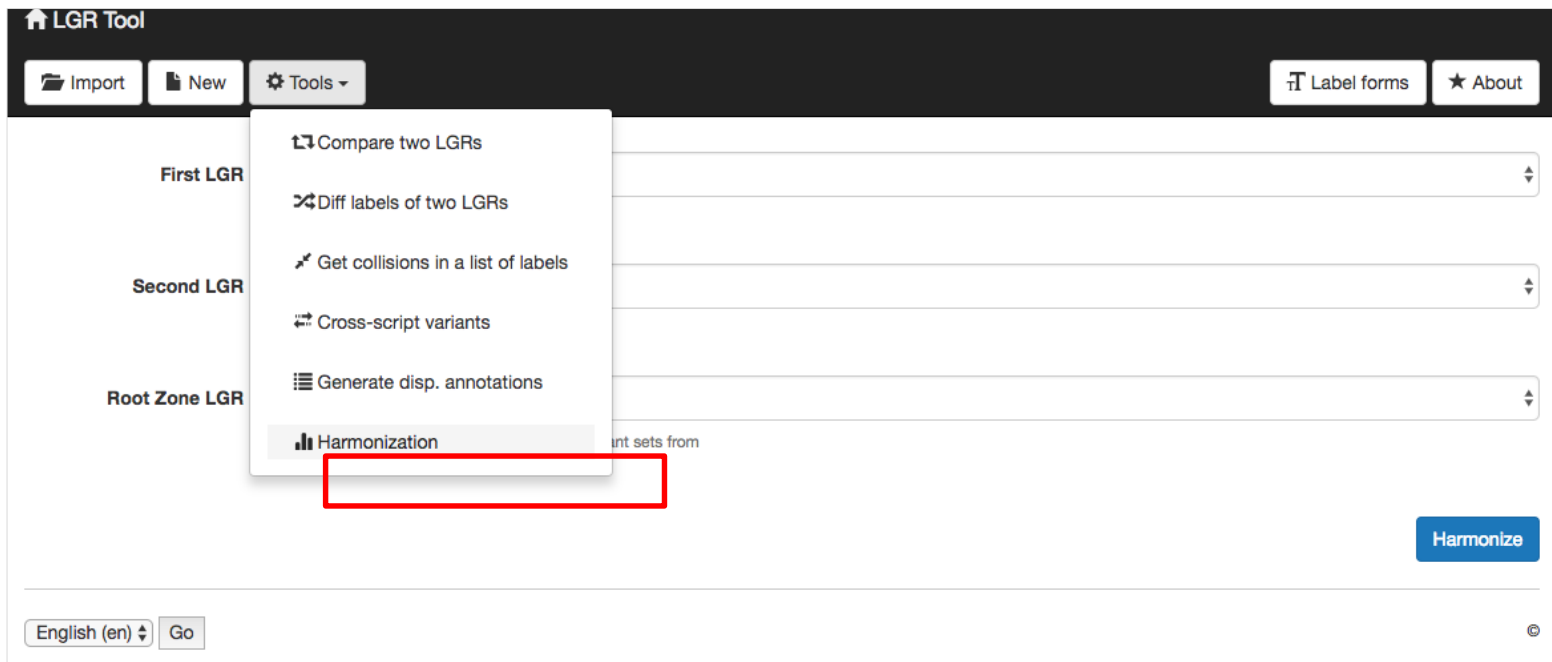
- ⦿ LGR toolset summary in one slide
- ⦿ New updates in Oct 2018 version
 - LGR harmonization tool
 - Interface improvements
 - Various improvements/ bug fixes

LGR Toolset Summary

- ⊙ Toolset to:
 - Create, update, use Label Generation Rules
 - Validate labels, generate variants, verify collisions
- ⊙ Available at <https://www.icann.org/resources/pages/lgr-toolset-2015-06-21-en>
 - Opensource
 - Online as a service
- ⊙ As
 - Command line and libraries in python
 - Web interface

LGR Tool - Harmonization

- Compare two LGRs and harmonize them
 - Same variant mappings for shared code points between both LGRs with respect to symmetry and transitivity
 - Variant mapping discovering in RZ-LGR



LGR Interface – Add WLE Rules to Selected Code Points

🏠 LGR Tool / test01

📁 Import 📄 New ⚙️ Tools ▾

✓ Validate label 📄 Validate LGR 🔄 Output ▾ 🏷️ Label forms ★ About

Code points References Meta data Tags

✓ ⇒ Apply batch action... 🔍 Search 📄 Show 500 entries

Add WLE Add Tags

1 U+0E04 (๓) 0 Variant(s) THAI CHARACTER KHOMUT

U+0E05 (๓) 0 Variant(s) THAI CHARACTER HONGSAWANG

U+0E06 (๓) 0 Variant(s) THAI CHARACTER KHO RAKHANG

U+0E07 (๓) 0 Variant(s) THAI CHARACTER NGO NGU

2 **Add Rule**

when

not-when

This rule will be added to 3 codepoint(s)

Next

See code point

See code point

Code point	Character	Action
U+0E01 (๓) 0 Variant(s)	THAI CHARACTER HONGSAWANG	<input type="checkbox"/>
U+0E02 (๓) 0 Variant(s)	THAI CHARACTER KHOMUT	<input type="checkbox"/>
U+0E03 (๓) 0 Variant(s)	THAI CHARACTER HO TIN	<input type="checkbox"/>
U+0E04 (๓) 0 Variant(s)	THAI CHARACTER KHOMUT	<input checked="" type="checkbox"/>
U+0E05 (๓) 0 Variant(s)	THAI CHARACTER HONGSAWANG	<input checked="" type="checkbox"/>
U+0E06 (๓) 0 Variant(s)	THAI CHARACTER KHO RAKHANG	<input checked="" type="checkbox"/>
U+0E07 (๓) 0 Variant(s)	THAI CHARACTER NGO NGU	<input type="checkbox"/>

