

The background of the image is a night-time photograph of a cityscape in Abu Dhabi. On the left, several tall, modern skyscrapers with glass facades are illuminated. In the center, there is a large, ornate fountain structure. On the right, a large, domed building with a prominent archway is lit up. The foreground is filled with colorful flower beds in shades of red, pink, and white, with a paved walkway leading through them. The sky is a deep red color, suggesting a sunset or sunrise. The text is overlaid on the right side of the image in a white box.

ICANN
ANNUAL GENERAL

60

ABU DHABI

28 October–3 November 2017

IDN Root Zone LGR Workshop



ICANN 60
1 November 2017

Agenda

- ⦿ Overview of Root Zone LGR version 2 - Marc Blanchet
- ⦿ How to Use RZ-LGR-2 - Michel Suignard
- ⦿ Update on LGR Toolset - Audric Schiltknecht
- ⦿ Community Updates
 - Chinese GP Update - Kenny HUANG, Wei WANG
 - Japanese GP Update - Hiro Hotta
 - Korean GP Update - KIM Kyongsok
 - Greek GP Update - Panagiotis Papaspiliopoulos
- ⦿ Q/A

Root Zone LGR version 2 (RZ-LGR2)

Marc Blanchet
Integration Panel

What is RZ-LGR-2?

- ⦿ Set of normative XML files and informative documents for 6 scripts
- ⦿ Governs the way the root zone is operated for a given set of scripts
- ⦿ Determines which Unicode code points are permitted in U-labels
- ⦿ Determines which variants are allocatable or blocked
- ⦿ Output is used by other procedures determining whether a label is allocated, delegated

Script Based

- ⦿ Each label in root zone belongs to a single script
- ⦿ In the future, for cases like Japanese and Korean, the script concept will be extended to cover well-defined sets of primary scripts (as Kana and Kanji (i.e. Han) for Japanese)
- ⦿ RZ-LGR is released in stages to allow some LGRs to be available sooner in the root zone
 - Version 1 was Arabic only
 - Version 2 adds five more scripts
 - Many more in the pipeline

Scripts Covered

Script	Name in script
Arabic	العربية
Ethiopic	ፊደል
Georgian	ქართული ენა
Khmer	ភាសាខ្មែរ
Lao	ອັກສອນລາວ
Thai	อักษรไทย

RZ-LGR-2 Content

- ⦿ Normative
 - One XML per script (element XML)
 - A common XML file
- ⦿ Informative
 - Overview document (PDF)
 - HTML representation of the 7 XML files
 - Repertoire table (PDF)
 - Example: Thai

Thai

	0E0	0E1	0E2	0E3	0E4	0E5	0E6	0E7
0		๐	๑	๒	๓	๔		
1	๕	๖	๗	๘	๙	๐		
2	๑	๒	๓	๔	๕	๖		
3	๗	๘	๙	๐	๑	๒		
4	๓	๔	๕	๖	๗	๘		
5	๙	๐	๑	๒	๓	๔		
6	๕	๖	๗	๘	๙	๐		
7	๑	๒	๓	๔	๕	๖		
8	๗	๘	๙	๐	๑	๒		
9	๓	๔	๕	๖	๗	๘		
A	๙	๐	๑	๒	๓	๔		
B	๕	๖	๗		๘	๙		
C	๐	๑	๒		๓			
D	๔	๕	๖		๗			
E	๘	๙	๐		๑			
F	๒	๓	๔	๕	๖	๗		

List of XML Files

- All files in <https://www.icann.org/sites/default/files/lgr/>

Script	File name in URL
Common	lgr-2-common-26jul17-en.xml
Arabic	lgr-2-arabic-script-26jul17-en.xml
Ethiopic	lgr-2-ethiopic-script-26jul17-en.xml
Georgian	lgr-2-georgian-script-26jul17-en.xml
Khmer	lgr-2-khmer-script-26jul17-en.xml
Lao	lgr-2-lao-script-26jul17-en.xml
Thai	lgr-2-thai-script-26jul17-en.xml

- Also provided, a mechanically generated and non-normative HTML presentation using same file name with .html extension (example: <https://www.icann.org/sites/default/files/lgr/lgr-2-arabic-script-26jul17-en.html>)

How Was It Created?

- ⊙ Script based XML files generated by the Generation Panels
- ⊙ These files reviewed and integrated by the Integration Panel
- ⊙ Common file created by IP is the cumulative set of all integrated LGRs for:
 - Repertoire
 - Variants all in 'blocked' type
 - Character classes (renamed to avoid collision)
 - WLE rules (also renamed)
 - Actions

How to Use It?

- ⦿ Covered in another presentation

LGR-2 Is Here, What Can You Do With It?

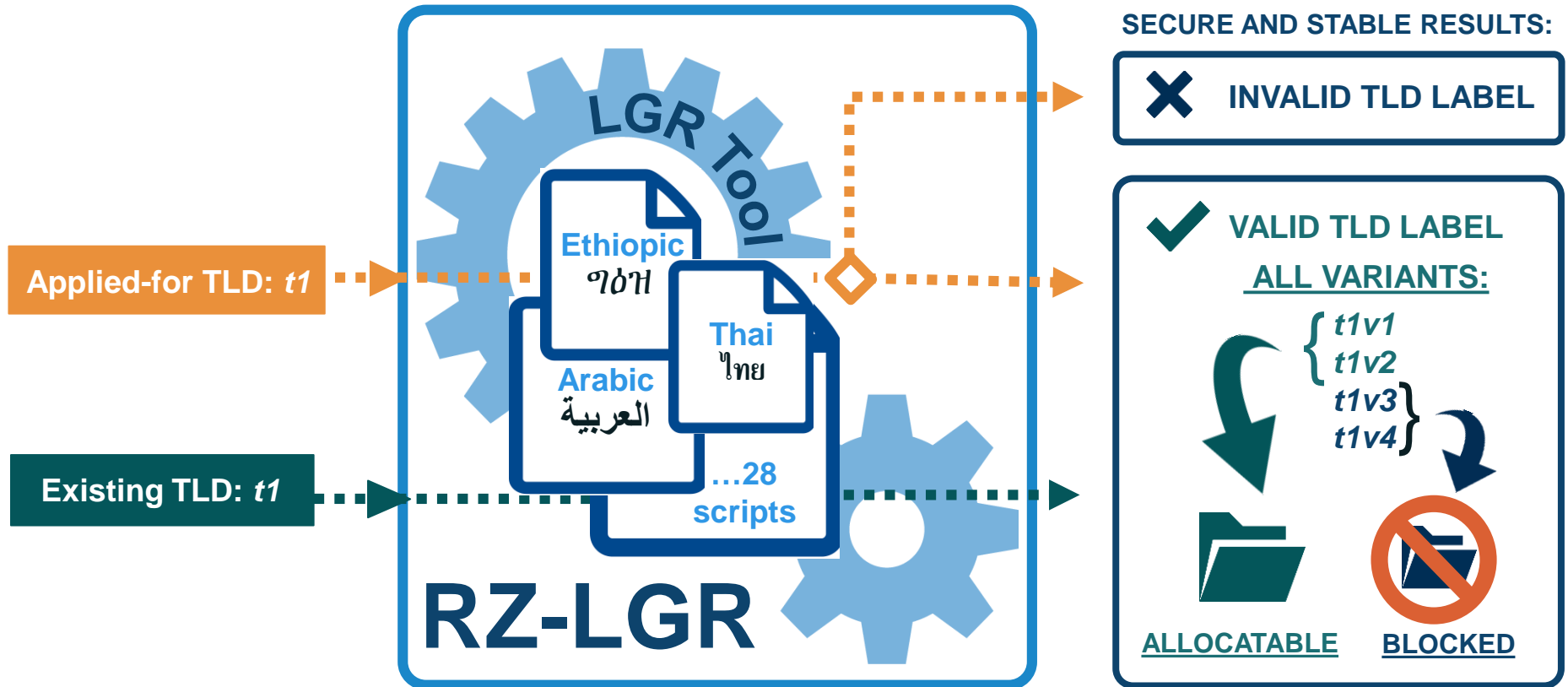
Thank You and Questions

- ⦿ Root Zone Label Generation Rules (RZ-LGR-1 and RZ-LGR-2)
 - <https://www.icann.org/resources/pages/root-zone-lgr-2015-06-21-en>
- ⦿ ICANN IDN Document Repository
 - <https://community.icann.org/display/croscomlgrprocedure/Document+Repository>
- ⦿ RFC 7940 Representing Label Generation Rulesets Using XML
 - <https://www.rfc-editor.org/info/rfc7940>
- ⦿ RFC 8228 Guidance on Designing Label Generation Rulesets (LGRs) Supporting Variant Labels
 - <https://www.rfc-editor.org/rfc/rfc8228.txt>

How to Use RZ-LGR-2

Michel Suignard
Integration Panel

RZ-LGR-2 Is Here, What Can You Do with It?



RZ-LGR-2 Is Here, What Can You Do with It?

- ⊙ The Root Zone LGR is split into
 - One **Element LGR** per script (N files)
 - A single **Merged LGR** (1 file)
 - Each of these files in XML (normative) and HTML (informative)
 - Plus documentation files (Overview and Code Tables)

- ⊙ Three tasks in using the LGR in applying for a label
 - Validate
 - Generate allocatable variants
 - Check for collisions with delegated labels and their variants

- ⊙ Walking through these tasks will explain why (N+1 files are used)

Applying for a Label

- ⦿ Each label in the Root Zone is in a single script and the application defines which script a label is applied for
- 1. Select script based **Element LGR** corresponding to application
- 2. Use Element LGR to validate label
 - check code points in label against repertoire
 - check context constraints on code points
 - check whole label constraints
- 3. Generate allocatable variant labels (if any)

Note: Element LGRs are based on the submitted script LGRs

- Lightly edited for consistency and common conventions
- Guaranteed match results of Script LGR proposals

Applying for a Label (cont.)

4. Check for collisions
 - Use **Merged** file (aka. Common)
 - Merged file contains all non-reflexive variant mappings
 - All variant mapping types are set to “blocked” as needed for collision checking.
 - Variants mappings are symmetric and transitive:
 - Any label or variant label is part of only one variant label set
 - In each set, all labels are variants of each other
 - Calculate index variant
 - Map each code point to its smallest variant code point
 - Compare to index variants for delegated labels
 - If index variants match, either the labels or their variants collide

Example

- ◉ Example: existing label for TLD "همراه"
- ◉ Script: Arab
- ◉ Process with Element-LGR: [lgr-2-arabic-script-26jul17-en.xml](#)
 - ◉ Details of steps in evaluation are described in RFC 7940

- ◉ This label has

- ◉ 1 original label "همراه" (0647 0645 0631 0627 0647)
- ◉ 1 allocatable variant "ہمراه" (06C1 0645 0631 0627 06C1)
- ◉ 268 blocked variants
- ◉ 50 invalid variants

- For example mixed use of 0647 vs. 06C1:

(0647 0645 0631 0627 06C1)

Other Tasks

- ⦿ Other tasks for Common LGR
 - Verification against script LGRs
 - merged file contains all variant mappings
 - merged file contains all context/WLE rules

Notes:

- ⦿ Merged file is derived from script LGR files
- ⦿ Differences from script LGR files
 - Certain items (tags, classes, rules) renamed to avoid collisions
 - Comments/descriptions edited for consistency
 - References point to Element LGRs not source documents
- ⦿ N+1 format (scripts + merged) could be useful for any other zones that support multiple scripts in parallel

Other Files

- ⦿ Overview describes
 - How LGR is laid out
 - Review of submissions that were integrated
 - How to use the LGR
 - The main features for each script LGR (summary)

- ⦿ Code Tables give visual overview of repertoire in relation to MSR-2 and set of IDNA2008 PVALID code points

- ⦿ The informative HTML versions of the Element and Merged LGR
 - Easier to read for human reviewers
 - Provide some calculated info (counts, etc.)
 - Show glyph shapes for code points, Unicode names, etc.

