ICANN 58

COMMUNITY FORUM

COPENHAGEN

11–16 March 2017

Agenda

⊙ Root Zone LGR-2   – Asmus Freytag, IP

⊙ Community Updates

- Chinese GP   – Wang Wei
- Japanese GP   – Hiro Hotta
- Korean GP   – KIM Kyongsok
- Thai GP   – Jitti Kunphruk

⊙ Q&A

Root Zone Label Generation Rules - 2

Asmus Freytag
Integration Panel

ICANN

LGR-2: What's in It - Scripts

- ⊙ Added for LGR-2
  - o Georgian
  - o Khmer
  - o Lao
  - o Thai

- ⊙ Retained from LGR-1
  - o Arabic

ქართული

ភាសាខ្មែរ

ລາວ

ไทย

عربى

- ⊙ Except for Georgian, all additions are SEA complex scripts
- ⊙ Khmer, Lao and Thai integrated together because related/similar
- ⊙ Armenian continues to be deferred until related scripts are available

- ⊙ "Integration" doesn't mean that LGR-2 will be a single file

LGR-2: What's in It - Files

- ⊙ Overview document (background)

- ⊙ Code tables (graphical overview of repertoire)

- ⊙ Merged LGR with
  - integrated repertoire
  - common variant mappings (all blocked)
  - common WLE rules and actions

- ⊙ One Element LGR per script
  - per-script repertoire
  - per-script variant mappings
  - using per-script variant types
  - WLE rules and actions that can be triggered by code points in per-script repertoire

  - plus default rules and actions for the Root Zone (defined in MSR-2)

Georgian

LGR-2 Development

- Each Element LGR:
  - derived from LGR proposal submitted by Generation Panel (GP)
  - Integration Panel (IP) does copy editing of descriptions and annotations for consistency
  - repertoire and rules are unchanged

- Merged LGR:
  - created by mechanical merge (see above)
  - common description
  - annotations replaced by pointers to element LGRs for details
  - rules, classes and tag values are prefixed by script to avoid name collisions

LGR-2 Verification

⊙ Each GP submits test labels that can be used to verify its LGR
  o Both valid and invalid labels
  o Aim is reasonable coverage of expected contexts

⊙ IP verifies mechanically that Proposal and final Element LGR give same label disposition
⊙ IP verifies mechanically that merged LGR correctly integrates the Element LGRs

⊙ Public comments allow further review and double check by community

LGR-2: How to Use It

Idealized steps in processing a label application:

1. Application defines a script context
   - Labels in the Root Zone are in a single script

2. Process application using Element LGR for that script

3. Determine validity of the applied-for label
   - fits repertoire subset
   - matches context rules on code points
   - matches whole label rules
   - doesn't have reflexive variants that make the label ineligible (e.g. out-of-repertoire-var)

LGR-2: How to Use It (continued)

4. Use Merged LGR to check for collisions
  - o create "index" variant for label based on merged LGR
  - o compare to stored index variants for delegated labels
  - o any match: applied label and all its variants are in conflict

5. Use Element LGR in final step to
  - o generate list of allocatable variants

⊙ Other steps in the process will determine which labels and which allocatable variants will actually be delegated

LGR-2 Q&A

⊙ This is an extended Q&A session:
IP invites questions both on LGR-2 as well as on future Root Zone LGR versions or pending LGR proposals

Update by the Chinese GP

Wang Wei and Kenny Huang
Chinese GP Co-Chairs

ICANN

CGP LGR Proposal

- Repertoire
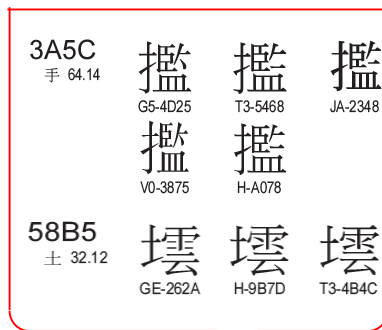
0	CDNC	19561=19559 + 2 (non MSR)
1	dotAsia	124=62+62 (Plane2)
2	NHCU	18
3	IICORE∩JGP IICORE∩KGP	43
		19746