LGR Interface – Add Tags to Selected Code Points

🏠 LGR Tool / test01

📁 Import 📄 New ⚙️ Tools ▾

✓ Validate label 📄 Validate LGR 🔄 Output ▾ 🗒 Label forms ★ About

Code points References Meta data Tags Rules

✓ → Apply batch action...
Add WLE
Add Tags

Show

1

2

3

Tags

space-separated tags

These tags will be added to 3 codepoint(s)

Next

<input type="checkbox"/>	U+0E01 (n) 0 Variant(s)				
<input type="checkbox"/>	U+0E02 (๓) 0 Variant(s)				
<input type="checkbox"/>	U+0E03 (๓) 0 Variant(s)				
<input checked="" type="checkbox"/>	U+0E04 (๓) 0 Variant(s)				
<input checked="" type="checkbox"/>	U+0E05 (๓) 0 Variant(s)				
<input checked="" type="checkbox"/>	U+0E06 (๓) 0 Variant(s)	THAI CHARACTER KHO RAKHANG			See code point
<input type="checkbox"/>	U+0E07 (๓) 0 Variant(s)	THAI CHARACTER NGO NGU			See code point

List Tags and Their Associated Code Points

🏠 LGR Tool / proposed-lgr-telu-20180808 - Proposed LGR for Telugu

📁 Import 📄 New ⚙️ Tools ▾

✓ Validate label 📄 Validate LGR ↻ Output ▾ 🗒️ Label forms ★ About

[Code points](#) [References](#) [Meta data](#) **[Tags](#)** [Rules](#)

Existing tags

Tag name	Associated code points	Action
Anusvara	U+0C02 (ఱ)	
Visarga	U+0C03 (ఱః)	
Vowel	U+0C05 (అ) U+0C06 (ఆ) U+0C07 (ఇ) U+0C08 (ఈ) U+0C09 (ఊ) U+0C0A (ఋ) U+0C0B (ౠ) U+0C0E (ఎ) U+0C0F (ఐ) U+0C10 (ఔ)	View all (13)
Consonant	U+0C15 (క) U+0C16 (ఖ) U+0C17 (గ) U+0C18 (ఘ) U+0C19 (ఙ) U+0C1A (చ) U+0C1B (ఛ) U+0C1C (జ) U+0C1D (ఝ) U+0C1E (ఞ)	View all (34)
Nasal-Consonant	U+0C23 (ణ) U+0C28 (న) U+0C2E (మ) U+0C19 (జ) U+0C1E (ఞ)	
Matra	U+0C40 (్) U+0C41 (్ఱ) U+0C42 (్ఱఱ) U+0C43 (్ఱ్) U+0C44 (్ఱ్ఱ) U+0C46 (్రి) U+0C47 (్రి్) U+0C48 (్రి్ఱ) U+0C4A (్రి్) U+0C4B (్రి్ఱ)	View all (13)
Halant	U+0C4D (్)	

📘 To create a new tag: Add it to a code point first to find it back in this list

English (en) ▾ Go

©

LGR Improvements – Populate Variants

🏠 LGR Tool / test02

📁 Import 📄 New ⚙️ Tools ✓ Validate label 📄 Validate LGR 🔄 Output 🗒 Label forms ★ About

Code points References Meta data Tags Rules

Populate variants Add code point(s)

➔ Select code point(s) to apply batch action... Show 500 entries Search:

<input type="checkbox"/>	Code point	Character Name	Tags	Comments	Action
<input type="checkbox"/>	U+0E01 (n) 1 Variant(s)	THAI CHARACTER KO KAI			See code point
<input type="checkbox"/>	U+0E16 (n) 0 Variant(s)	THAI CHARACTER THO THUNG		Automatically added from out-of-repertoire variant	See code point

Previous 1 Next

English (en) Go

LGR Improvements – Display Label Forms

↑ LGR Tool / test02

Import New Tools

Validate label Validate LGR Output

Label forms About

Display all label forms

Label

Unicode version

The unicode version used

Display forms

Code point sequence	U-label	A-label
U+0E2A U+0E27 U+0E31 U+0E2A U+0E14 U+0E35	สวัสดี	xn--l3c1bib8a0a

LGR Improvements – Adding Variants

- Automatically add a code point, reflexive variant and symmetric variant upon adding a variant for a non-existing code point

🏠 LGR Tool / test02 / U+0E16 (n) THAI CHARACTER THO THUNG

📁 Import 📄 New ⚙️ Tools ▾

✓ Validate label 📄 Validate LGR ↻ Output ▾ 🗒 Label forms ★ About

Code points / U+0E16 (n) THAI CHARACTER THO THUNG

Code point appeared in Unicode version: 1.1.0.0

Variants

Code point: Override repertoire: ①

Code point	Type	Comments	When	Not When	References	Action
U+0E01 (n) THAI CHARACTER KO KAI Age: 1.1.0.0	<input type="text" value="blocked"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>		Delete variant Edit references

LGR Improvements – Adding Variants

- Automatically add a code point, reflexive variant and symmetric variant upon adding a variant for an non-existing code point

LGR Tool / test02 / U+0E16 (๓) THAI CHARACTER THO THUNG

Import New Tools Validate label Validate LGR Output Label forms About

Automatically added codepoint U+0E20 from out-of-repertoire variant **2**

New variant U+0E20 added

Code points / U+0E16 (๓) THAI CHARACTER THO THUNG

Code point appeared in Unicode version: 1.1.0.0

Variants

Code point: 0E20 Override repertoire: Add variant

Code point	Type	Comments	When	Not When	References	Action
U+0E01 (๓) THAI CHARACTER KO KAI Age: 1.1.0.0	blocked					Delete variant Edit references
U+0E20 (๓) THAI CHARACTER PHO SAMPHAO Age: 1.1.0.0	blocked	3				Delete variant Edit references