What Should GPs Note for Future RZ-LGRs?

- ⦿ Script LGR proposals are archived
 - Sole documentation of design decision behind LGR
 - Make sure all decisions are documented and cite references
- ⦿ Script LGRs are input to Element LGRs
 - Following common templates and conventions
 - Section numbering, comment convention for XML
 - Conventions for organizing <description>
 - Minimizes need to adjust these
- ⦿ Provide required out-of-repertoire variants
 - Any cross-script homoglyphs
 - Any in-repertoire variants

Thank You and Questions

- ⦿ Packaging the MSR and RZ-LGR

<https://community.icann.org/download/attachments/43989034/Packaging-the-MSR-and-LGR-2017-09-15.pdf>

- ⦿ RFC7940

<https://tools.ietf.org/rfc/rfc7940>

- ⦿ IDN TLD portal (for RZ-LGR-2 files)

<https://www.icann.org/resources/pages/root-zone-lgr-2015-06-21-en>

LGR Toolset Update

Audric Schiltnknecht, Julien Bernard, Marc Blanchet
Viagénie

Contents

- ⦿ LGR Toolset summary in one slide
- ⦿ New features in 2017
 - Sets of LGR files
 - Label validation
 - HTML export
 - Interface improvements

LGR Toolset Summary

- ⊙ Toolset to
 - Create, update, use Label Generation Rules
 - Validate labels, generate variants, verify collisions

- ⊙ Available(*)
 - Opensource
 - Online as a service

- ⊙ As
 - Cmdline and librairies in python
 - Web interface

<https://www.icann.org/resources/pages/lgr-toolset-2015-06-21-en>

- ⊙ Implements the Root Zone LGR structure as a set of LGRs(*)
 - Repertoire: cumulative repertoire of all the Element LGRs
 - Variants: union of the variant mappings from all the Element LGRs, with « blocked » type
 - Classes: union of the character classes from all the Element LGRs. Name are prefixed by script of Element LGR
 - WLE: cumulative set of Whole Label Evaluation rules and actions for all Element LGRs. Name are prefixed by script of Element LGR

* As described in Root Zone Label Generation Rules — RZ-LGR-2

LGR Sets - Interface

Import existing LGR

To import LGR set, select the XML files composing the LGR set. The set will automatically be created.

Select file(s)

2 files selected.
If you select more than one file, this will create a LGR set

Validating repertoire

Code points will be limited to the selected repertoire

LGR set name

The name of the set

⚠ Note that importing large LGR files may take significant time to load on your browser.



LGR tool / root-zone - Root Zone LGR

Code point	Character Name	Comments	Action
-10D0 (ⴀ) 0 Variant(s)	GEORGIAN LETTER AN	Georgian	<input type="button" value="See code"/>
-10D1 (ⴁ) 0 Variant(s)	GEORGIAN LETTER BAN	Georgian	<input type="button" value="See code"/>
-10D2 (ⴂ) 0 Variant(s)	GEORGIAN LETTER GAN	Georgian	<input type="button" value="See code"/>
-10D3 (ⴃ) 0 Variant(s)	GEORGIAN LETTER DON	Georgian	<input type="button" value="See code"/>

LGR Sets – Interface (cont.)

🏠 LGR Tool / root-zone - Root Zone LGR

📁 Import 📄 New ⚙️ Tools ▾

✓ Validate label 📄 Summary 👁️ View XML 🌐 HTML Output ⬇️ Download

Code points References Meta data Rules Embedded LGRs

List of LGRs in the set

- lgr-2-georgian-script-01jun17-en
- lgr-2-khmer-script-01jun17-en

English (en) ▾ Go

©

🏠 LGR Tool / root-zone / lgr-2-georgian-script-01jun17-en - Root Zone LGR

📁 Import 📄 New ⚙️ Tools ▾

✓ Validate label 📄 Summary 👁️ View XML 🌐 HTML Output ⬇️ Download

Code points References Meta data Rules

Code point	Character Name	Comments	Action
U+10D0 (ⴀ) 0 Variant(s)	GEORGIAN LETTER AN	Georgian	See code point
U+10D1 (ⴁ) 0 Variant(s)	GEORGIAN LETTER BAN	Georgian	See code point
U+10D2 (ⴂ) 0 Variant(s)	GEORGIAN LETTER GAN	Georgian	See code point
U+10D3 (ⴃ) 0 Variant(s)	GEORGIAN LETTER DON	Georgian	See code point

Label Validation on a LGR Set

The screenshot shows the 'LGR Tool / khmer-georgian - Root Zone LGR' interface. A 'Validate label' dialog box is open, featuring a 'Label' input field and a 'Validate' button. Below the input field, it states 'Maximum length: 63 code points'. The dialog contains two main sections, both highlighted with red boxes:

- Allocated Set labels:** Includes a 'Browse...' button and the text 'No file selected.' Below this, it explains: 'Labels already allocated in the LGR set, that will be used to check for collisions when evaluating labels on the LGR set.'
- Script:** A dropdown menu is open, showing 'und-Geor' (highlighted in red) and 'und-Khmr'.

On the left side of the slide, there are two text labels with arrows pointing to the dialog:

- 'Delegated labels' points to the 'Allocated Set labels' section.
- 'Scripts from Elements LGR' points to the 'Script' dropdown menu.

LGR Validation - Process

- ⦿ Element LGR is retrieved from selected script
- ⦿ Label is validated against this Element LGR
- ⦿ Collisions are checked using the Common-Merged LGR between
 - Input label (and its variants)
 - List of delegated labels (and their variants)
- ⦿ If no collision, then generates variants using Element LGR

LGR Validation – Delegatable Use Case

Label valid in
Element LGR
(Khmer)

No collisions with
delegated labels

Validate label

Label

Maximum length: 63 code points

Allocated Set labels No file selected.
Labels already allocated in the LGR set, that will be used to check for collisions when evaluating labels on the LGR set

Script
The script used to validate the label

✓ VALID

U-label	Disposition	Code point sequence	
ចេងណាំ xn-- m2eb2a2m0b7a	valid	U+1785 (ច) U+17C1 (ាំ) U+1784 (ង) U+178E (ណ) U+17B6 (ាំ) U+17C6 (ាំ)	<input type="button" value="Show / hide rule"/>

✓ No collision

Collision

No collision.

Variant labels (including original as last)

LGR Validation – Non-Delegatable Use Case

Validate label

Label

Maximum length: 63 code points

Allocated Set labels
Labels already allocated in the LGR set, that will be used to check for collisions when evaluating labels on the LGR set

Script
The script used to validate the label

✓ VALID

U-label	Disposition	Code point sequence	
ចេតនា xn-- m2eb2a2m0b7a	valid	U+1785 (ច) U+17C1 (ន) U+1784 (ា) U+178E (ណ) U+17B6 (រ) U+17C6 (័)	<input type="button" value="Show / hide rule"/>

✗ Collision(s) found

Collision

The label is in the LGR set labels.

Variant labels

No variant labels generated.

Label valid in Element LGR (Khmer)

Collide with one of the delegated labels

LGR Sets - Tools

- ⦿ Tools have been updated to handle LGR Sets (where applicable)
- ⦿ New « Allocated Set labels » parameter: (optional) list of labels allocated in the given set. Used e.g. to check collisions
- ⦿ New parameter for label disposition: « Script » of the label(s)
- ⦿ New tool: cross-script variants

HTML Export

- ⦿ Create a human-readable static version of the LGR
- ⦿ Support LGR Sets
- ⦿ Accessible from web interface or as a standalone cmdline (but still need some configuration)

HTML Export - Metadata

Lgr-2-Arabic-Script-2017-02-23-En

LGR Version	2
Date	2017-02-24
Language(s)	und-Arab
Scope(s)	domain: .
Unicode Version	6.3.0

This document is mechanically formatted from the XML file for the LGR. It provides additional summary data and explanatory text. The XML file remains the sole normative specification of the LGR.

Table of Contents

- 1 Description
- 2 Repertoire
- 3 Variant Sets
- 4 Classes, Rules and Actions
 - 4.1 Character Classes
 - 4.2 Whole label evaluation and context rules
 - 4.3 Actions
- 5 Table of References

Unicode version: 6.3.0

Description

```
<h1>Root Zone Label Generation Ruleset for the Arabic Script</h1>
<h2>Overview</h2>
<p>This file contains a set of Label Generation Rules (LGR) for Arabic as would be appropriate for the Root Zone.
For more details on this LGR and its development see
TF-AIDN, "Proposal for Arabic Script Root Zone LGR", Version 3.4, 2015 November 18 [Proposal].</p>
<h2>Repertoire</h2>
<p>The repertoire is based on Section 3.2 in [Proposal] and
only includes code points used by languages that are actively written in the Arabic script. It
excludes code points for which TF-AIDN was unable to find sufficient evidence of use (see Appendix E in [Proposal]). The repertoire is
```

Description type: text/html

Description

Root Zone Label Generation Ruleset for the Arabic Script

Overview

This file contains a set of Label Generation Rules (LGR) for Arabic as would be appropriate for the Root Zone. For more details on this LGR and its development see TF-AIDN, "Proposal for Arabic Script Root Zone LGR", Version 3.4, 2015 November 18 [Proposal].

Repertoire

The repertoire is based on Section 3.2 in [Proposal] and only includes code points used by languages that are actively written in the Arabic script. It excludes code points for which TF-AIDN was unable to find sufficient evidence of use (see Appendix F in [Proposal]). The repertoire is based on [MSR-2], which is a subset of Unicode 6.3 [Unicode 6.3].

HTML Export - Repertoire

Repertoire

Summary

Number of elements in repertoire	385
Number of ranges in repertoire	0
Number of code point sequences	7

Repertoire by Code Point

The following table lists the repertoire by code point (or code point sequence). The data in the Script and Name columns are extracted from the Unicode character database. Where the comment in the original LGR is equal to the character name, it has been suppressed.

For any code point or sequence for which a variant is defined, the link to the associated variant set, or if mapped to itself, the variant type of that mapping is provided in the Variants column.

#	Code Point	Glyph	Script	Name	Tags	Required Context	Variants	Comment	References
1	U+0061	a	Latin	LATIN SMALL LETTER A	Latin				[0]
2	U+0062	b	Latin	LATIN SMALL LETTER B	Latin				[0]
3	U+0063	c	Latin	LATIN SMALL LETTER C	Latin				[0]

Link to WLE

Link to variant set

382	U+17D0	័	Khmer	KHMER SIGN SAMYOK SANNYA	Khmer,und-Khmr-sign,sc:Khmr	when: und-Khmr-follows-consonant-shifter			[2], [207], [208], [209]
383	U+17D2	៑	Khmer	KHMER SIGN COENG	Khmer,sc:Khmr,und-Khmr-coeng	when: und-Khmr-coeng-context			[2], [4]
384	U+17D2 U+178A	៑ ឝ	Khmer	KHMER SIGN COENG KHMER LETTER DA		when: und-Khmr-follows-consonant	set 18	KHMER CONSONANT SIGN COENG DA	[2]
385	U+17D2 U+178F	៑ ឝ	Khmer	KHMER SIGN COENG KHMER LETTER TA		when: und-Khmr-follows-consonant	set 18	KHMER CONSONANT SIGN COENG TA	[2]

HTML Export – Variant Sets

Variant Sets

Summary

Number of variant sets	26
Largest variant set	8
Ordinary Variants by Type	blocked (243)

The following tables list each pair of variant mappings on one row.