IICORE∩JGP
IICORE∩KGP   43

TGSCC 18

dotAsia 124

CDNC
19561
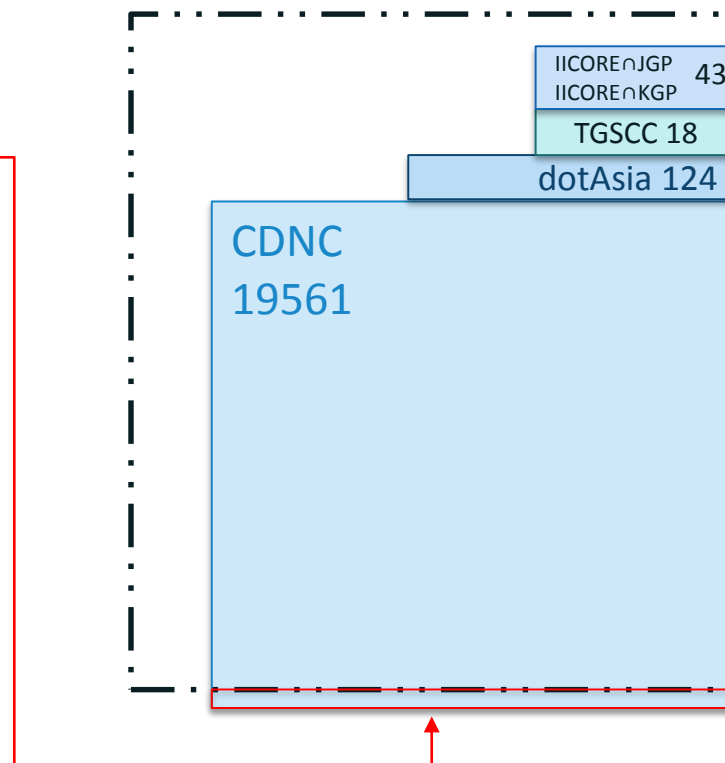
JGP Kanji: 6358

KGP Hanja: 4819

overlap
4161

4809

6267

CGP R3: 19746
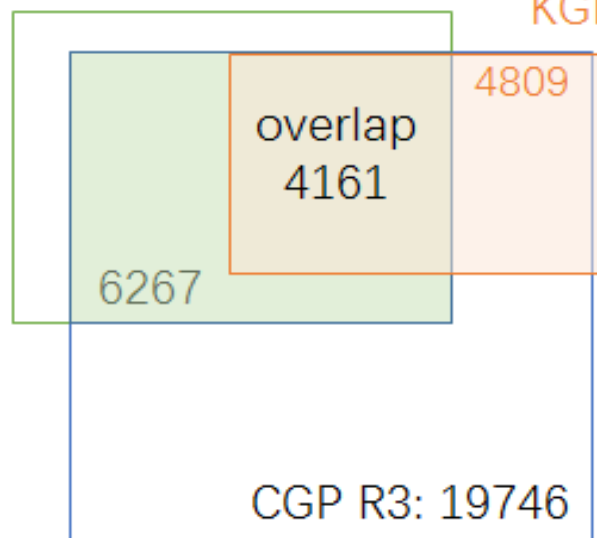
CGP LGR Proposal

• Variant Mappings：CGP interior coordination

```
┌─────────────────────┐
│        CDNC         │
│  Variant Mappings   │
└─────────────────────┘

┌─────────────────────┐
│ 172 Non-CDNC Chars  │                              ┌─────────────────────┐
│ Variant Mappings    │ ───────────────────────────► │        CGP          │
│ Review              │                              │  Variant Mappings   │
│ Sept 2015           │                              └─────────────────────┘
└─────────────────────┘                                        ▲
                                                               │
┌─────────────────────┐              69 Plane0 chars           │
│   7 dotAsia Chars   │              62 Plane2 chars           │
│ Variant Mappings    │                                        ▼
│ Review              │                              ┌─────────────────────┐
│ May 2016            │                              │      dotAsia        │
└─────────────────────┘                              │  Variant Mappings   │
                                                     └─────────────────────┘
```
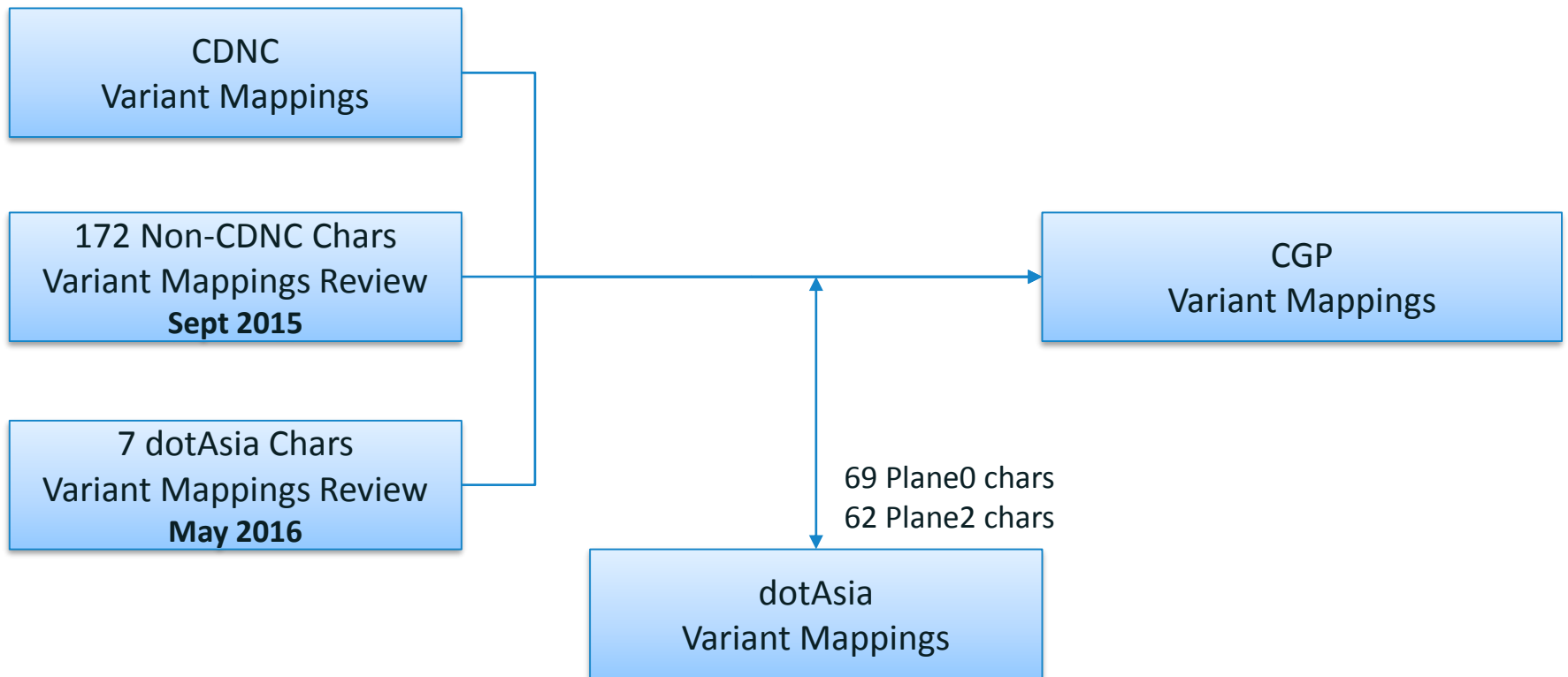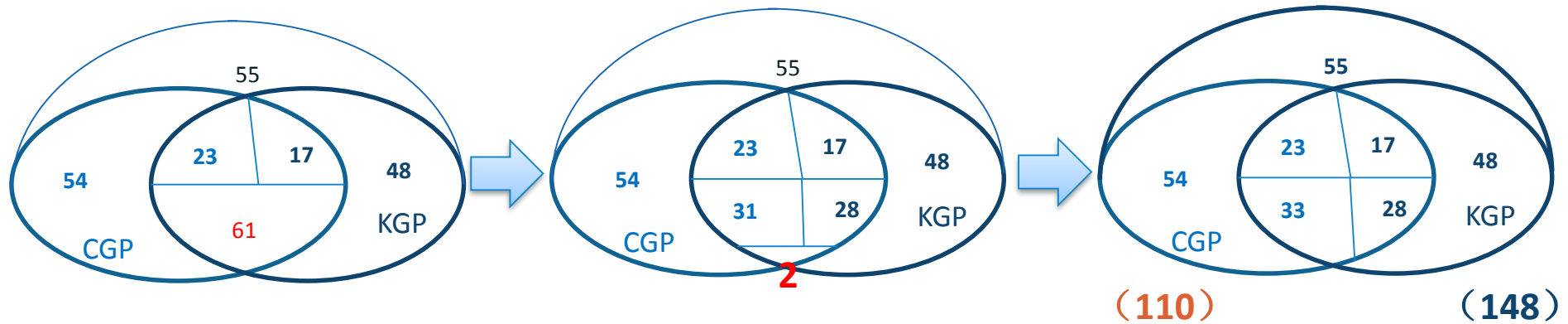
CGP LGR Proposal

- Variant Mappings：CJK coordination
  - J accept all C variant mappings
  - C & K coordination on 258 disagreed variant groups