LGR Improvements – Adding Code Points from a Script

LGR Tool / sample-french - Sample LGR for French

Import New Tools

Code points References Met

Code point Code point range **Code point from script** Import from file

Script: Latn

Validating repertoire: msr-3-wle-rules-28mar18-en

Manual import

Next

Select code point(s) to apply batch

<input type="checkbox"/>	Code point
<input type="checkbox"/>	U+0061 (a) 1 Variant(s)
<input type="checkbox"/>	U+0062 (b) 0 Variant(s)
<input type="checkbox"/>	U+0063 (c) 0 Variant(s)
<input type="checkbox"/>	U+0064 (d) 0 Variant(s)
<input type="checkbox"/>	U+0065 (e) 0 Variant(s)
<input type="checkbox"/>	U+0066 (f) 0 Variant(s)
<input type="checkbox"/>	U+0067 (g) 0 Variant(s)

LGR Tool / sample-french - Sample LGR for French

Import New Tools

Code points References Met

Code point Code point range Code point from script Import from file

Code points:

- U+00DF LATIN SMALL LETTER SHARP S
- U+00E0 LATIN SMALL LETTER A WITH GRAVE
- U+00E1 LATIN SMALL LETTER A WITH ACUTE
- U+00E2 LATIN SMALL LETTER A WITH CIRCUMFLEX
- U+00E3 LATIN SMALL LETTER A WITH TILDE
- U+00E4 LATIN SMALL LETTER A WITH DIAERESIS
- U+00E5 LATIN SMALL LETTER A WITH RING ABOVE
- U+00E6 LATIN SMALL LETTER AE
- U+00E7 LATIN SMALL LETTER C WITH CEDILLA
- U+00E8 LATIN SMALL LETTER E WITH GRAVE
- U+00E9 LATIN SMALL LETTER E WITH ACUTE
- U+00EA LATIN SMALL LETTER E WITH CIRCUMFLEX
- U+00EB LATIN SMALL LETTER E WITH DIAERESIS
- U+00EC LATIN SMALL LETTER I WITH GRAVE

Select code point(s) to apply batch

<input type="checkbox"/>	Code point
<input type="checkbox"/>	U+0061 (a) 1 Variant(s)
<input type="checkbox"/>	U+0062 (b) 0 Variant(s)
<input type="checkbox"/>	U+0063 (c) 0 Variant(s)
<input type="checkbox"/>	U+0064 (d) 0 Variant(s)
<input type="checkbox"/>	U+0065 (e) 0 Variant(s)
<input type="checkbox"/>	U+0066 (f) 0 Variant(s)
<input type="checkbox"/>	U+0067 (g) 0 Variant(s)

LATIN SMALL LETTER G

See code point

LGR Improvements – Misc.

- ⦿ Improve performances
 - Improve loading time for large LGRs
- ⦿ Full Python3 support
- ⦿ Output on failing rules is more explicit
- ⦿ Supporting for Unicode 5.2.0-10.0.0 and MSR-2, MSR-3
- ⦿ Display combined form of sequences
- ⦿ Bug fixes

Availability and Links

- ⦿ Online deployment at: <https://lgrtool.icann.org/>
- ⦿ Open source package(s) released with BSD license at GitHub: [picu](#), [lgr-core](#), [lgr-django](#), [munidata](#)
- ⦿ Toolset information and user manual at: <https://www.icann.org/resources/pages/lgr-toolset-2015-06-21-en>