In a properly specified LGR, all members of each variant set are variants of each other, a property called transitivity. Because of that, all variant sets are necessarily disjoint. In each set, shading is used to group mappings from the same source code point or sequence.

Variant Set 1 — 3 Members

#	Source	Glyph	Target	Glyph	Type(s)	References	Comment
1	U+0067	g	U+0067	g	blocked		Identity
2	U+0067	g	U+0581	g	blocked		Homoglyph
3	U+0581	g	U+0067	g	blocked		Latin small letter G

Variant Set 2 — 8 Members

#	Source	Glyph	Target	Glyph	Type(s)	References	Comment
1	U+0068	h	U+0068	h	blocked		Identity
2	U+0068	h	U+04BB	h	blocked		Homoglyph
3	U+0068	h	U+0570	h	blocked		Homoglyph
4	U+04BB	h	U+0068	h	blocked		Homoglyph
5	U+04BB	h	U+04BB	h	blocked		Identity
6	U+0570	h	U+0068	h	blocked		Homoglyph
7	U+0570	h	U+04BB	h	blocked		Homoglyph
8	U+0570	h	U+0570	h	blocked		Identity

HTML Export – Classes

Classes, Rules and Actions

Character Classes

Computed members

The following table lists all top-level classes with their definition and the regular expression defining their members.

Name	Definition	Count	Members	References	Comment
Latin		181	U+0067 U+0068 U+006E U+006F U+0071 U+0075 U+0269		
Greek		77	U+03B7 U+03B9 U+03BF		
Cyrillic		80	U+043E U+0448 U+04BB		
Armenian		53	U+0561-U+0586		
sc:Armn		53	U+0561-U+0586		
Arabic		1214	U+0620-U+063A U+0641-U+064A U+0672 U+0679-U+0681 U+0683-U+068F U+0691 U+0693 U+0695-U+069A U+069F U+06A0 U+06A2 U+06A4 U+06A6-U+06AB U+06AD-U+06B1 U+06B3 U+06B5 U+06BA-U+06BE U+06C0-U+06C4 U+06C6 U+06CB-U+06D2 U+06D5 U+0751 U+0752 U+0756 U+0760 U+0762 U+0763 U+0766-U+0768 U+076A U+076E-U+0771 U+08A0 U+08A2- U+08A9		
			U+0620-U+063A U+0641-U+064A U+0672 U+0679-U+0681 U+0683-U+068F U+0691 U+0693 U+0695-U+069A U+069F U+06A0 U+06A2 U+06A4 U+06A6-U+06AB U+06AD-U+06B1 U+06B3		

und-Thai-c-av-bv		198	([:und-Thai-consonant:]U[:und-Thai-above-vowel:]U[:und-Thai-below-vowel:])		
und-Thai-ct		117	([:und-Thai-consonant:]U[:und-Thai-tone:])		
und-Thai-ctaa		176	([:und-Thai-consonant:]U[:und-Thai-tone:]U[:und-Thai-sara-aa:])		

Class combination

Legend

Members or Ranges: Lists the members of the class as code points (xxx) or as ranges of code points (xxx-yyy). Any class too numerous to list in full is elided with "...".

Tag=ttt: An anonymous class implicitly defined based on tag value.

[::] - named character set: Reference to a named character set [:name:].

(n, u, \, \Delta) - set operators: Sets may be combined by set operators (n = intersection, u = union, \ = difference, \Delta = symmetric difference).

HTML Export – WLE

Whole label evaluation and context rules

The following table lists all the top-level, or named rules defined in the LGR and indicates whether they are used as trigger in an action or as context (when or not-when) for a code point. (Any use of context rules for variants is not indicated).

Regex used to evaluate rule

Name	Regular Expression	Used as Trigger	Used as Context	Anchor	References	Comment
Common-leading-combining-mark	(start) ([[:class property:gc=Mn:]]U[:class property:gc=Mc:])	True	False	False		default WLE rule matching labels with leading combining marks ☼
und-Arab-no-mix-kaf-keheh	(U+0643 (any){0,} U+06A9 U+06A9 (any){0,} U+0643)	True	False	False	[100]	WLE Rule 1: do not mix Arabic letters KAF and KEHEH in the same label
und-Arab-no-mix-kaf-swash	(U+0643 (any){0,} U+06AA U+06AA (any){0,} U+0643)	True	False	False	[100]	WLE Rule 2: do not mix Arabic letters KAF and SWASH KAF in the same label
und-Arab-no-mix-alef-maksura-farsi-yeh	(U+0649 (any){0,} U+06CC U+06CC (any){0,} U+0649)	True	False	False	[100]	WLE Rule 3: do not mix Arabic letters ALEF MAKSURA and FARSI YEH in the same label
und-Arab-no-mix-heh-goal	(U+0647 (any){0,} U+06C1 U+06C1 (any){0,} U+0647)	True	False	False	[100]	WLE Rule 4: do not mix Arabic letters HEH and HEH GOAL in the same label
und-Arab-no-mix-heh-goal-ae	(U+06C1 (any){0,} U+06D5 U+06D5 (any){0,} U+06C1)	True	False	False	[100]	WLE Rule 5: do not mix Arabic letters HEH GOAL and AE in the same label
und-Arab-no-mix-heh-ae	(U+0647 (any){0,} U+06D5 U+06D5 (any){0,} U+0647)	True	False	False	[100]	WLE Rule 6: do not mix Arabic letters HEH and AE in the same label
und-Arab-no-mix-heh-doachashmee	(U+0647 (any){0,} U+06BE U+06BE (any){0,} U+0647)	True	False	False	[100]	WLE Rule 7: do not mix Arabic letters HEH and HEH DOACHASHMEE in the same label
und-Arab-no-mix-teh-marbuta-goal	(U+0629 (any){0,} U+06C3 U+06C3 (any){0,} U+0629)	True	False	False	[100]	WLE Rule 8: do not mix Arabic letters TEH MARBUTA and FEH WITH DOT MOVED BELOW in the same label
und-Arab-no-mix-noon-with-three-dots-above-yeh-with-three-dots-below	(U+06BD (any){0,} U+06D1 U+06D1 (any){0,} U+06BD)	True	False	False	[100]	WLE Rule 9: do not mix Arabic letters NOON WITH THREE DOTS ABOVE and YEH WITH THREE DOTS BELOW in the same label

Interface Improvements



Welcome to the LGR (Label Generation Ruleset) Tool

This application provides a convenient interface for browsing and editing [LGR's](#) conforming to the [Representing Label Generation Rulesets using XML](#) specification.

Previously loaded LGR file(s)

Previously, you edited the following LGR file(s). Click on its title to resume your editing session.

LGRs

- [View lgr-2-arabic-script-2017-02-23-en](#)

LGR sets

- [View rootzone](#) [View embedded LGRs](#)

List of loaded LGR and LGR Sets

Your saved results

The following files contains your tools computation results.

▲ Note that these files could be cleaned up regularly.

- [Download 20171004_202015_annotation_sample-french.txt.gz](#)
- [Download 20171004_202010_annotation_sample-french.txt.gz](#)

Start a new LGR file or import an existing one

Import an existing XML file

▲ Note that importing large LGR files may take significant time to load on your browser.

Start with a New blank XML file

Start from a built-in LGR

The following LGRs are pre-installed in the system. You may use them as a starting point for your own LGR. To do so, just click on it to make a copy that you can then edit.

- [Open Sample-French](#)

Remember to save your work regularly by downloading a copy of the XML file.

Please send any feedback to support@viagenie.ca.

Regroup create/import buttons

Tool outputs (hidden if none)

English (en)

©

Thank You and Questions

- ⦿ RFC 7940 Representing Label Generation Rulesets Using XML
 - <https://www.rfc-editor.org/info/rfc7940>
- ⦿ LGR Toolset
 - <https://lgrtool.icann.org>
- ⦿ LGR Toolset User Manual
 - <https://www.icann.org/resources/pages/lgr-toolset-2015-06-21-en>