Taiwan
September 2016

Seoul
November 2016

December 2016

CGP LGR Proposal

- Variants
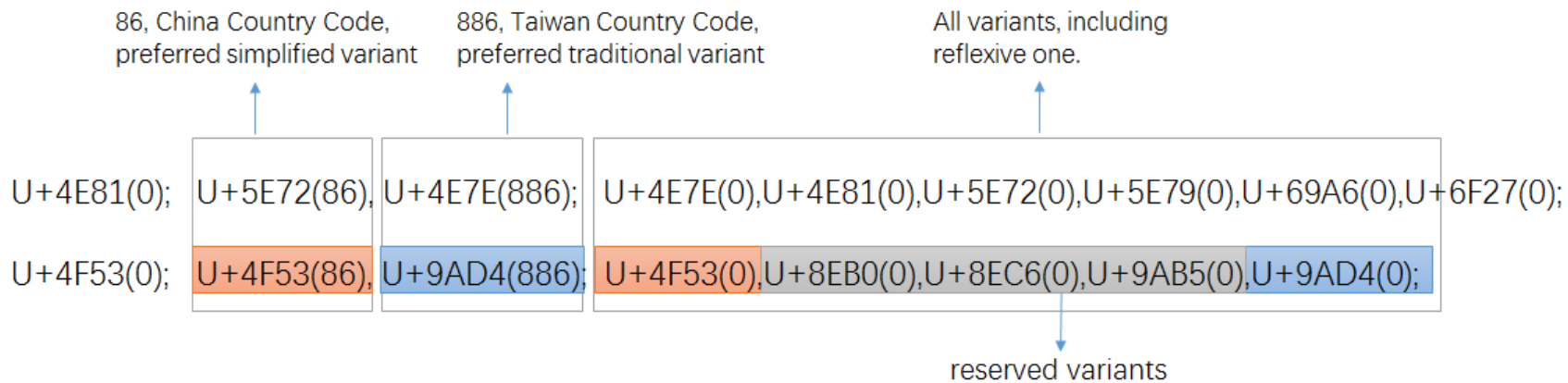
Size of variant group	Number of variant groups	Number of chars
1	11884	11884
2	2800	5600
3	507	1521
4	119	476
5	35	175
6	10	60
7	2	14
8	2	16
		19746

CGP LGR Proposal

- Variant Types and Sub-Types

"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

U+4E81(0); | U+5E72(86), | U+4E7E(886); | U+4E7E(0),U+4E81(0),U+5E72(0),U+5E79(0),U+69A6(0),U+6F27(0);

U+4F53(0); | U+4F53(86), | U+9AD4(886); | U+4F53(0),U+8EB0(0),U+8EC6(0),U+9AB5(0),U+9AD4(0);

86, China Country Code, preferred simplified variant

886, Taiwan Country Code, preferred traditional variant

All variants, including reflexive one.

reserved variants

```
<char cp="4E81" tag="sc:Hani" >
     <var cp="4E7E" type="trad" />
     <var cp="4E81" type="r-neither" comment="identity" />
     <var cp="5E72" type="simp" />
     <var cp="5E79" type="blocked" />
     <var cp="69A6" type="blocked" />
     <var cp="6F27" type="blocked" />
</char>
<char cp="4F53" tag="sc:Hani" >
     <var cp="4F53" type="r-simp" comment="identity" />
     <var cp="8EB0" type="blocked" />
     <var cp="8EC6" type="blocked" />
     <var cp="9AB5" type="blocked" />
     <var cp="9AD4" type="traded" />
</char>
```

CGP LGR Proposal

- Multiple preferred/allocatable variants

	chars	number of allocatable variants
multiple simplified allocatable variants	4	2
multiple traditional allocatable variants	128	2
	5	3
	1	4

CGP LGR Proposal

- WLE Rules

Allocatable labels = original input + all simplified + all traditional

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

Next Steps

- Provide the rationale for the number of repertoire
  - 19746 is a big number compared with other GPs repertoire

- Limit the number of allocatable labels
  - Limit the number of allocatable variants
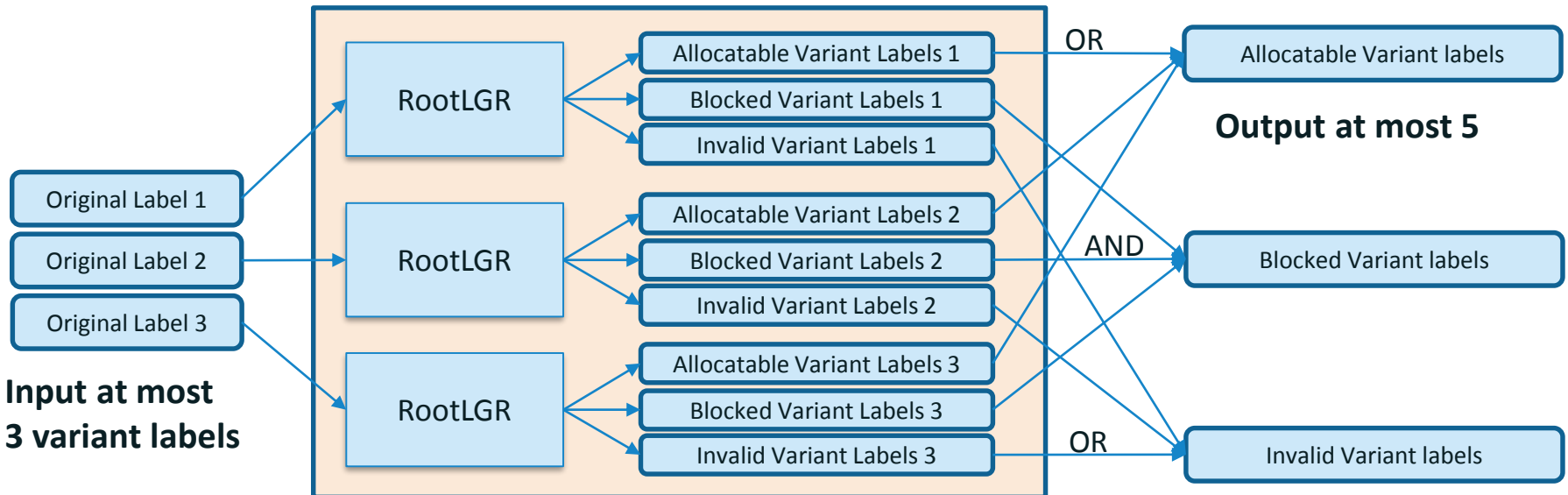 - Multiple LGR execution process

CGP LGR Proposal

Limit the number of allocatable variants

复旦 >> 复旦; 复旦，復旦，複旦

复旦 >> 复旦; 复旦
復旦 >> 复旦; 復旦
複旦 >> 复旦; 複旦

Sub-Type	Type
"simp-m"	Blocked
"trad-m"	Blocked
"both-m"	Blocked
"r-simp-m"	Blocked
"r-trad-m"	Blocked
"r-both-m"	Blocked

