ICANN **61**
COMMUNITY FORUM

SAN JUAN
10–15 March 2018

IDN Root Zone LGR Workshop

Pitinan Kooarmornpatana
IDN Program Manager

ICANN 61
14 March 2017

Overview of Session Presentations

- Update by LGR Integration Panel - Marc Blanchet

- 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

- Q/A

Update by LGR Integration Panel

Marc Blanchet
Integration Panel

Integration Panel Scope

⊙ "The Integration Panel is a panel of independent experts tasked with reviewing proposals presented by the Generation Panels and, if accepted, integrating them into a consistent set of Label Generation Rules for the Root Zone. The decisions by the integration Panel are required to be unanimous."

- ○ https://community.icann.org/display/croscomlgrprocedure/Integration+Panel

⊙ "The integration panel must take into account any public comments submitted in response to the posting of the generation panel's output."

- ○ Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels, Version 2013-03-20b, March 20th 2013.

IP Activities Summary: Review

- ⊙ GP formation proposals
  - ○ Sinhala

- ⊙ Draft LGRs (some of them multiple iterations)
  - ○ Cyrillic (final drafts)
  - ○ Korean (final drafts)
  - ○ Chinese
  - ○ Japanese
  - ○ Neo-Brahmi
    - • Devanagari
    - • Kannada
    - • Gurmukhi
    - • Gujarati
    - • Telugu

IP Activities Summary: Other LGR Related

- Discussions regarding homoglyphs
  - Latin GP

- Reviewed LGRs after their public comment period
  - Cyrillic

- Han variants analysis
  - Significant amount of work for IP
  - Various sources of data and usage
    - Sets are not necessarily similar depending on the reference

- Started preliminary (prototyped) integration of expected upcoming LGRs, such as CJK and Neo-Brahmi
  - Enables to find issues before doing real integration for LGR-3
  - Large scripts, many cross-script variants

IP Activities Summary: MSR-3

- ⊙ MSR-3
  - ○ Call for proposals for additions/changes to MSR
    - • Issued Nov 25th 2017
    - • Deadline Jan 15th 2018
  - ○ Goal is to do a minor update
    - • No additional scripts
    - • Addition of code points needed for some scripts
  - ○ Received the following requests
    - • From Japanese GP: U+20B9F (叱) (Joyo-Kanji)
    - • From Latin GP: U+0268 (ɨ), U+0272(ɲ), U+01C0 (ǀ), U+01C1 (ǁ), U+01C2 (ǂ), U+1E3D (ḽ)
      U+01C3 (ǃ) mentioned but not requested
    - • From CGP: U+3A5C (㩜) and U+58B5 (墵)

IP Activities Summary: MSR-3 (cont.)

- ⊙ MSR-3
  - ○ Result
    - • Addition of three code points in Hani:
      U+20B9F (𠮟), U+3A5C (㩜) and U+58B5 (墵)
    - • Addition of three code points in Latin:
      U+0268 (ɨ), U+0272(ɲ), U+1E3D (ḽ)
    - • Did not add:
      U+01C0 ǀ LATIN LETTER DENTAL CLICK
      U+01C1 ǁ LATIN LETTER LATERAL CLICK
      U+01C2 ǂ LATIN LETTER ALVEOLAR CLICK
      U+01C3 ǃ LATIN LETTER RETROFLEX CLICK
      Security concerns: they are related to non-PVALID/punctuation
  - ○ Public comments started on 15 January 2018, deadline 26 February 2018

- ⊙ MSR-4
  - ○ Currently not scheduled. Might include additional scripts when needed.

IP Activities Future Work

⊙ Depending on timely delivery of script LGRs (including public comments), Integration Panel shall be producing a new Root Zone LGR (LGR-3)

  ○ Target: Q3 CY2018

⊙ Three groups of LGRs are considered (not all in a single release)

  ○ Han (CJK)

  ○ Neo-Brahmi

  ○ Latin-Cyrillic-Greek-Armenian

Update on LGR Toolset

Audric Schiltknecht
Julien Bernard
Marc Blanchet
Viagénie

Contents

- ⊙ LGR toolset summary in one slide

- ⊙ New features in 2018

LGR Toolset Summary

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

- ⊙ Available(*)
  - ○ Open source
  - ○ Online as a service

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

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

New Release in 2018

- ⊙ New Features
  - ○ Python 3
  - ○ MSR-3
  - ○ Significantly improve performance (especially for large LGRs)
  - ○ Check if multiple LGRs are harmonized: variant code points are transitive and symmetric in each LGR and also variant code points or code point sequences in one LGR cannot be non-variants in another
  - ○ Print the specific rule for which the label fails
  - ○ Print the line number with error message when the syntax of an XML is incorrect
  - ○ Automatically add when assigning a variant code point from another script