Chinese GP Update

Kenny HUANG
Wei WANG

Agenda

1

CGP Work Review

2

CGP Proposal Draft
201707

3

CJK Coordination

4

Repertoire
Size

5

Allocatable Label
Number

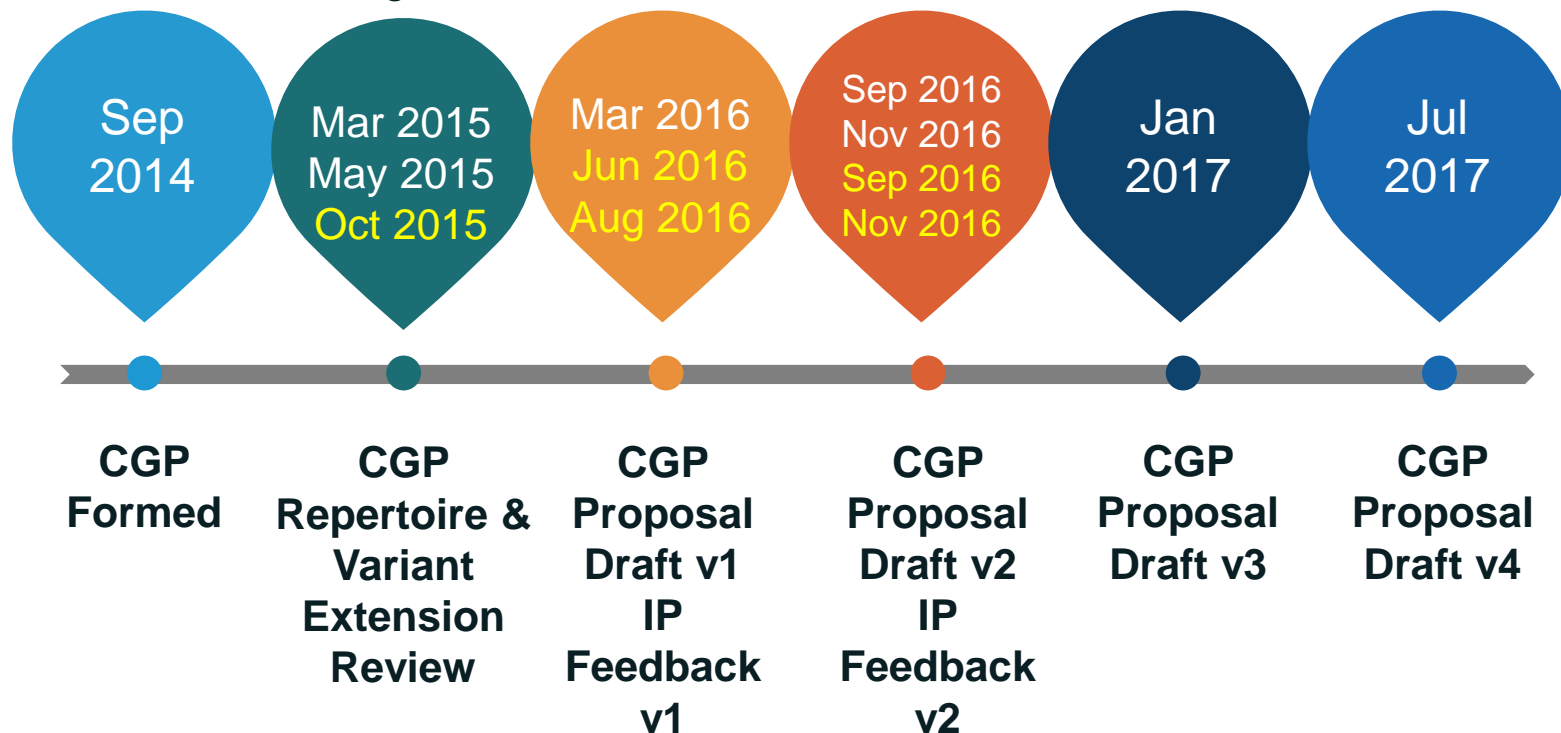
6

Next Step &
Suggestions

CJK Meeting@Beijing

CJ Meeting@Dallas
CJK Meeting@Seoul

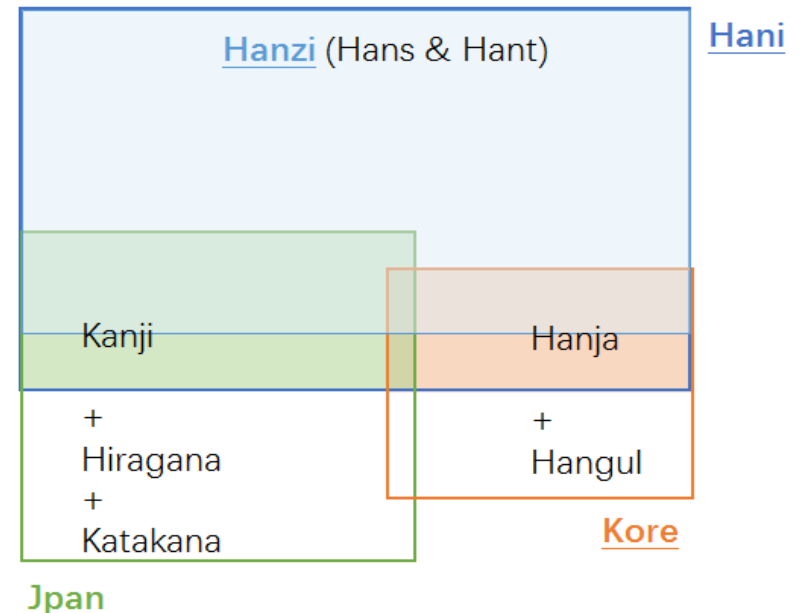
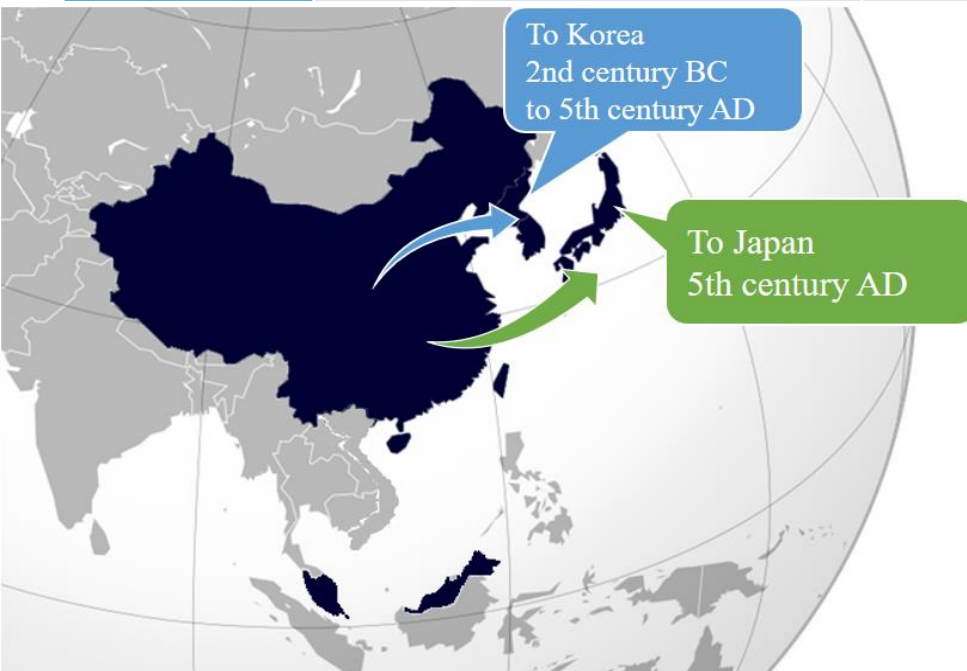
CK Meeting@Taipei
CK Meeting@Seoul



Next Steps

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



Team and Work Process

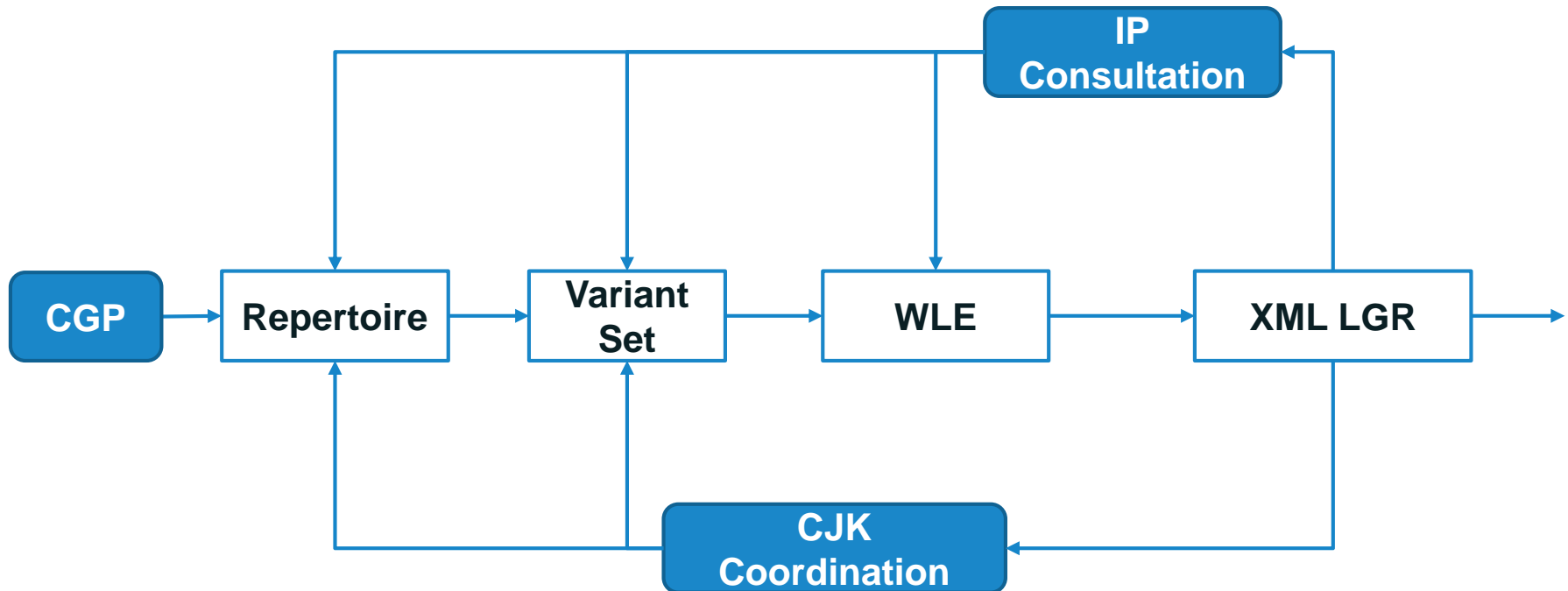
Members, 23 experts from 10 countries/regions

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

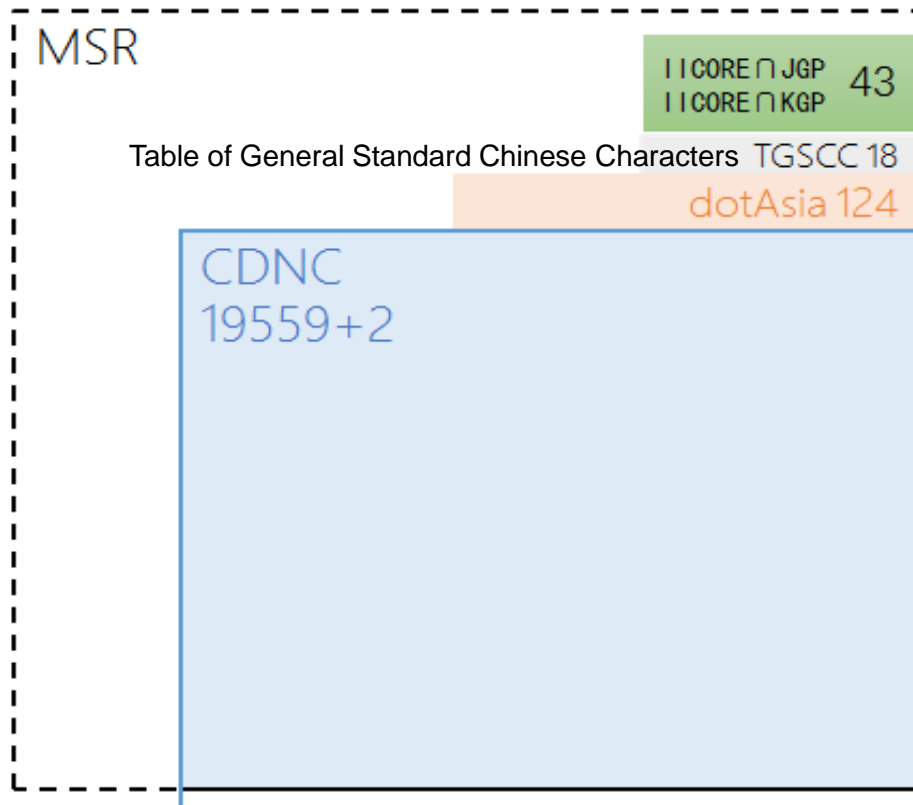
Advisor, Edmon CHUNG

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

CJK coordination working group

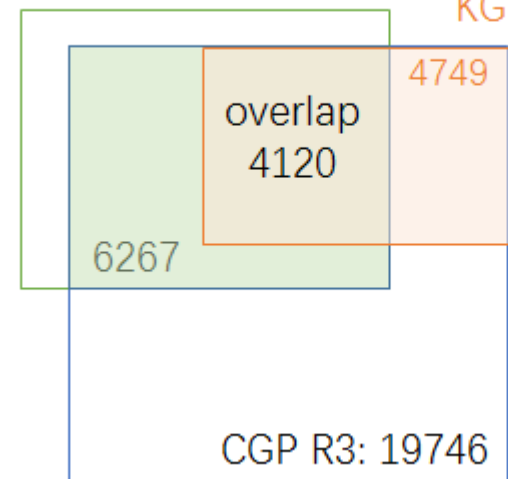


- ⊙ Code Point Repertoire
- ⊙ 19,746 Characters/Code Points



JGP Kanji: 6356

KGP Hanja: 4758



Note: two chars not included in current MSR

- Code Point Variants

- Definition

- “characters with different visual forms but with the same pronunciations and with the same meanings as the corresponding official forms in the given language contexts.”