Multiple LGR execution process

Input at most 3 variant labels

Output at most 5

CGP LGR Proposal 2

- WLE 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" only-variants="simp simp-m r-simp r-simp-m both
both-m r-both r-both-m" comment="multiple simplified label" />
<action disp="blocked" only-variants="trad trad-m r-trad r-trad-m both
both-m r-both r-both-m" comment="multiple traditional label"/>
<action disp="blocked" any-variant="simp trad both r-simp r-trad r-both
simp-m trad-m both-m r-simp-m r-trad-m r-both-m r-neither" comment="block
any other mixed labels" />
<action disp="allocatable" comment="catch-all" />
</rules>
```

Thanks

Q&A

Update by the Japanese GP

Hiro Hotta
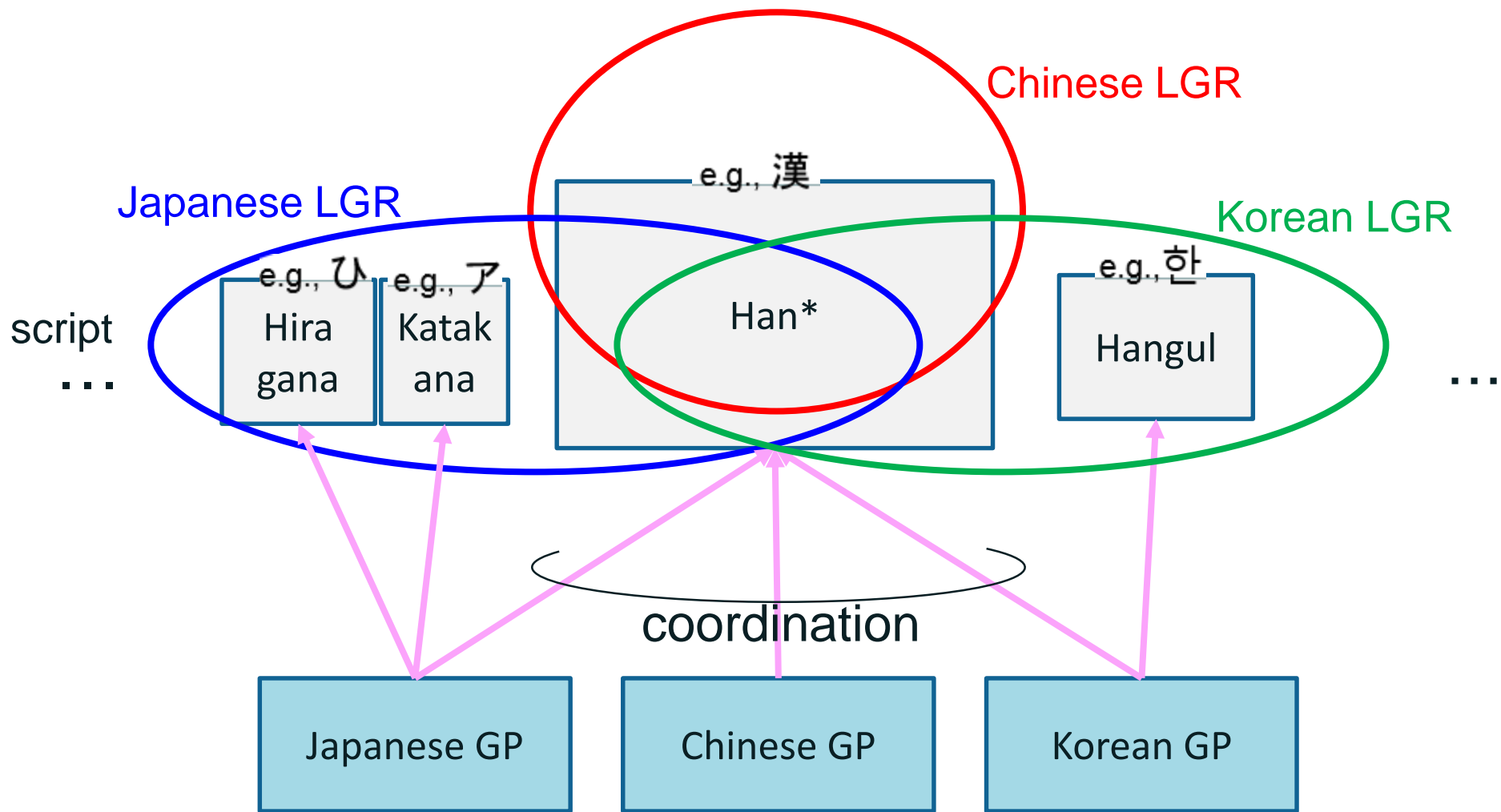Japanese GP Chair

ICANN

JGP Meetings & Related Events

- 2014
  - August 29   preparatory JGP meeting (1)
  - September 12   preparatory JGP meeting (2)
  - September 24   JGP meeting (1)
  - October 24   JGP meeting (2)
  - November 26   JGP meeting (3)
  - December 18   JGP meeting (4)
- 2015
  - January16   JGP meeting (5)
  - February 4   JGP meeting (6)
  - February 6   submission of JGP proposal to ICANN
  - February 20   JGP meeting (7)
  - March 10   JGP establishment approved by ICANN
  - March 18   JGP meeting (8)
  - April 15   JGP meeting (9)
  - May 15-16   CJK coordination meeting in Seoul
  - May 20   JGP meeting (10)
  - June 17   JGP meeting (11)
  - June 21-25   CJK coordination meeting during ICANN
  - September 29   JGP meeting (12)
  - October18-22   CJK coordination meeting during ICANN

JGP Meetings & Related Events (continued)
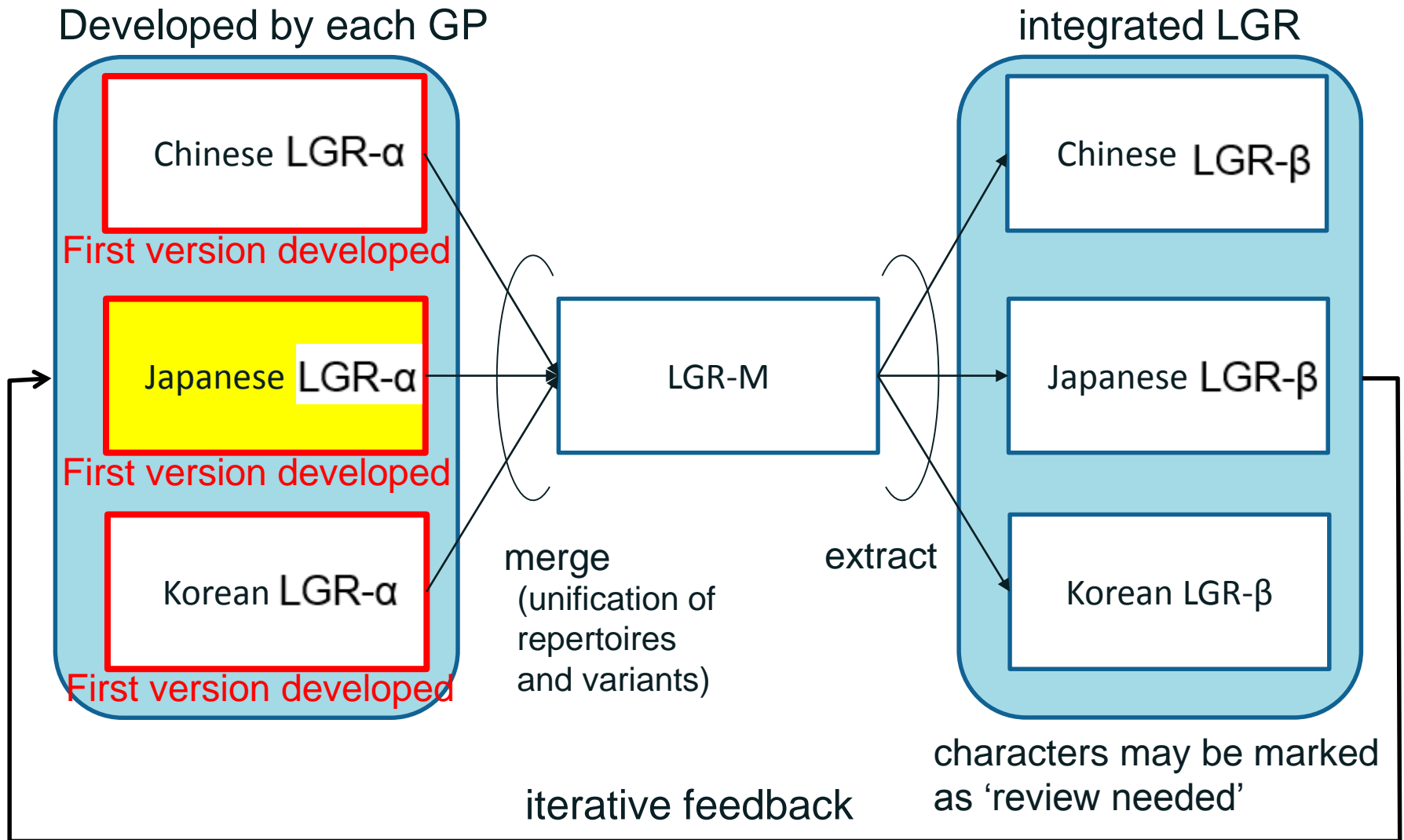
- 2016
  - March 6-10   CJK coordination meeting during ICANN
  - March 20   -21   CJK coordination meeting in Beijing
  - June 27-30   CJK coordination meeting during ICANN
  - September 24   J   GP meeting (13)
  - September 29-30   CJK coordination meeting in Taipei
  - November 4-8   CJK coordination meeting during ICANN
  - November 13-17   CJK coordination meeting during IETF

Relationship Among CJK Language LGRs

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

Developed by each GP

integrated LGR

Chinese LGR-α

First version developed

Japanese LGR-α

First version developed

Korean LGR-α

First version developed

LGR-M

merge
(unification of
repertoires
and variants)

extract

Chinese LGR-β

Japanese LGR-β

Korean LGR-β

characters may be marked
as 'review needed'

iterative feedback

Overview of Japanese LGR-α (J-LGR-α)