Update by RZ-LGR Study Group

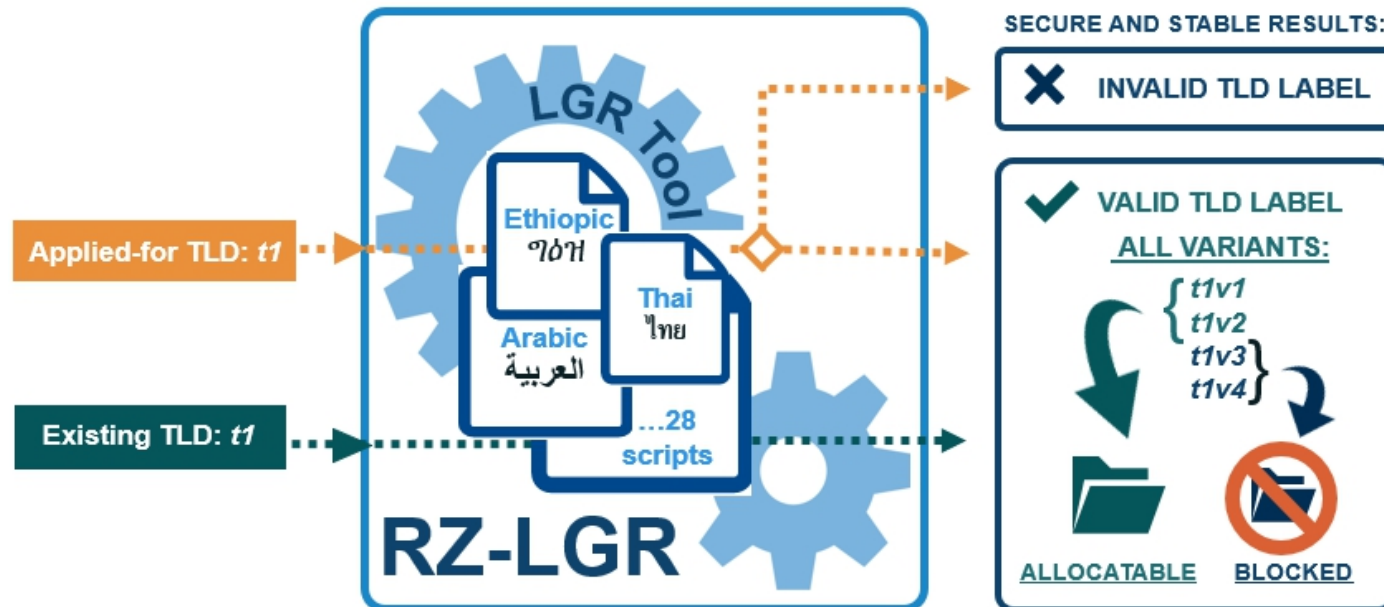
Dennis Tanaka

Agenda

- ⦿ Background
- ⦿ Scope of work
- ⦿ Current status
- ⦿ Next steps

Background

- ⦿ RZ-LGR available through the LGR Procedure
 - Several scripts already integrated; many others in-progress
- ⦿ Need of a harmonized way to use the RZ-LGR for ccTLDs and gTLDs
 - Single source to validate top-level labels and calculate variant labels
- ⦿ Need for a technical assessment of the implementation of the RZ-LGR
 - Technical considerations for subsequent policy



Background – Study Group Members

	Name	Organization	Sponsoring Organization
1	Mirjana Tasic	.rs and .cpб	ccNSO
2	Edmon Chung	.asia	GNSO
3	Gaurav Vedi	Dominion Registries	GNSO
4	Dusan Stojicevic	Gransy	GNSO
5	Dennis Tan Tanaka	Verisign	GNSO
6	Wei Wang	KNET	GNSO
7	Ajay Data	XGENPLUS	GNSO
8	Alireza Saleh	IRNIC	IAB
9	Dessaiegn Yehuala	Addis Abab Univ. and Ethiopic Generation Panel	

Scope of Work

1

WHO will use it?

- TLD applicant (ccTLD, gTLD)
- Generation and Integration Panels
- Other stakeholders

2

WHAT does it do?

- Syntax validation
- Calculation of variant labels and disposition values
- What if RZ-LGR calculation is not accepted?

3

WHY is it important?

- Single source and/or repository, for consistency and predictable results
- But, what about scripts not yet integrated in the LGR? What are the technical issues subsequent policy would need to address

4

WHEN do you apply it?

- Existing TLDs and new TLD applications
- gTLDs: application window
- ccTLDs: Fast Track process (rolling)
- Reserved TLD labels

5

WHERE do you find it?

- Implementation (i.e, specs, test cases)
- Maintenance (e.g., update to repertoire, variant rules, etc.)
- Repository of normative XML

6

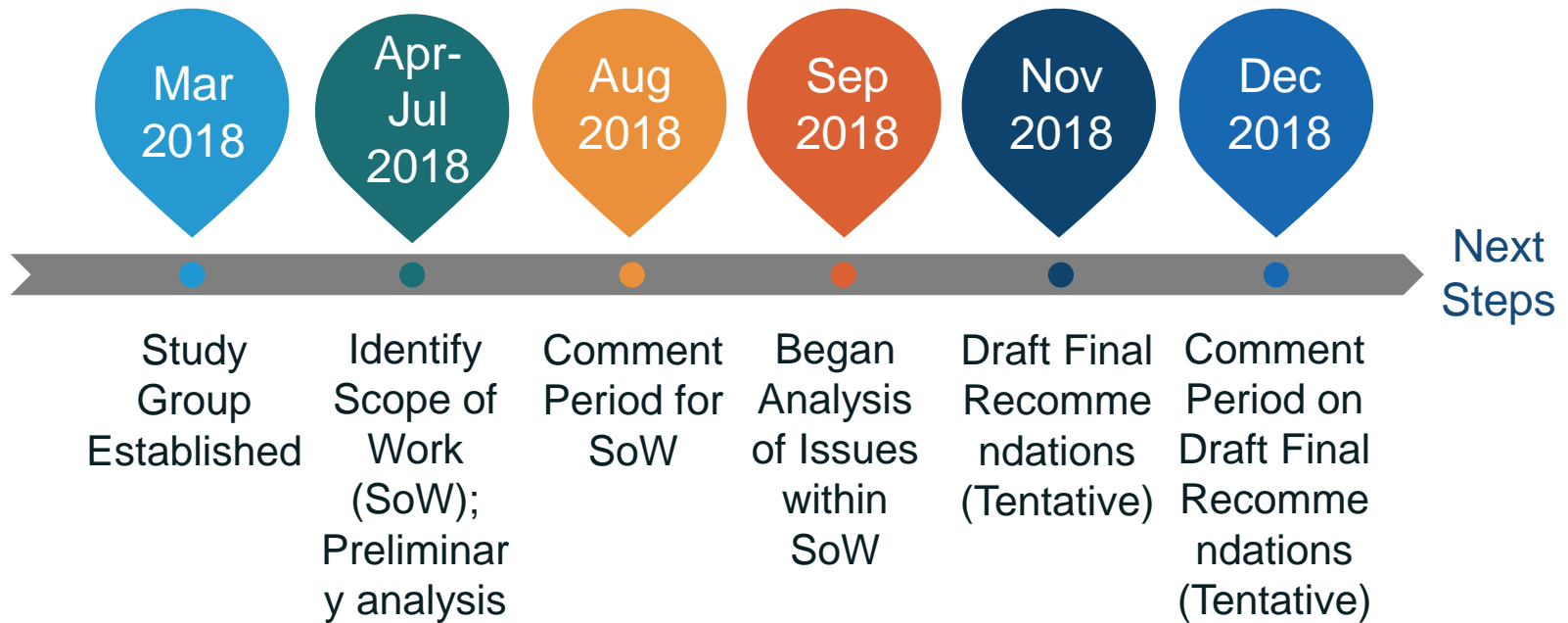
Other Considerations

- Variant states and transition among states
- Limits on allocatable variant labels
- Other security and stability considerations (e.g., single character IDN TLDs)

Not in Scope

- ⦿ Semantic validation
 - IDN ccTLD, Geo-Names, Brand, Community, etc.
- ⦿ Limiting number of allocable variant TLDs
- ⦿ How to process TLD applications whose script is not yet supported by the Root Zone LGR.