- Simplified and Traditional

- Every code point in the CGP repertoire has its preferred simplified variant(s), preferred traditional variant(s), and reserved variant(s)
- A code point might have a reflexive preferred S/T variant
- A code point might have no reserved variant

Sub-Type	Type	Comment
“simp”	Allocatable	preferred simplified variant char;
“r-simp”	Allocatable	reflexive preferred simplified variant char;
“trad”	Allocatable	preferred traditional variant char
“r-trad”	Allocatable	reflexive preferred traditional variant char
“both”	Allocatable	preferred simplified and traditional variant chars are the same
“r-both”	Allocatable	reflexive preferred simp and trad variant chars are the same
“r-neither”	Blocked	Non-allocatable reflexive/original char
“blocked”	Blocked	Non-allocatable variant char

⦿ Whole Label Generation Rules

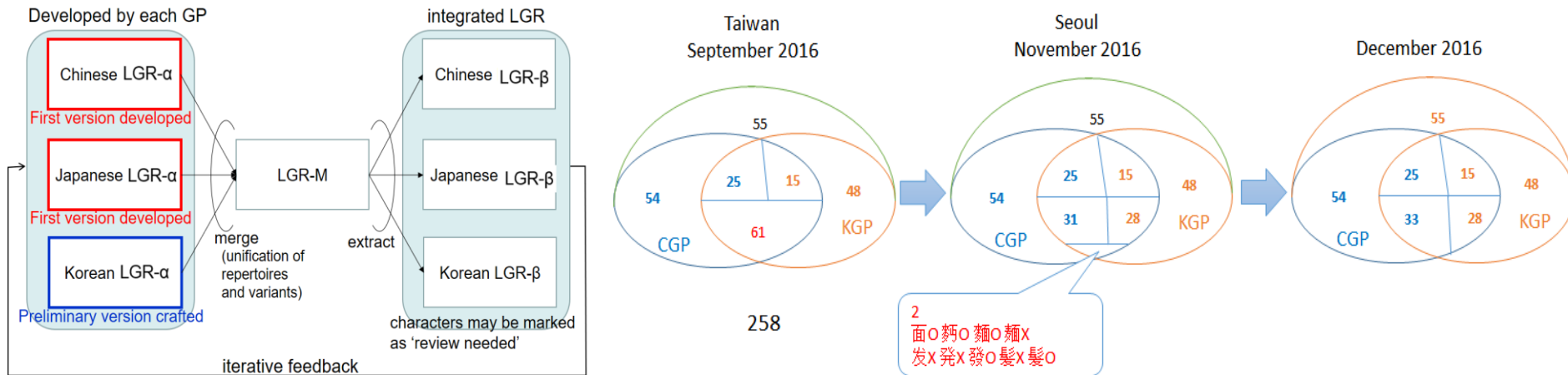
```
<rules>
<!--Action elements - order defines precedence-->
<action disp="invalid" match="leading-combining-mark" comment="labels with leading combining
marks are invalid" />
<action disp="blocked" any-variant="blocked" comment="default action for blocked variant"/>

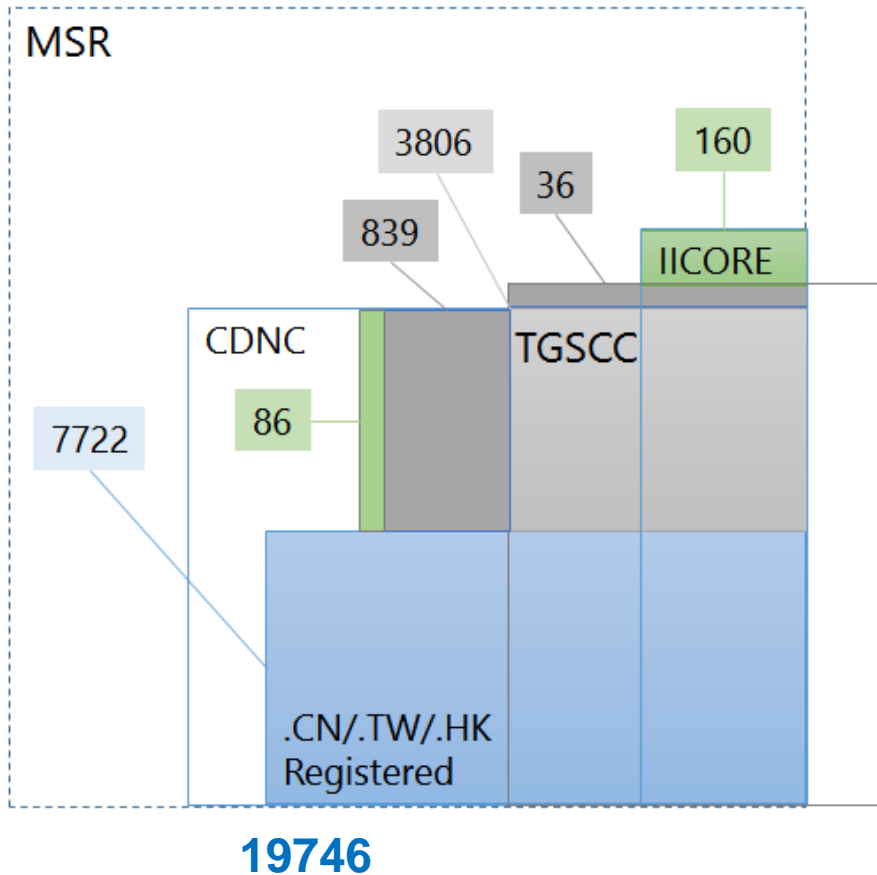
<action disp="allocatable" only-variants="simp r-simp both r-both" comment="simplified label" />
<action disp="allocatable" only-variants="trad r-trad both r-both" comment="traditional label"/>
<action disp="allocatable" only-variants="r-simp r-trad r-both r-neither" comment="original label"/>

<action disp="blocked" any-variant="simp trad both r-simp r-trad r-both r-neither" comment="block
any other mixed labels" />

<action disp="allocatable" comment="catch-all" />
</rules>
```

- ⊙ Coordination within CGP
 - CDNC variants
 - TGSCC and IICORE variants review (172)
 - dotAsia variant coordination (69)
- ⊙ Coordination between **C**, **J** and **K**
 - 445 variant mappings (146 variant groups)





Dictionary

- 16th-11th century BC 商甲骨文 3,500 ~ 4,500
- 202 BC-220 AD 汉朝 说文解字 9,353
- 960-1279 AD 宋朝 类篇 31,319
- 1710 清朝 康熙字典 47,035
- 1959 Japanese 大汉和辞典 49,964
- 1994 China Mainland 中华字海 87,019
- 2004 Taiwan 異體字字典 106,230

Paper

- 2007, A Survey on the Usage of Chinese Characters and Phrases in the Newspapers, Radio, TV, and Web, [8128](#)
- 2010, survey on Chinese Weblog Wording, [20923](#)

Standard

- BIG5 13,053
- CNS11643 76,067
- GBK 21,886
- GB18030 70,244
- Unicode 80,388

Legislation & Regulation

- China's Ministry of Civil Affairs issued Notification 2016[33], requiring government departments to update the naming related information system in public service and administration areas, to cover the characters in national standard GB13000 (20902 chars) or GB18030 (70244 chars)

5 The Issue of Allocatable Label Number

In the CGP variant mapping table (Appendix I), for all 19746 characters, there are 3 characters with 2 PSVs, 127 with 2 PTVs, 5 with 3 PTVs and 1 with 4 PTVs, which means, all together 136 characters have multiple preferred variant characters. These 136 characters will generate multiple all-simplified labels or all-traditional labels, which will lead to an **over-production of allocatable labels** issue at the root zone level.

Original Char	Allocatable Simplified Variant	Allocatable Traditional Variant	Blocked Variant
A	B	C	
D	E	F	G
H	I, J	K	
L	M, N	O	P

Input: AD
Allocatable: BE; CF
Blocked: BF, CE

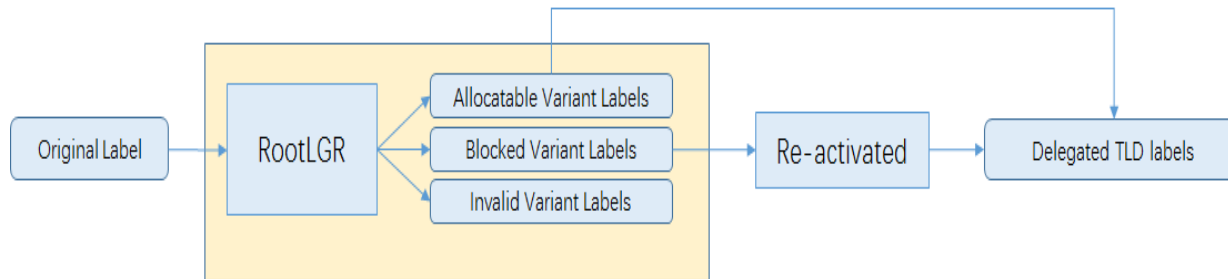
Input: AH
Allocatable: BI, CJ; CK
Blocked: BK, CI, CJ

Input: HL
Allocatable: IM, IN, JM, JN; KO
Blocked:

CGP tend to keep the **multiple variant mappings** under the current LGR framework.

Instead

CGP would propose to introduce new **types** and **rules** to **reduce the number of allocatable labels**.



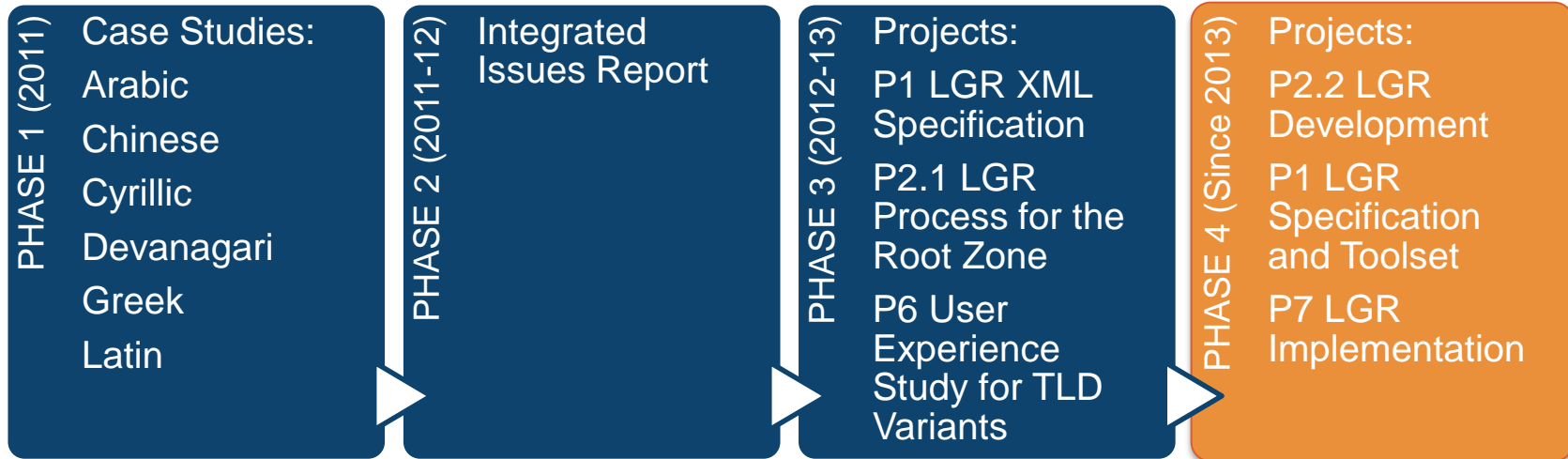
Original Char	Allocatable Simplified Variant	Allocatable Traditional Variant	Blocked Variant
A	B	C	
D	E	F	G
H	I, J	K	
L	M, N	O	P