- Repertoire
  - Consists of characters from 3 scripts (Han, Hira and Kana – Jpan in ISO 15924)

Script	# of characters
Han	6358
Hira	85
Kana	89
Total	**6532**

- Variants & their types
  - No variants
  - Types of imported variants will be investigated and determined after LGR-α from CGP and KGP are proposed
- WLE
  - Rules (although not very many) are under discussion

- Reduction of the number of allocatable labels
  - Variant labels will exist by importing CGP variant and KGP variant characters, although JGP defines no variants
  - Strings containing any combinations of characters are allowed as natural Japanese words
  - Therefore, many variant labels may exist for a label
    - Waiting for Chinese and Korean GP's definition of variant characters, which CGP and KGP are coordinating to converge => making significant progress

  - IP (Integration Panel) requests JGP to reduce the number of allocatable labels
    - Trying some ideas to reduce allocatable labels

- After consultation with ICANN and IP, the idea was temporarily made frozen
  - Because such change is not promised to be adopted
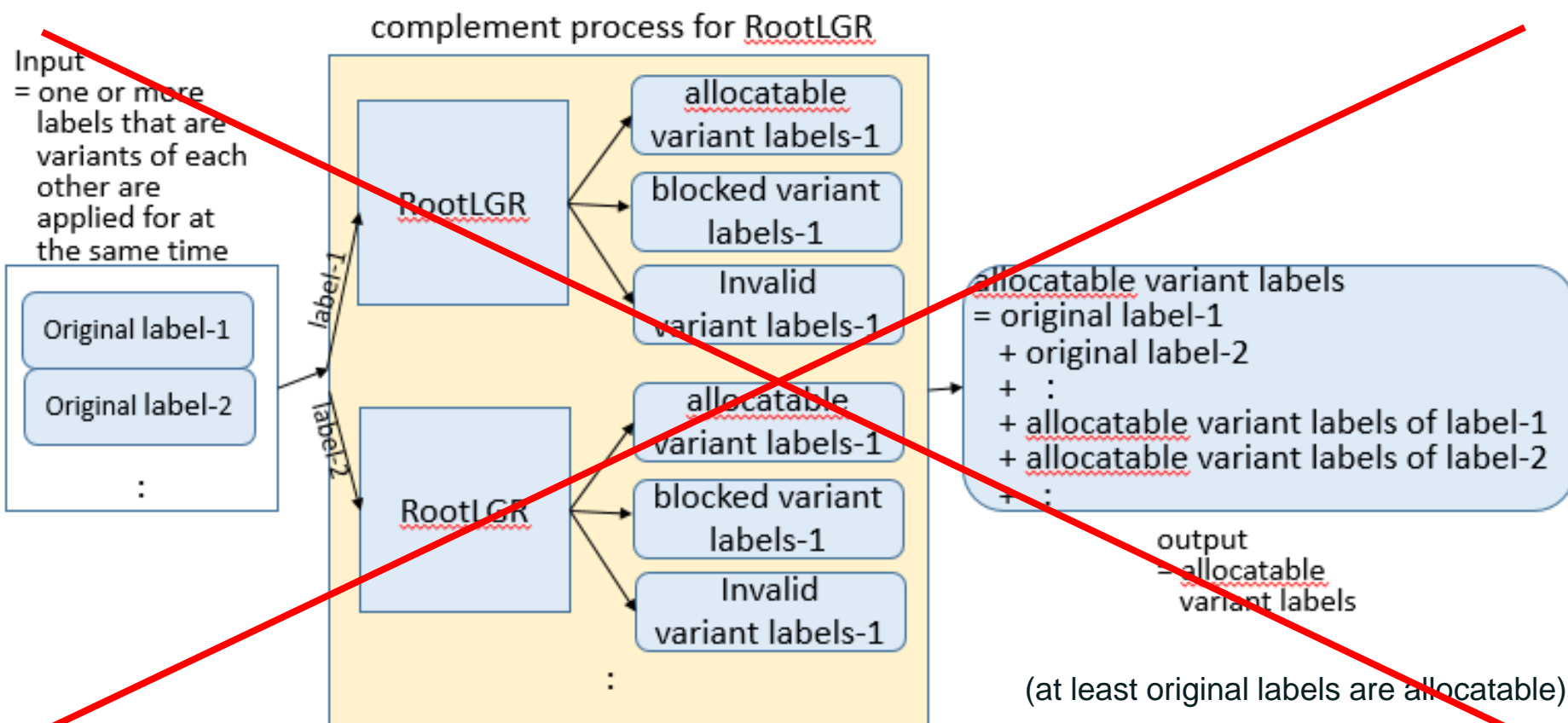
Update by the Korean GP

KIM Kyongsok
Korean GP Chair

0. Agenda

⊙ Introduction

⊙ A list of Hangul Syllables, Hanja characters for K-LGR v0.7 (2017.03.03.)

⊙ Review of K (Korean) and C (Chinese) Variant Groups (Sets)

⊙ K-LGR v0.7 (2017.03.03.), # Hanja chars: 4819 -> 4758

⊙ 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 (2017.03.03.)
 - 11172 Hangeul syllables
 - 4758 Hanja chars, 152 variant groups

⊙ In Jan. 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 v0.5 (2016.09.28.)

- **A list of Hangul Syllables for K-LGR v0.7 (2017.03.03.)**
  - 11172 Hangul Syllbles (U+AC00 ~ U+D7A3) ← KS X ISO/IEC 10646

- **A list of Hanja characters for K-LGR v0.7 (2017.03.03.)**

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 (2017.03.03.): Hanja List (Union of 1) and 2))	4758

- ⊙ **3475 C vg's in C-LGR (2017.02.23.) analyzed based on K chars in K-LGR v0.7TMP (2017.02.23.)**
  - ○ K-LGR v0.7TMP (2017.02.23.): 4819 Hanja chars and 168 variant groups
  - ○ C-LGR (2017.02.23.): 19744 Hanzi chars and 3475 variant groups (a variant group (set) is composed of two or more variant chars)
- ⊙ **Analysis of 3475 C (Chinese) variant groups (sets)**
  - ○ 168 C variant groups where there are two or more K characters
    - K character is a character belonging to K-LGR v0.7TMP (2017.02.23.)