Status



Resource



<https://community.icann.org/display/croscomlgrprocedure/Study+Group+on+Technical+Use+of+RZ-LGR>



<https://mm.icann.org/pipermail/rz-lgr-sg/>

Chinese Generation Panel Update

Kenny Huang
Wei Wang

Agenda

1

CGP Work Review

2

CJK Coordination

3

CGP Proposal Draft
201808

4

Visual Similarity

5

Next Step

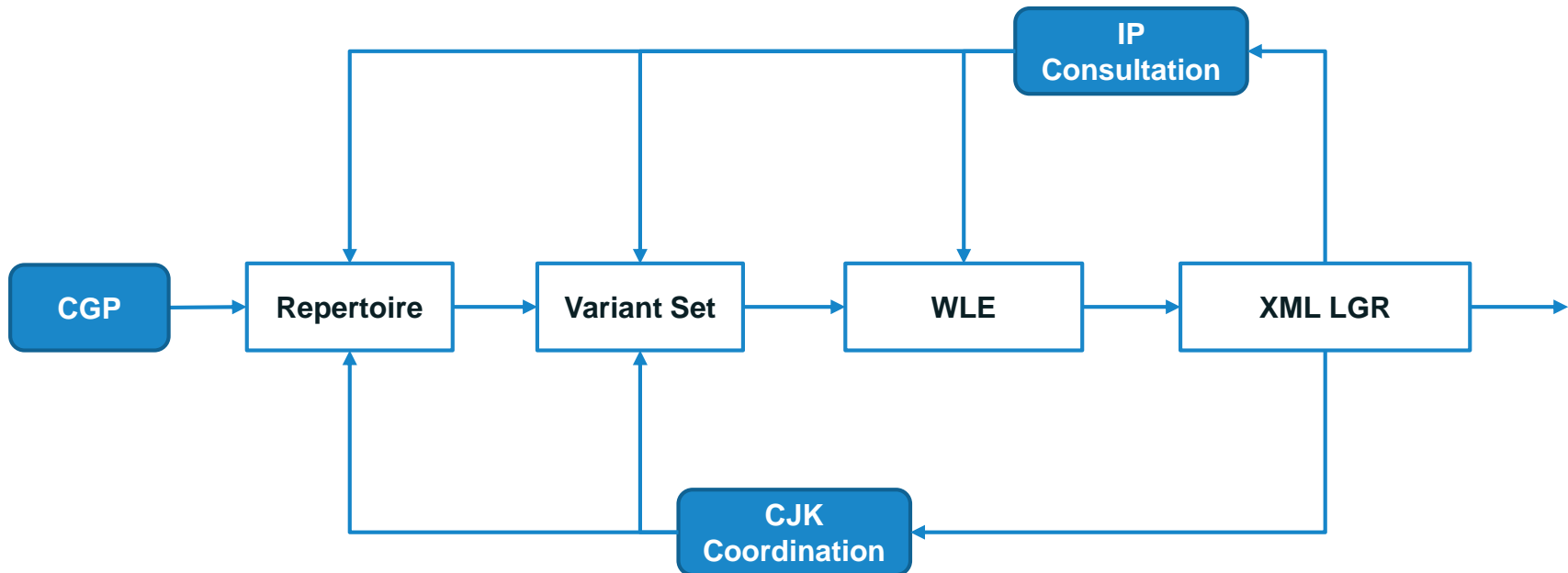
Members, 23 experts from 10 countries/regions

China mainland, Taiwan, Hong Kong, Macau, Singapore, Malaysia, as well as members from Europe and North America.

Advisor, Edmon CHUNG

CEO of dotAsia and Co-Chair of the Universal Acceptance Steering Group

CJK coordination working group





Agenda

1

CGP Work Review

2

CJK Coordination

3

CGP Proposal Draft
201712

4

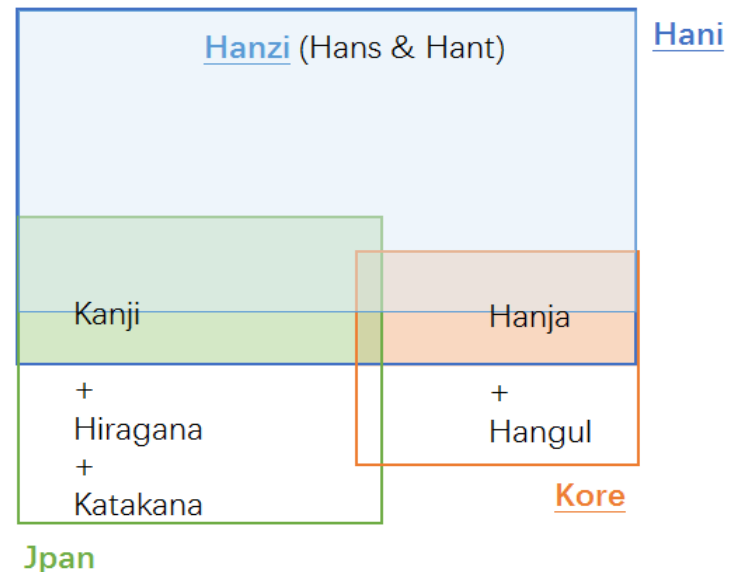
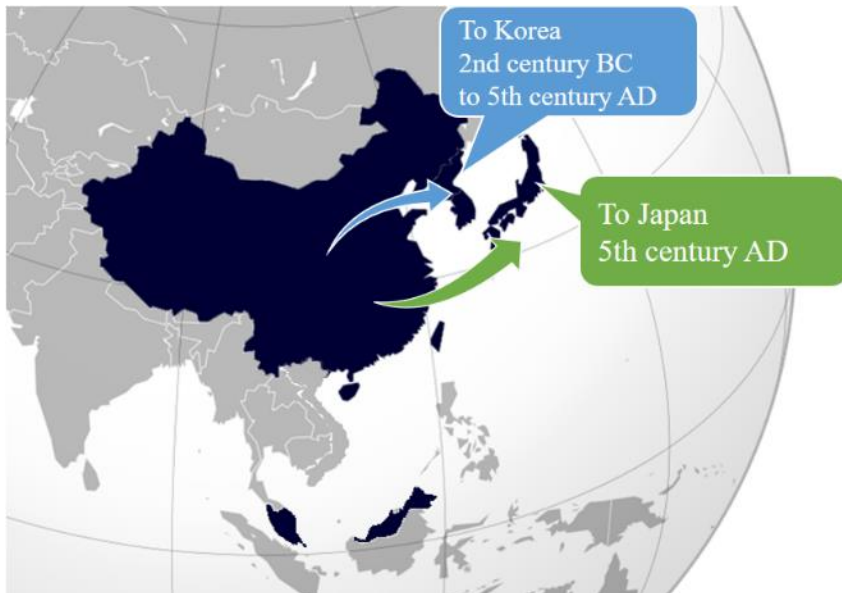
Visual Similarity