New Release in 2018

- ⊙ User Interface/Experience Improvement
  - ○ Button to populate symmetric and transitive variants
  - ○ 'Tag' page for tag management
  - ○ Function to assign code points for each WLE rules
  - ○ Select a script and automatically populates the code point range
  - ○ Summary output more readable
  - ○ Update the variant table heading to show both the number of members and number of mappings
  - ○ Links in "Table of References" as hyperlinks in HTML output
  - ○ View all forms of a label: A-label, U-label and Unicode sequence
  - ○ Reduce rows in the variant table by using arrows in HTML output
  - ○ Improve display the combined form of the code point sequence

New Release in 2018

- ⊙ Fix Bugs
  - ○ Variant type cannot be changed from the default "block"
  - ○ The default type should be 'blocked'
  - ○ Incorrect tag count in character classes
  - ○ Correctly validate label for unexpanded ranges
  - ○ Annotation tool on a test file does not display correctly the first label
  - ○ Variant label generated should also be validated based on existing rules in an LGR
  - ○ Update HTML output to show number of code points in the largest set

New Release in 2018

- ⊙ Planned for June release

- ⊙ More information for LGR toolset:
  https://www.icann.org/resources/pages/lgr-toolset-2015-06-21-en

- ⊙ Link to the toolset:
  https://www.icann.org/en/system/files/files/lgr-toolset-user-guide-24oct17-en.pdf

- ⊙ Link to the user manual:
  https://www.icann.org/en/system/files/files/lgr-toolset-user-guide-24oct17-en.pdf

Community Updates

- Chinese GP
- Japanese GP
- Korean GP

Chinese Generation Panel (CGP) Update

Kenny Huang
Wei Wang
Co-chairs of Chinese GP

Agenda

1 CGP Work Review

2 CJK Coordination

3 CGP Proposal Draft 201712

4 Feedback from IP

5 Next Step
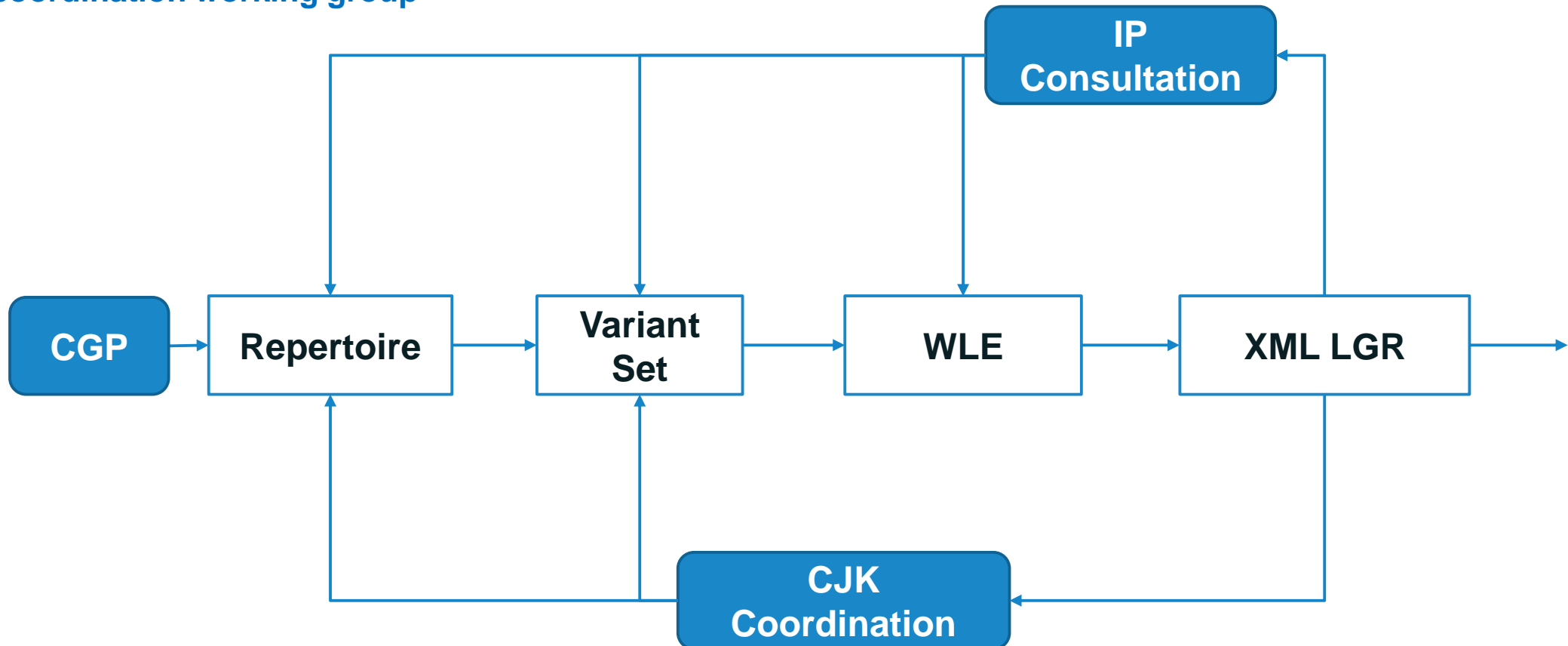
1. CGP Team and Work Process

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

1. CGP Work Review

CJK Meeting@Beijing

CJ Meeting@Dallas
CDNC Meeting@Tapei
CJK Meeting@Seoul

CK Meeting@Taipei
CK Meeting@Seoul

Repertoire Size
Allocatable Labels number

| Sep 2014 | Mar 2015 May 2015 Jul 2015 Oct 2015 | Mar 2016 Jun 2016 Aug 2016 | Sep 2016 Nov 2016 Sep 2016 Nov 2016 | Jan 2017 Jul 2017 | Dec 2017 Feb 2018 |