  - ○ No or just one K char in the remaining 3307 (= 3475 – 168) C vg's
- ⊙ **No more conflict in variant groups between K-LGR v0.7TMP (2017.02.23.) and C-LGR (2017.02.23.)**

⊙ KGP and CGP coordinated to resolve conflicts of variant groups between K-LGR and C-LGR.
  - Three possible scenarios in resolving conflicts:

 scenario 1) C vg is kept without any modification (K accepts C position):
  - [K: indep (C1), indep (C2)]; [C: vg (C1, C2)] → [K & C: vg (C1, C2)]

 scenario 2) C vg is fully split (i.e., no two K chars are included in a variant group)
  - [K: indep (C1), indep (C2)]; [C: vg (C1, C2)] → [K & C: indep (C1), indep (C2)]
  - [K: indep (C1), indep (C2), indep(C3)]; [C: vg (C1, C2, C3)] → [K & C: indep (C1), indep (C2), indep(C3)];

scenario 3) C vg is partially split. (usually) One variant char is split from C vg and that char becomes an independent char.
  - a new variant group still contains two (or more) K chars.

  - [K: vg (C4, C5), indep (C6)]; [C: vg (C4, C5, C6)];
   → [K & C: vg (C4, C5), indep (C6)]

3. Review of K and C Variant Groups (4)

- ⊙ Coordination Results (K and C agreed in principle on 2017.02.23. in Beijing)
- ○ #1a = #1b + #1c
- 1a) 304 Cvg which contain two or more K chars
- 1b)   46 Kvg are the same as Cvg (no conflict)
- 1c) 258 Cvg in conflict (= 304 – 46) (in conflict)

- ○ 258 Cvg resolved as follows: 258 = #2a + #2b + #2c
- 2a) 110 Cvg kept without any modification (K accepts C position)
- 2b)   12 Cvg partially split (a new variant group still contains two (or more) K chars)
- 2c) 136 C vg fully split (i.e., no two K chars are included in a variant group)

- · #Kvg = #1b + #2a + #2b = 46 + 110 + 12 = 168 (K-LGR v0.7TMP (2017.02.23.))
- · #Cvg fully or partially split = #2b + #2c = 12 + 136 = 148
- · #Cvg in C-LGR (2017.02.23.): 3475

- ⊙ **K-LGR v0.6 (2016.11.28.) and K-LGR v0.7TMP (2017.02.23.):**
    - ○ 4819 Hanja chars

- ⊙ **K-LGR v0.7 (2017.03.03.): 4758 Hanja chars**
    - ○ only 4758 chars in K0 (= KS X 1001) and/or IICORE/K column kept
    - ○ 61 chars in KP0 (= KPS 9566) and/or Hanja Test deleted

- ⊙ **No conflict in variant groups between K-LGR v0.7 (2017.03.03.) and C-LGR (2017.02.23.)**

5. History of KGP Activities (1)

2013   Dec: Korean LGP organized

2014   Mar: CJK coordination meeting ICANN 49 @ Singapore
Jun: : CJK coordination meeting ICANN 50 @ London; KGP status update
Jun: 1st KGP meeting
Aug: 2nd KGP meeting
Oct: CJK coordination meeting during ICANN 51 LA; KGP status update