Variants imported from J & K: 43 characters

Unicode	Char		All Variant Chars
967A	險	J	硷(7877)碱(78B1)硷(7906)险(9669)險(967A)險(96AA)鹼(9E7C)
7E4A	織	J	織(5B45)緯(7E34)織(7E4A)織(7E8E)織(7E96)纤(7EA4)
9421	鐵	J	鉄(9244)鋈(9295)鐵(9421)鐵(9435)铁(94C1)
9D8F	鷄	J	雞(96DE)鸡(9CEE)鷄(9D8F)鷄(9DC4)鸡(9E21)
4FAD	俛	J	俛(4FAD)儘(5118)尽(5C3D)盡(76E1)
6442	撰	J	捫(6315)撰(6442)撮(6444)撮(651D)
685F	棧	J	栈(6808)棧(685F)棧(68E7)輶(8F4F)
7E4B	繫	J	繫(7E4B)繫(7E6B)
81D3	臟	J	脏(810F)臟(81D3)臟(81DF)髒(9AD2)
8217	舖	J	舖(8216)舖(8217)舖(92EA)铺(94FA)
9039	達	J	达(8FBE)达(8FD6)達(9039)達(9054)
9271	鉍	J	矿(77FF)礦(7926)鉍(9271)鑛(945B)
3960	悞	K	悞(3960)誦(8ADD)誦(8C1E)
784F	研	K	擘(63C5)研(7814)研(784F)
663B	昂	K	昂(6602)昂(663B)
7A36	穢	K	穢(7A22)穢(7A36)

To ICANN



There should be more interaction between P2.2, P1 and P7.
 The generation rule, algorithm and system design will never be able to replace rational thought and case analysis from Applicant Evaluation Panel.
 P2.2 might not be able to fix all issues related to IDN variant.
 P7 are wanted to join the discussion together with GPs and IP.

Thank You All

- Integration Panel
- Edmon Chung
- Sarmad Hussain
- ICANN Singapore

Q&A

Japanese GP Update

Hiro Hotta

JapaneseGP (JGP) update

1 November, 2017

Hiro Hotta <hotta@jprs.co.jp>

Development from March 2017 is written in red

What JGP (Japanese Generation Panel) does

- Mandate
 - Proposing LGR for TLDs of Japanese language/scripts
 - that can co-exist with LGRs for other languages/scripts
- Steps
 - Step1 : Populate JGP with diverse experts
 - Step2 : Define the requirements and basic framework of Japanese LGR based on the expertise and experience of Japanese IDNs
 - Step3 : Coordinate with other language Generation Panels especially of those languages interrelated with Japanese
 - Step4 : Finalize LGR following necessary consultation with IP and Japanese community

What JGP should care about

- Consultation with the local community
 - Post JGP discussions/activities on JGP web-site
 - Presentation and discussion about LGR
 - with Japan Trademark Association
 - IGCJ (Internet Governance Conference Japan) events
 - Public comment forum
- Coordination among CJK communities
- Coordination with IP
- Coordination with global communities
- Alignment with the rules for second-level domain labels
 - Under .jp (Japanese TLD)
 - Under gTLDs (such as .asia and .com with second-level registration of IDNs)

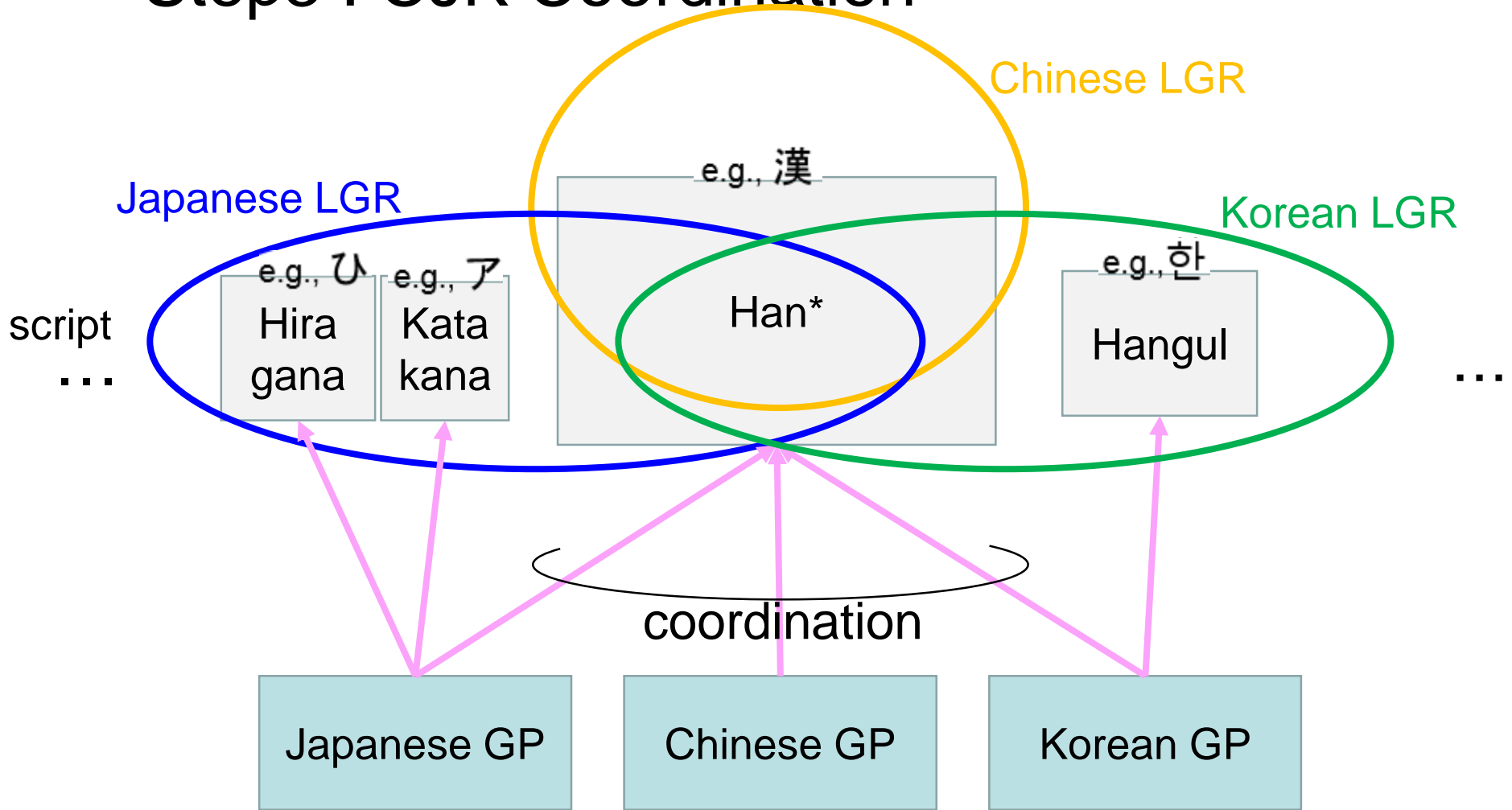
Step1 : Manning JGP

- Members and their expertise
 - Hiro Hotta chair
 - Policy/business aspects of registry/registrar
 - Akinori Maemura vice chair
 - Internet governance and domain name in general
 - Shigeki Goto
 - Internet in general
 - Kazunori Konishi
 - Internet in general
 - Tsugizo Kubo
 - Trademarks and domain names
 - Yoshitaka Murakami
 - Trademarks and gTLD markets from registry/registrar perspective
 - Shuichi Tashiro
 - Character codes
 - Yoshiro Yoneya
 - Technical aspects of IDN, LGR

Step2 : First Version of Japanese LGR

- Scopes of the character codes
 - Kanji, Hiragana, Katakana
 - For Kanji
 - JIS (Japanese Industrial Standard) level-1 and level-2
- Variants
 - For Kanji
 - Japanese LGR will define no variants for itself
 - Final Japanese LGR will import (= passively adopt) variants of Chinese LGR and Korean LGR
- WLE (whole label evaluation)
 - Japanese LGR may have no or very limited number of tiny rules for the usage of characters even if defined

Step3 : CJK Coordination



* "Han" is called "Kanji" in Japan, "Hanja" in Korea

Coordinated definition of variants has been completed

Step4 : Consultation with IP and Japanese community

- Reduction of the number of allocatable labels
 - Any combination of characters is allowed in Japanese strings
 - It may make the number of variant strings very huge, as many variant groups are defined by importing Chinese variants
 - E.G., 慶応大学 has 3 variant strings – 慶應大学/慶応大學/慶應大學
 - JGP needs to reduce the number of allocatable labels
 - JGP is trying to solve it by limiting allowed strings by employing the notion that “allocatable labels basically consists of day-use Japanese characters”
 - However - it seems Japanese community is not comfortable with this solution because most gTLDs in Japanese scripts may not be general nouns but trade names or geo names that often encompass personal names or geo names, which sometimes contain characters that are not “day-use Japanese characters”
 - JGP is still seeking the way to reduce allocatable labels

Korean GP Status Update

KIM Kyongsok
Korean GP Chair

0. Agenda

- ⊙ Introduction
- ⊙ K-LGR v0.7 (2017.03.03.)
 - 11172 Hangeul Syllables, 4758 Hanja characters, 152 Variant groups
- ⊙ K-LGR proposal and XML sent to IP
- ⊙ History of KGP activities
- ⊙ Timeline of KGP activities

1. Introduction

- ⊙ Characters included for “Kore” (Korean Label)
 - Both Hangeul (Hangul) syllables and Hanja chars are included in K-LGR
- ⊙ K-LGR v0.7 (3 March 2017)
 - 11172 Hangeul syllables
 - 4758 Hanja chars, 152 variant groups
- ⊙ In January 2017, the Name and Address Committee of Korea Internet Governance Alliance (KIGA, <http://kiga.or.kr>) formally created a WG for allowing Hanja at the second level under .KR/.한국
 - The WG started working

2. K-LGR v0.7 (3 March 2017)

- ⊙ A list of Hangul Syllables for K-LGR v0.7 (3 March 2017)
 - 11172 Hangul Syllables (U+AC00 ~ U+D7A3) ← ISO/IEC 10646
- ⊙ A list of Hanja characters for K-LGR v0.7 (3 March 2017)

Source of Hanja Character Set	# chars
1) KS X 1001 (268 comptb. chars excluded)	4620
2) IICORE - K column marked	4743
K-LGR v0.7 (3 March 2017): Hanja List (Union of 1) and 2))	4758