5

Next Step

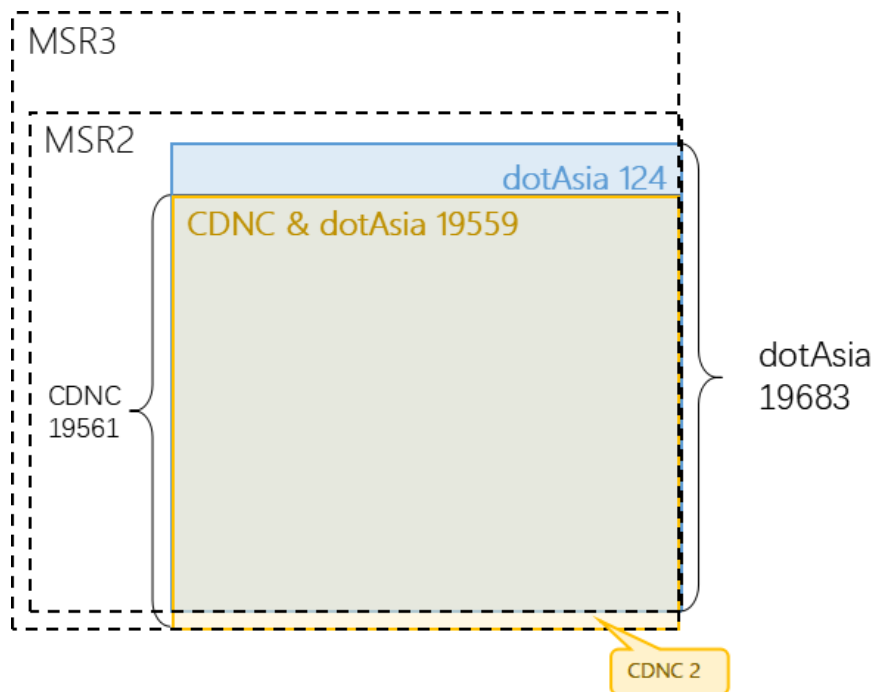
Script and Languages Covered

Language	ISO 15924 Code	Countries	Local Names of the Script
Chinese	cdo, cjt, cmn, cpx, czh, czo, gan, hak, hsn, lzh, mnp, nan, wuu, yue, zho	China	汉字 Hanzi
Japanese	jpn	Japan	漢字 Kanji
Korean	kor	Korea	한자 Hanja



Coordination within CGP

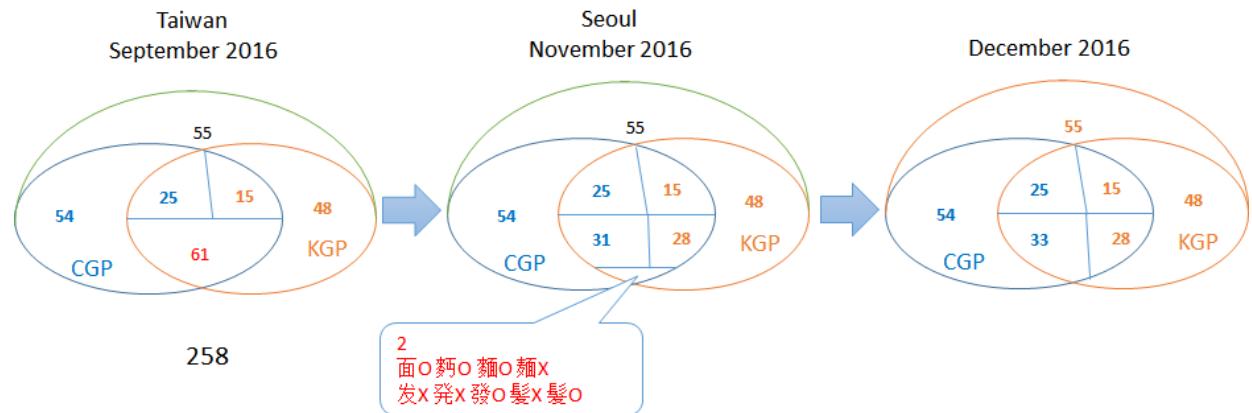
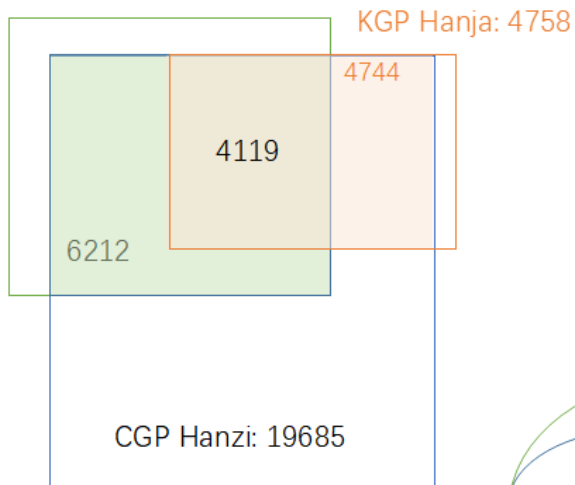
- 19030 variant mappings are kept as same as CDNC-2015 and dotAsia
- 64 are the same as CDNC-2015 but different with dotAsia
- 131 variant mapping entries from dotAsia
- 36 variant mapping entries from CGP internal review