2015   Jan: 3rd KGP meeting; KGP reorganized
Feb: CJK coordination meeting ICANN52 @ Singapore; KGP status update
Apr: 4th and 5th KGP meetings; KGP reorganized
May: 6th and 7th KGP meetings (K-LGR-1 v0.1); CJK Coordination meeting in Seoul, Korea
Jun: 8th KGP meeting (K-LGR, v0.2 CJK coordination meeting during 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

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 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 ICANN56 @ Helsinki; KGP status update
Jul: 18th KGP meeting
Aug: 19th KGP meeting
Sep: 20th KGP meeting; (K-LGR v0.5); CJK coordination meeting @ Taipei, Taiwan

5. History of KGP Activities (3)

2016

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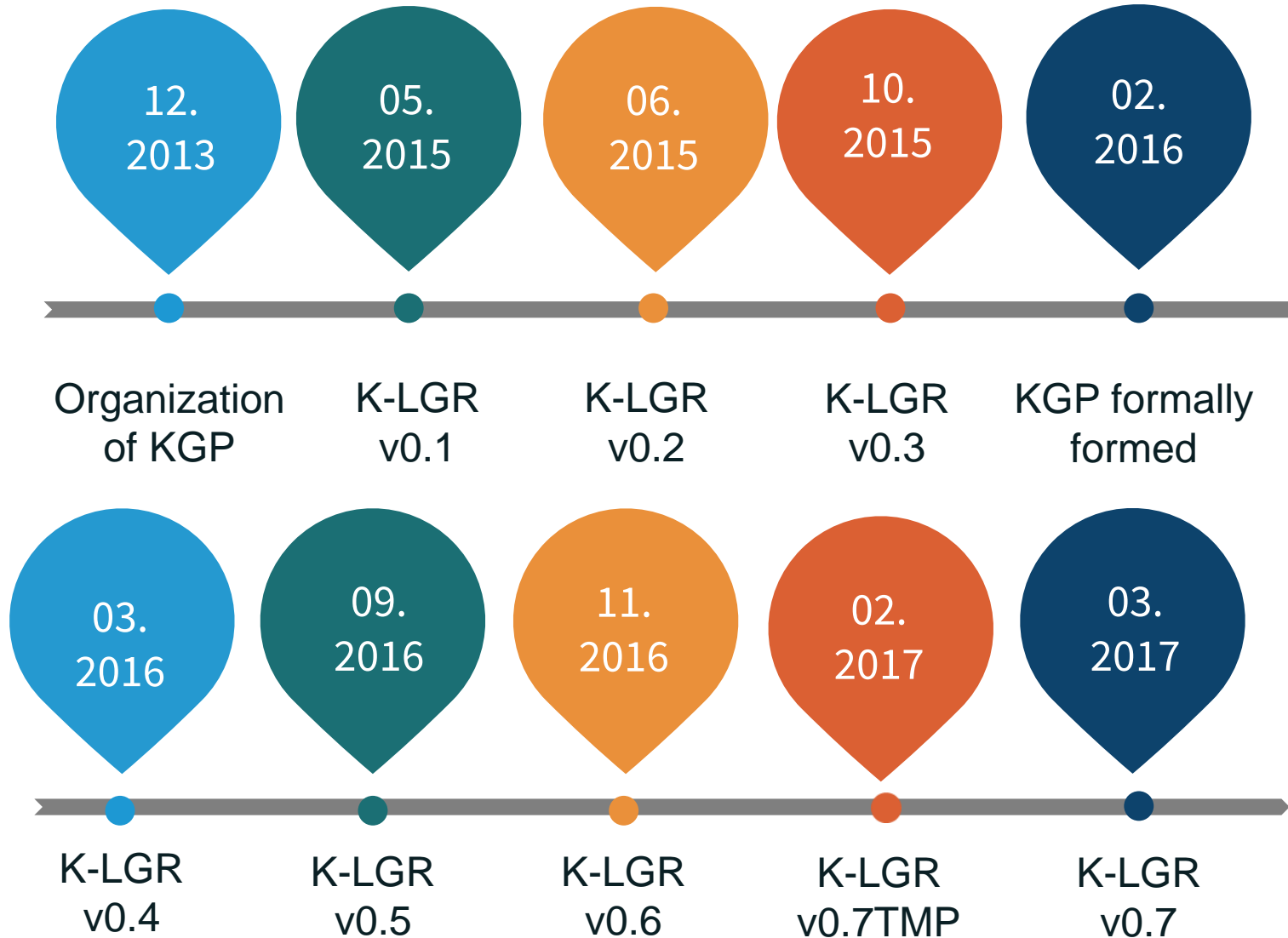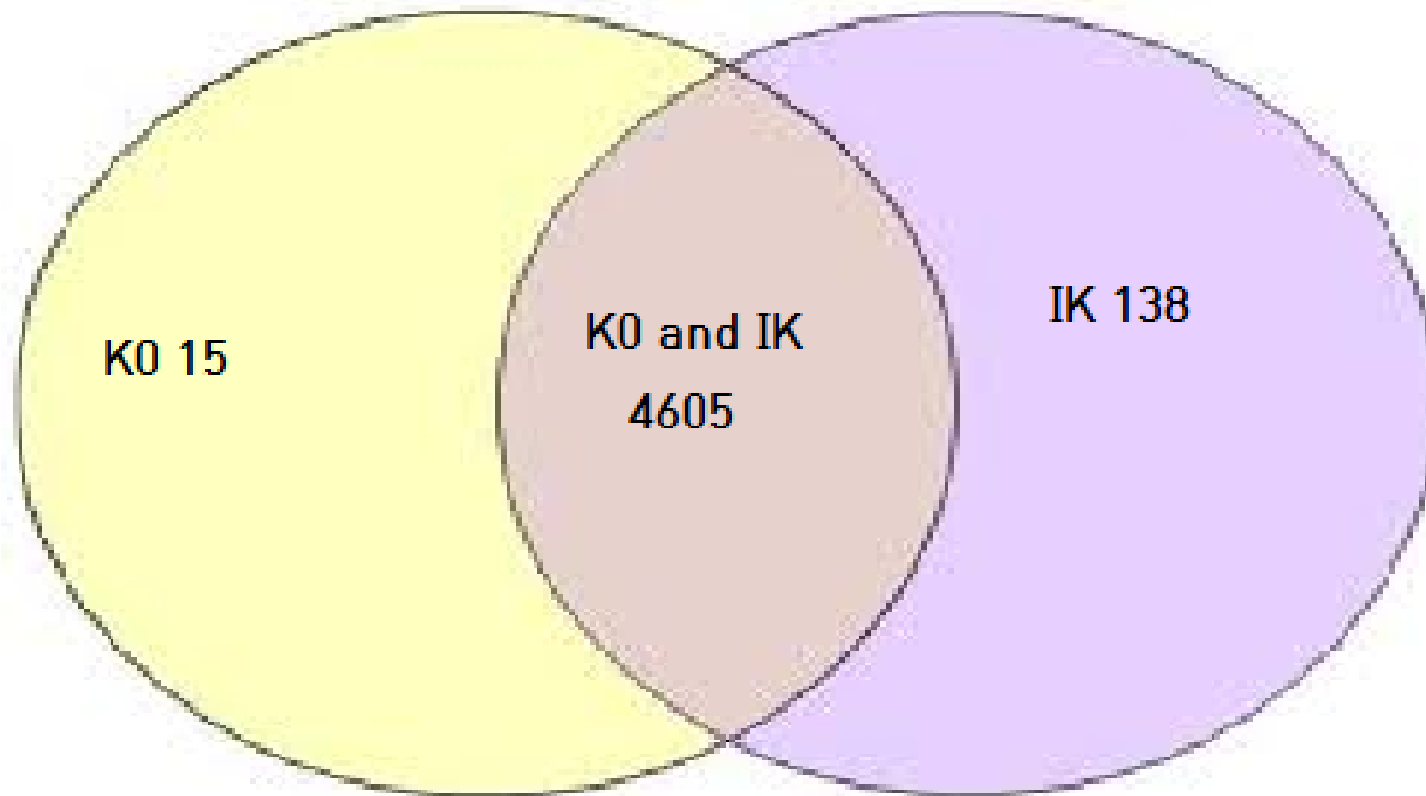Nov: CJK Coordination Meeting @ Seoul during IETF #97 (K-LGR v0.6)
Dec: 23rd KGP meeting

2017

Feb: CK coordination meeting during CDNC meeting @ Beijing, China
         K-LGR v0.7TMP (2017.02.23.)
Mar: K-LGR v0.7 (2017.03.03.): # Hanja chars 4819 -> 4758 (61 chars deleted)

12. 2013 — Organization of KGP

05. 2015 — K-LGR v0.1

06. 2015 — K-LGR v0.2

10. 2015 — K-LGR v0.3

02. 2016 — KGP formally formed

03. 2016 — K-LGR v0.4

09. 2016 — K-LGR v0.5

11. 2016 — K-LGR v0.6

02. 2017 — K-LGR v0.7TMP

03. 2017 — K-LGR v0.7

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

Update by the Thai GP

Jitti Kunphruk
Thai GP Coordinator

ICANN

Agenda

- ⊙ Background on Script and Principal Languages

- ⊙ Code point repertoire

- ⊙ Whole Label Evaluation (WLE) rules

- Thai script is an abugida script, written left-to-right, without spaces between words
- No notion of uppercase and lowercase characters
- Some vowels are written before and after the main consonant
- Certain vowels, tone marks, and diacritics are written above and below the main character

1

ISO 15924
ISO 15924 – Code: Thai
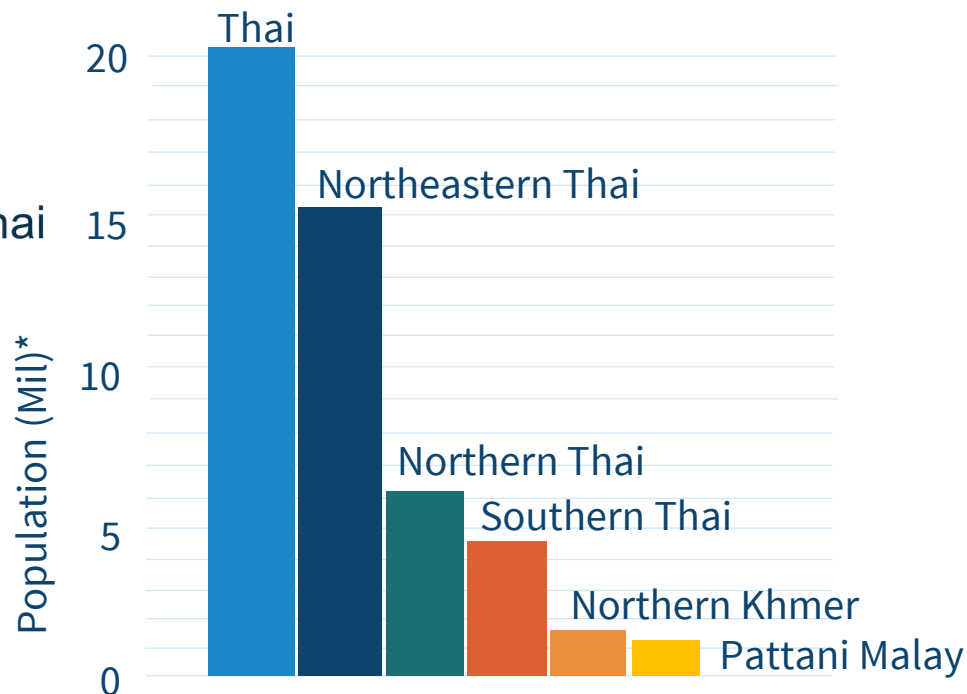ISO 15924 – Number: 352
ISO 15924 – English name: Thai

2

Unicode Range:
U+0E00 – U+0E7F

3

Writing systems that use Thai script

35 languages

Selected Languages written in **Thai script**

Thai

Northeastern Thai

Northern Thai

Southern Thai

Northern Khmer

Pattani Malay

Population (Mil)*

20

15

10

5

0

Code Point Repertoire

- The Thai GP takes code points shortlisted in MSR-2 as a starting point for Thai Script analysis for Root Zone Label Generation Rules