- K-LGR v0.7 (3 March 2017): 4758 Hanja chars, 152 variant groups
 - 61 chars in KP0 (= KPS 9566) and/or Hanja Test deleted
- No conflict in variant groups between K-LGR v0.7 (3 March 2017) and C-LGR (31 March 2017)

3. K-LGR Proposal and XML Sent To IP

- ⊙ 11 May 2015, repertoire and variant groups in K-LGR v0.1 sent to IP
 - 21 June 2015, IP comments received by KGP
- ⊙ 20 August 2016, XML for K-LGR v0.4 sent to IP
 - 14 September 2016, IP comments received by KGP
- ⊙ 21 December 2016, Proposal and XML for K-LGR v0.6 sent to IP
 - 2 February 2017, IP comments received by KGP
- ⊙ 24 May 2017, Proposal and XML for K-LGR v0.7 sent to IP
 - 24 July 2017, IP comments received by KGP
- ⊙ ? December 2017, Rev. Proposal and XML for K-LGR v0.7 will be sent to IP

4. History of KGP Activities (1)

- 2013** Dec: Korean GP organized
- 2014** Mar: CJK coordination meeting during ICANN 49 @ Singapore
Jun: CJK coordination meeting during ICANN 50 @ London; KGP status update
Jun: 1st KGP meeting
Aug: 2nd KGP meeting
Oct: CJK coordination meeting during ICANN 51 @ LAX; KGP status update
- 2015** Jan: 3rd KGP meeting; KGP reorganized
Feb: CJK coordination meeting during ICANN52 @ Singapore; KGP status update
Apr: 4th and 5th KGP meetings; KGP reorganized
May: 6th and 7th KGP meetings (K-LGR, v0.1); CJK coordination meeting in Seoul, Korea
Jun: 8th KGP meeting (K-LGR, v0.2) CJK coordination meeting, ICANN53@Buenos Aires
Jul: 9th KGP meeting and workshop; participated in APriGF @ Macau
Aug: 10th KGP meeting (K-LGR, v0.3)
Sep: 11th KGP meeting
Oct: Call for formal Generation of KGP (Korean Script Generation Panel) to ICANN;
CJK coordination meeting during ICANN54 @ Dublin

4. History of KGP Activities (2)

2015 | Nov: 12th KGP meeting

2016 | Jan: 13th KGP meeting

Feb: The Korean Community “formally” Forms Generation Panel for Developing the Root Zone Label Generation Rules (LGR), 2016-02-01

Mar: CJK coordination meeting during ICANN55 @ Marrakesh; KGP status update

Mar: 14th KGP meeting (K-LGR, v0.4)

Mar: CJK coordination meeting @ Beijing, China

Apr: 15th KGP meeting

May: 16th KGP meeting

Jun: 17th KGP meeting

Jun: CJK coordination meeting during ICANN56 @ Helsinki; KGP status update

Jul: 18th KGP meeting

Aug: 19th KGP meeting

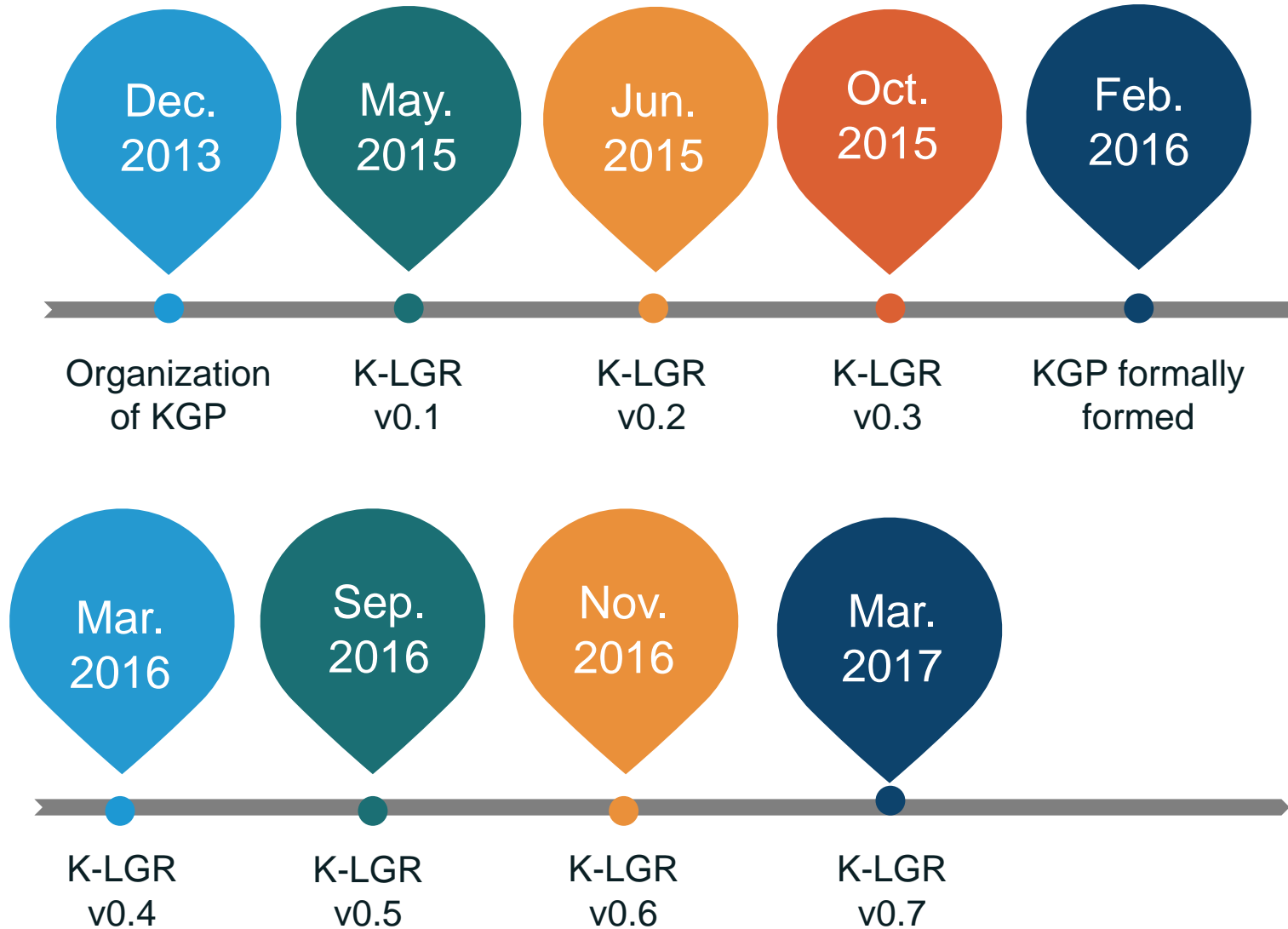
Sep: 20th KGP meeting; (K-LGR, v0.5); Public Hearing for K-LGR @KriGF Workshop)

4. History of KGP Activities (3)

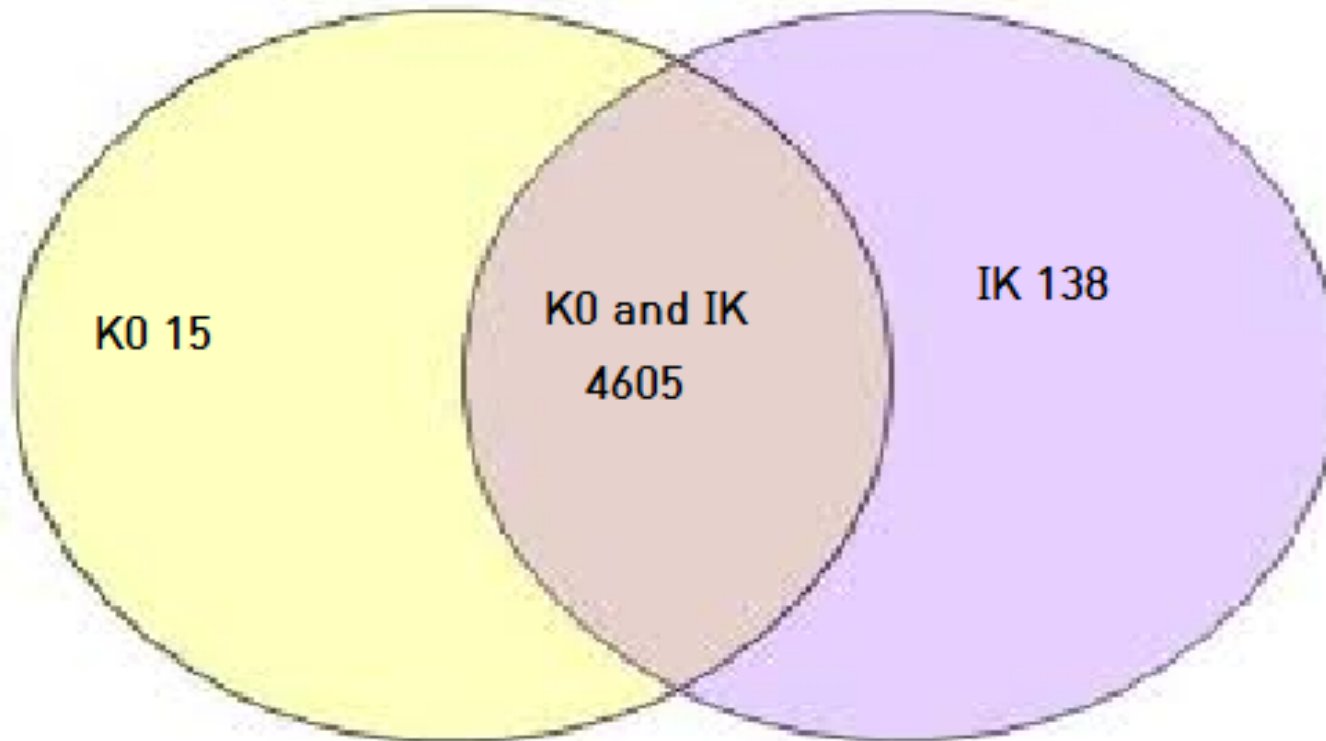
- 2016**
- Sep: CJK coordination meeting @ Taipei, Taiwan
 - Oct: 21st and 22nd KGP meetings
 - Nov: CJK coordination meeting during ICANN 57 @ Hyderabad, India
 - Nov: CJK Coordination Meeting @ Seoul during IETF #97 (K-LGR, v0.6)
 - Dec: 23rd KGP meeting

- 2017**
- Feb: CK coordination meeting during CDNC meeting @ Beijing, China
 - Mar: 24th KGP meeting
 - K-LGR, v0.7 (3 March 2017): # Hanja chars 4819 -> 4758 (61 chars deleted)
 - Mar: CJK coordination meeting during ICANN58 @ Copenhagen; KGP status update
 - Apr: 25th KGP meeting
 - Sep: 26th KGP meeting
 - Sep: Public Hearing for K-LGR @KriGF Workshop
 - Oct: KGP status update, ICANN60 @ Abu Dhabi, UAE

5. Timeline of KGP Activities



Appendix. Hanja in K-LGR v0.7 (2017.03.03.): K0 and IICORE/K (IK)



Venn Diagram of 2 sets showing number of Hanja chars:

K-LGR v0.7 (2017.03.03.): 4758 chars

K0 (KS X 1001): 4620 chars, IK (IICORE: K): 4743 chars

Greek GP Status Update

Panagiotis Papaspiliopoulos
Greek GP Chair

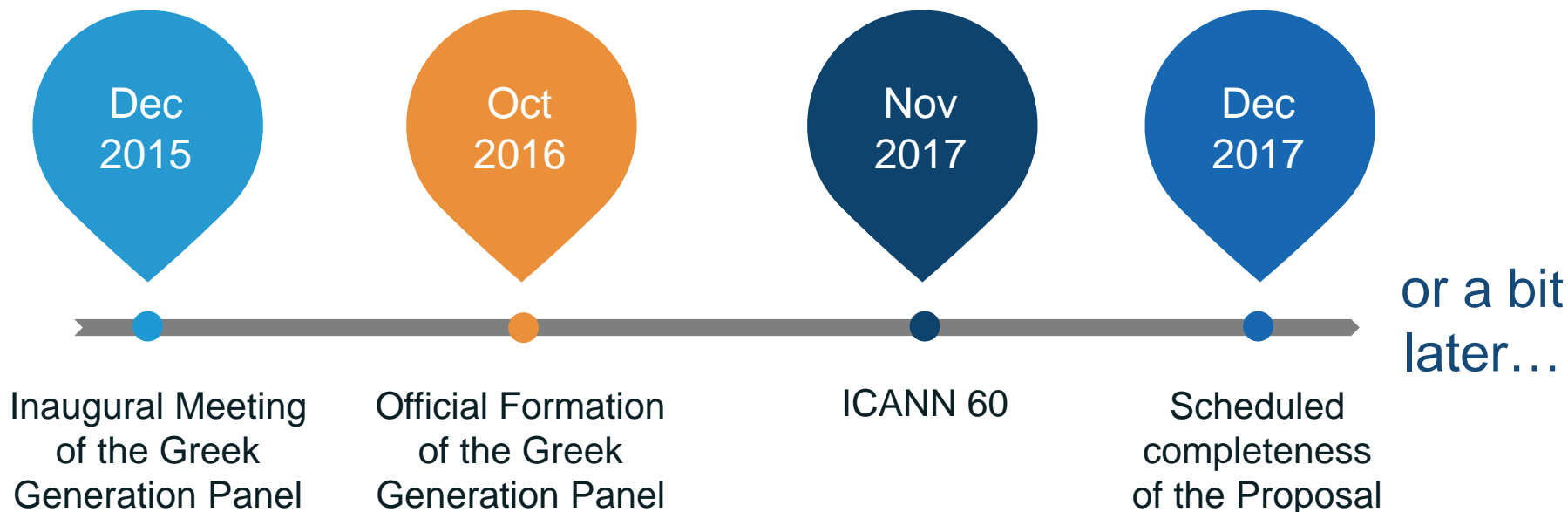
Script for Which The LGR Will Be Proposed

- ⊙ Greek
- ⊙ ISO 15924 Code: Grek
- ⊙ ISO 15924 Key No: 200
- ⊙ ISO 15924 English Name: Greek
- ⊙ Property Value Alias: Greek
- ⊙ Native Name of the script: Ελληνικά
- ⊙ Maximal Starting Repertoire (MSR) version: MSR-2
- ⊙ Unicode Standard 9.0: 0370-03FF

Composition of The Greek Generation Panel

- ⊙ Individuals actively involved in policy development processes related to electronic communication, research and development related to the Greek language, standardization, computing and maintenance of the domain name system in Greece and Cyprus
- ⊙ The panelists come from governmental and regulatory authorities, academia, private sector, ccTLD registries of Greece and Cyprus

Timeline



Moreover...

The Greek Generation Panel was officially formed with the Decision 54020/1088/25 Nov 2015 of the Minister of Infrastructure, Transport and Networks of the Hellenic Republic.

Structure of The Proposal

- ⊙ Background on script
 - Other languages using the Greek script (e.g. Pomak, Karamanlidika, Arvanitika etc)
 - The Greek language question
 - The Greek orthography
- ⊙ Overall development process and methodology
- ⊙ Repertoire
- ⊙ Variants
 - Within Script Variants
 - Cross Script Variants
- ⊙ Whole Label Evaluation rules

Useful Key Points

- ⊙ The Pomak language
 - Oral, Rup dialects spoken in Thrace (~30,000 people)
- ⊙ The Greek language question
 - Diglossia: Katharevousa vs. Dimotiki (e.g. Πειραιεύς - Πειραιάς)
- ⊙ The Greek orthography
 - Within Script Variants
 - Cross Script Variants
- ⊙ Whole Label Evaluation rules
 - Polytonic vs. Monotonic (e.g. the Lord's Prayer)
 - Πάτερ ἡμῶν ὁ ἐν τοῖς οὐρανοῖς, ἁγιασθήτω τὸ ὄνομά σου, ἐλθέτω ἡ βασιλεία σου, γεννηθήτω τὸ θέλημά σου, ὡς ἐν οὐρανῷ καὶ ἐπὶ τῆς γῆς
 - Πάτερ ημών ο εν τοις ουρανοίς, αγιασθήτω το όνομά σου, ελθέτω η βασιλεία σου, γεννηθήτω το θέλημά σου, ως εν ουρανώ και επί της γης

Proposed Characters for Registrations

- ⦿ Only Monotonic characters are to be allowed!
 - Monotonic characters are used today in the spelling of Greek words
 - Polytonic characters for TLD registration offer no significant advantage for the average user.
 - This recommendation stands only for the Greek TLDs; registries are encouraged to decide upon the use of the polytonic characters in their policy for the lower levels

- ⦿ The Pomak language does not affect the formation of domain names with the Greek characters' set

Within Script Variants: The Sigma and Final Sigma

- The sigma and final sigma
 - σ (small, U+03C3) e.g. ασπίδα (shield)
 - Σ (capital, U+03A3) e.g. Σοφία (female name)
 - ς (small final, U+03C2) e.g. τέλος (end)
- But..
 - IDNA2003: σ <-> Σ <-> ς (e.g. θησαυρός – treasure)
θησαυρός → ΘΗΣΑΥΡΟΣ → θησαυροσ → **not correct Greek!**
 - IDNA2008: σ and ς are different accepted characters,
treated separately – reverse mapping no possible!
- Conclusion: handle ‘σ’ and ‘ς’ as within script variants!

Within Script Variants: Vowels

Non-accented vowel	Vowel with tonos	Vowel with diaeresis	Vowel with tonos and diaeresis
α U+03B1	ά U+03AC		
ε U+03B5	έ U+03AD		
η U+03B7	ή U+03AE		
ι U+03B9	ί U+03AF	ϊ U+03CA	ΐ U+0390
ο U+03BF	ό U+03CC		
υ U+03C5	ύ U+03CD	ϋ U+03CB	Ϝ U+03B0
ω U+03C9	ώ U+03CE		

Issues to be Concerned: Cross Script Variants

- ⊙ Status: under process
- ⊙ Cases under examination
 - Greek and Latin
 - Greek and Cyrillic
 - Greek and Armenian
 - Greek and other scripts
- ⊙ Methodology:
 - Creation of tables per script (extensively)
 - Consideration of various fonts and sizes
 - Question: homoglyphs only or consider visual similarity cases too?
To what extend?
 - Key factor: consideration of other GPs conclusions

Cross Script Variants (Examples): Greek - Latin

Greek Letter		Latin Letter	
U+03B1	α	α	U+0251
Greek small letter ALPHA		Latin small letter ALPHA	
U+03B2	β	ß	U+00DF
Greek small letter BETA		Latin small letter SHARP S	
U+03B7	η	h	U+0068
Greek small letter ETA		Latin small letter H	
		n	U+006E
		Latin small letter N	
		ñ	U+0146
		Latin small letter N WITH CEDILLA	
		η	U+014B
		Latin small letter ENG	
		ħ	U+0266
Latin small letter H WITH HOOK			
ñ	U+1E47		
Latin small letter N WITH DOT BELOW			
U+03AE	ή	ñ	U+0144
Greek small letter ETA WITH TONOS		Latin small letter N WITH ACUTE	
		ñ	U+01F9
		Latin small letter N WITH GRAVE	
		ñ	U+1E45
		Latin small letter N WITH DOT ABOVE	

Cross Script Variants (Examples): Greek - Cyrillic

Greek Letter	Cyrillic Letter
U+03B2 β Greek small letter BETA	В U+0432 Cyrillic small letter VE
U+03B7 η Greek small letter ETA	И U+0438 Cyrillic small letter I h U+04BB Cyrillic small letter SHHA
U+03B8 θ Greek small letter THETA	е U+04E9 Cyrillic small letter BARRED O
U+03BA κ Greek small letter KAPPA	к U+043A Cyrillic small letter KA ќ U+049B Cyrillic small letter KA WITH DESCENDER κ U+04A1 Cyrillic small letter BASHKIR KA
U+03C4 τ Greek small letter TAU	т U+0442 Cyrillic small letter TE τ U+04AD Cyrillic small letter TE WITH DESCENDER

Cross Script Variants (Example): Greek - Armenian

Greek Letter	Armenian Letter
U+03B7 η Greek small letter ETA	ղ U+0572 Armenian small letter GHAD ն U+0578 Armenian small letter VO դ U+0564 Armenian small letter DA ռ U+057C Armenian small letter RA

Cross Script Variants (Under Question): Other Scripts

Greek Letter		Georgian Letter	
U+03BE	ξ	Ⴡ	U+10F0
Greek small letter XI		Georgian letter HAE	
		Ⴢ	U+10F4
		Georgian letter HAR	
U+03C6	φ	Ⴣ	U+10F6
Greek small letter PHI		Georgian letter FI	
Greek Letter		Ethiopic Syllable (Font: Ebrima)	
U+03BB	λ	ለ	U+1208
Greek small letter LAMDA		Ethiopic syllable LA	
		ለ	U+120B
		Ethiopic syllable LAA	
U+03C2	ς	ኆ	U+1293
Greek small letter FINAL SIGMA		Ethiopic syllable NAA	
Greek Letter		Myanmar Letter (Font size: 16)	
U+03C9	ω	ဝ	U+1003
Greek small letter OMEGA		Myanmar letter GHA	
		ဝ	U+101A
		Myanmar letter YA	

Issues to be Concerned: WLE Rules

- ⊙ Status: under process
- ⊙ Issues should be taken into consideration
 - Contextual information
 - Existing rules (since 2005)
 - User experience of the average Greek speaking Internet users

Drawbacks - Reasons for (rather) Slow Progress

- ⊙ Heavy work load of the members of the Greek Generation Panel due to other demanding professional obligations
- ⊙ Difficulty in settling face-to-face meetings (distance, little available time). Use of telephone and mailing lists only! → too much time to get all responses, conclude and move forward
- ⊙ Careful steps due to problems in previous projects (i.e. discussions about confusing similarity cases during the IDN ccTLD Fast Track Process for .ελ and Greek .eu)

But,

- ⊙ **We feel that we are moving to the final act !**

Thank You and Questions

- ⦿ Reach us at
 - GreekGP@icann.org
 - <https://community.icann.org/display/croscomlgrprocedure/Greek+Script+GP>

Engage with ICANN and IDN Program



Thank You and Questions

Reach us at: IDNProgram@icann.org

Website: icann.org/idn



twitter.com/icann



[gplus.to/icann](https://plus.google.com/icann)



facebook.com/icannorg



weibo.com/ICANNorg



linkedin.com/company/icann



flickr.com/photos/icann



youtube.com/user/icannnews



slideshare.net/icannpresentations