Coordination between C, J and K

- 445 variant mappings (146 unacceptable variant groups)
- 424 Variant mapping entries changed by C&K pre-integration

JGP Kanji: 6356



Agenda

1

CGP Work Review

2

CJK Coordination

3

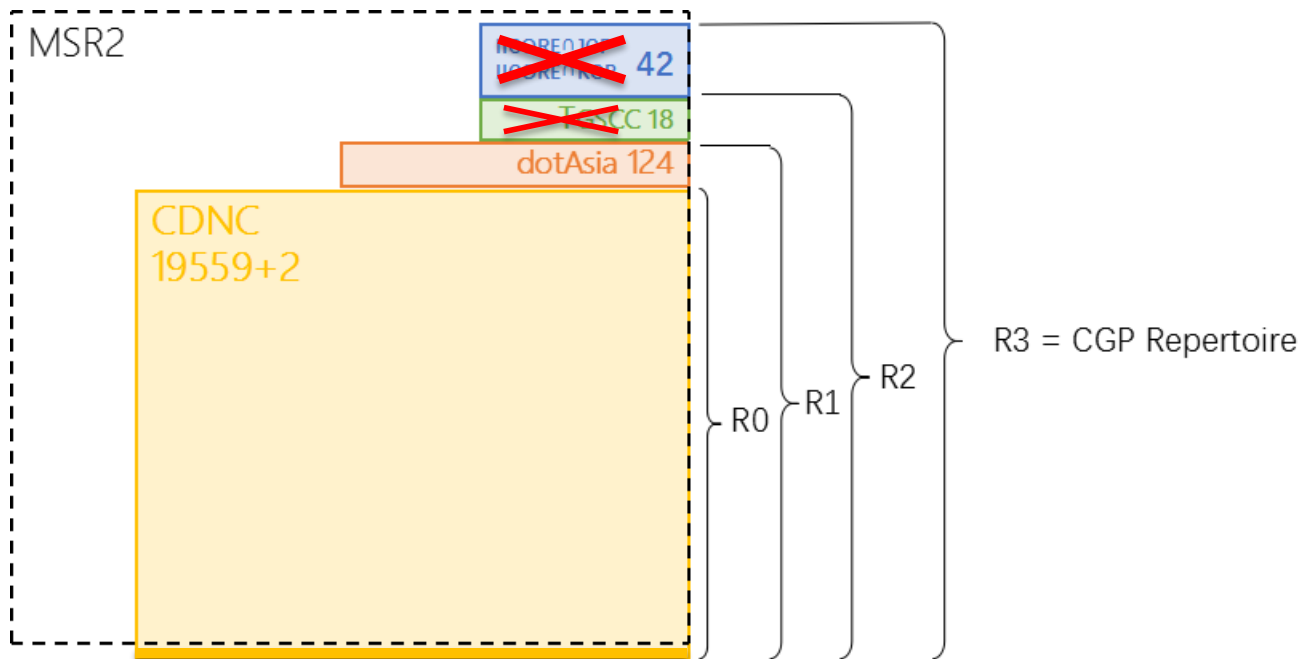
CGP Proposal Draft
201808

4

Visual Similarity

5

Next Step



19745 Code Points (201712)

>>

19685 Code Pionts (201808)

Accordingly, some variant groups and mappings changed due to the reduction.

Sub-Type	Type	Comment
“simp”	Allocable	preferred simplified variant char;
“r-simp”	Allocable	reflexive preferred simplified variant char;
“trad”	Allocable	preferred traditional variant char
“r-trad”	Allocable	reflexive preferred traditional variant char
“both”	Allocable	preferred simplified and traditional variants are the same
“r-both”	Allocable	reflexive preferred simp and trad variants are the same
“r-neither”	Blocked	Non-allocable reflexive/original char
“blocked”	Blocked	Non-allocable variant char
“out-of-repertoire-var”	Invalid	Non-CGP chars imported from other GPs

Agenda

1

CGP Work Review

2

CJK Coordination

3

CGP Proposal Draft
201808

4

Visual Similarity

5

Next Step

Unicode consortium's confusables list

<https://www.unicode.org/Public/security/11.0.0/confusables.txt>

Source	Glyph	Target	Glyph
3078	へ	30D8	へ
30A4	イ	4EBB	イ
30A8	エ	5DE5	工
30AB	カ	529B	力
30BF	タ	5915	夕
30C8	ト	535C	卜
30CB	ニ	4E8C	二
30CE	ノ	4E3F	ノ
30CF	ハ	516B	八
30ED	口	53E3	口
30ED	口	56D7	口
30FC	一	4E00	一
784F	研	7814	研
53E3	口	56D7	口
571F	土	58EB	土
58AB	樽	58FF	樽
676E	柿	67FF	柿
8D7F	越	8D86	越
9E42	鹵	9E43	鹵

Disposition Pinciple:

Non-modern used ones will be treated as visual identical variants

-- 58AB樽 & 58FF樽

Some will be kept with explanation

-- 571F土 & 58EB土

Some will be blocked as radical

-- 56D7 口

Respect the rules for kana-Kanji pairs made by JGP

Agenda

1

CGP Work Review

2

CJK Coordination

3

CGP Proposal Draft
201712

4

Visual Similarity

5

Next Step

Provide the detailed information of C&K coordination

Generate visual similarity list

Further interaction with IP

Korean Generation Panel Update

Dongman Lee

Agenda

1

Script(s) Covered
and where they
are used

2

Members of the
GP

3

Work achieved
to-date 1
(K-LGR v1.0)

4

Work achieved
to-date 2
(Public comments
reviewed)

5

Work achieved
to-date 3
(Brief history of KGP
activities)

6

Future Plan and
Schedule

Script(s) Covered by K-LGR and Where They Are Used

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

Members of the GP

- ⊙ 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: Eunjun JEON, Boknam YUN, Byeongil OH
- ⊙ Registry: Jinhyun CHO, Minjung PARK, Yunmi CHOI, Ryoung CHAE, Minjee KIM
- ⊙ Registration Agency: Seong-jin PARK, ChangKi JANG, Myungsoo LEE

Work achieved to-date by KGP – 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

Work achieved to-date by KGP – 2:

Public Comments Reviewed

- ⦿ A summary of public comments
 - 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/modified 4758 Hanja chars in K-LGR v1.0

- ⦿ Examples of issues raised by Mr. Byeon
 - 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

Work achieved to-date by KGP – 2:

Public Comments Reviewed

- ⊙ Requests by Mr. Byeon for specific details
 - Reviewed and discussed
 - Mostly accepted in principle and will be reflected in the next version of K-LGR

- ⊙ Hangeul-Hanja mixed labels
 - There is a general consensus to include Hanja in K-LGR repertoire
 - We have not reached a conclusion whether to allow Hangeul-Hanja mixed label

- ⊙ Because we think that it would be very difficult (or almost impossible) to get resolved as long as two schools of different views (i.e., Hangeul only vs. Hangeul-Hanja mixed) on Korean Language usage exist, it is expected that our stance would be to keep the current K-LGR (i.e., to allow Hangeul-Hanja mixed label)

Work achieved to-date by KGP – 2:

Plan and Next Steps

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

- ⊙ Hangul-Hanja mixed labels
 - Decide on a final conclusion

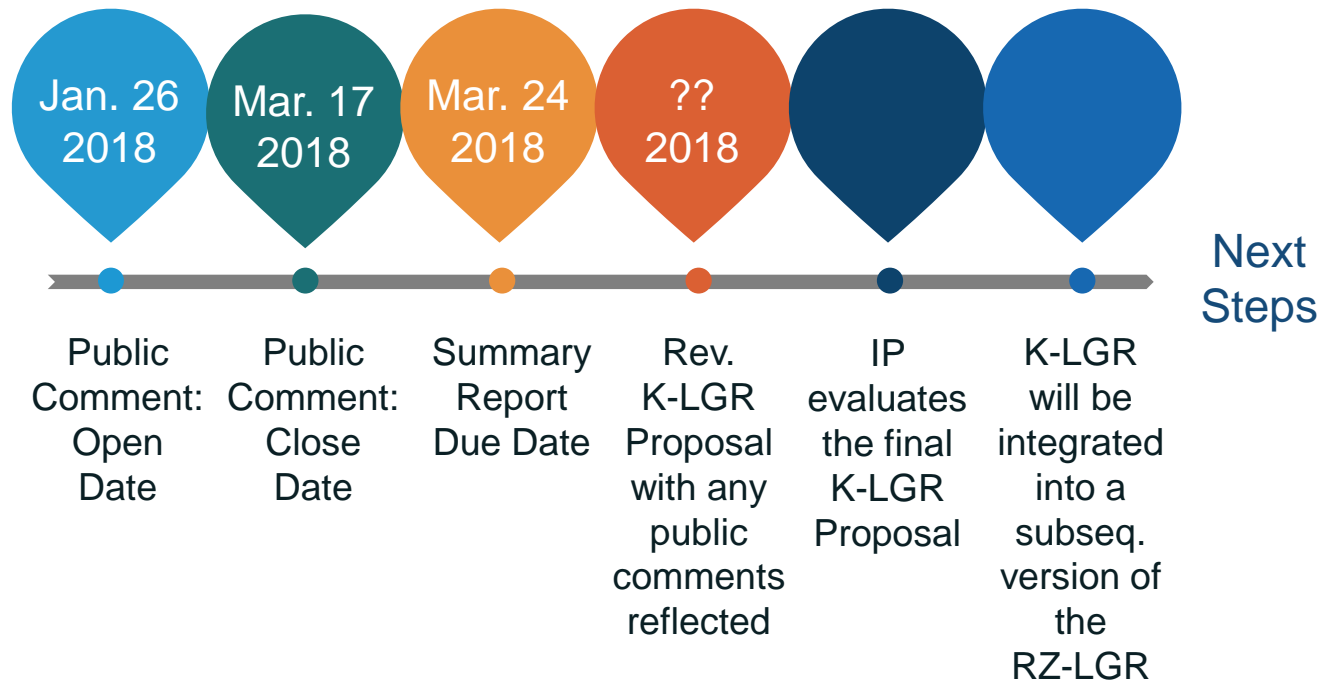
- ⊙ Revision of K-LGR 1.0
 - After the above issues are resolved, K-LGR will be revised and published

Work achieved to-date by KGP – 3:

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
- ⦿ Jan. ~ Mar. 2018: public comments for K-LGR v1.0
- ⦿ Mar. ~ Sep. 2018: public comments for K-LGR v1.0 reviewed for possible reflection in the next version of K-LGR
- ⦿ 33 KGP meetings
- ⦿ Several CJK coordination meetings during ICANN meetings 49 ~ 61
- ⦿ Several CJK coordination meetings in Rep. of Korea, China, and Taiwan.

Future Plan



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