- The Thai GP refers to Thai script writing system from Royal Institute of Thailand and refers to various standards such as:
  - TIS 620 series – Standard for Thai Character Codes for Computers
  - TIS 820 series – Layout of Thai Character Keys on Computer Keyboard
  - TIS 1566 – Thai Input/ Output Methods for Computers

Code Point Repertoire

- There are some of code point are no longer used in the dictionary words, it is still possible to use them in names or brands.
  - ฃ ฅ ฦ
- These have also been used in the movie "คนไฟบิน" (in English: Dynamite Warrior, 2006)

- The code point U+0E26 (LU, ฦ) appears in a few Thai words and is often recognized as archaic or poetic. For example, "ฦๅสาย" (reading "Lue Sai", meaning "great man, King").

  - It is not considered obsolete as its rarity in regular words makes it more popular for use in a person names, such as "นฤเดช" (read "Na Lu Dech" meaning "Powerful Brave Man")

Code Point Repertoire

- There are a few sets of similar code points that are possibly confusable by non-native script users , but are not considered confusable by Thai Script users as they are used commonly in everyday writing.

#	Glyph	Unicode Code Point	Unicode Code Point Name
1	ก	0E01	THAI CHARACTER KO KAI
	ถ	0E16	THAI CHARACTER THO THUNG
	ภ	0E20	THAI CHARACTER PHO SAMPHAO
2	ข	0E02	THAI CHARACTER KHO KHAI
	ฃ	0E03	THAI CHARACTER KHO KHUAT
3	ช	0E0A	THAI CHARACTER CHO CHANG
	ซ	0E0B	THAI CHARACTER SO SO
4	ค	0E04	THAI CHARACTER KHO KHWAI
	ฅ	0E05	THAI CHARACTER KHO KHON
	ด	0E14	THAI CHARACTER DO DEK
	ต	0E15	THAI CHARACTER TO TAO
5	ฌ	0E0C	THAI CHARACTER CHO CHOE
	ณ	0E13	THAI CHARACTER NO NEN
6	ฎ	0E0E	THAI CHARACTER DO CHADA
	ฏ	0E0F	THAI CHARACTER TO PATAK
7	ฑ	0E11	THAI CHARACTER THO NANGMONTHO
	ท	0E17	THAI CHARACTER THO THAHAN

Whole Label Evaluation (WLE) Rules

1. No leading combining mark (default WLE)
2. Every leading vowel must precede a consonant
3. Code points which must follow a consonant
4. Context of U+0E31 (MAI HAN-AKAT)
5. Context of U+0E30 (SARA-A)
6. Context of U+0E32 (SARA-AA)
7. Context of tone mark
8. Context of diacritic

Update on Whole Label Evaluation (WLE) Rules

Context of diacritic

- An above diacritic MAITAIKHU can follow consonant, above vowel and below vowel

- An above diacritic NIKHAHIT can follow consonant, above vowel and below vowel

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

facebook.com/icannorg

weibo.com/ICANNorg

linkedin.com/company/icann

flickr.com/photos/icann

youtube.com/user/icannnews

slideshare.net/icannpresentations