CGP Formed

CGP Proposal Draft v1
IP Feedback

Proposal v8
Proposal v9

Proposal v10
&
IP Feedback 201802

CGP Repertoire &
Variant Extension Review

CGP Proposal Draft v2
IP Feedback

Next Steps

2. CJK Coordination

Script and Languages Covered

Language	ISO 15924 Code	Countries	Local Names of the Script
Chinese	cdo, cjy, 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
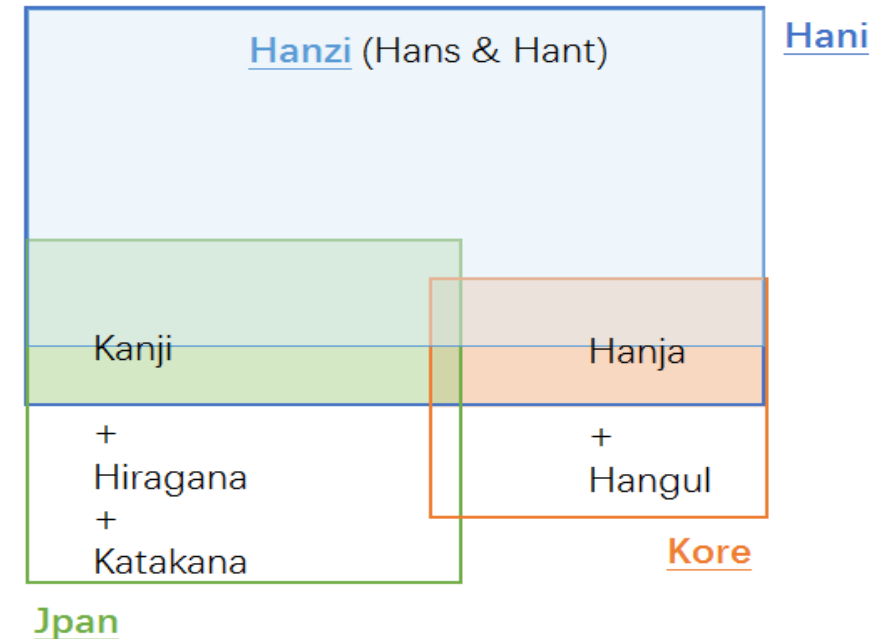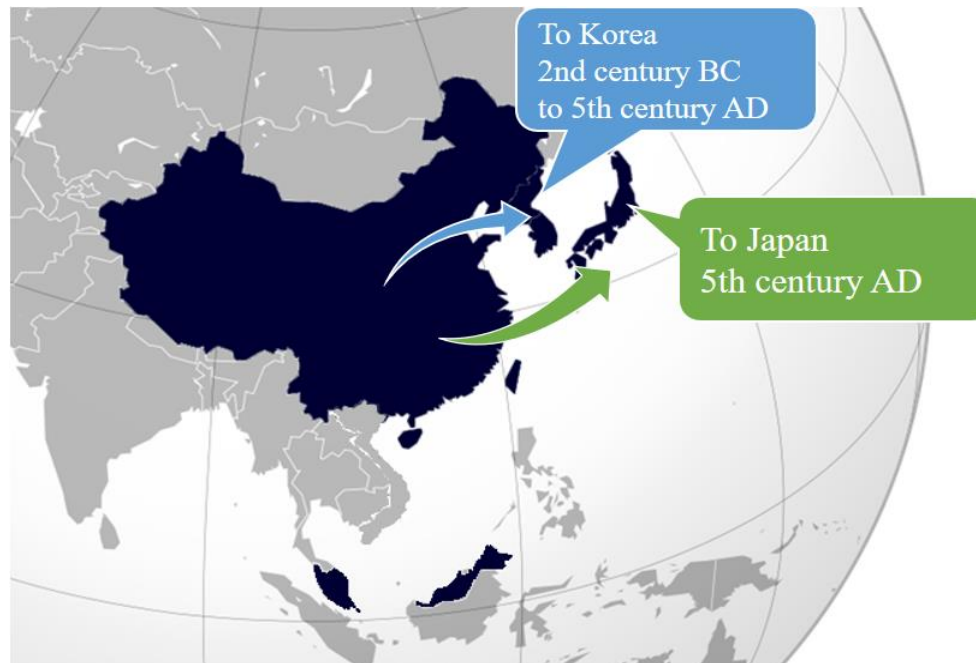
2. CJK Coordination

Coordination within CGP

⦿ CDNC variants

⦿ dotAsia variant coordination (69)

⦿ TGSCC and IICORE variants review (172)

TGSCC 10986

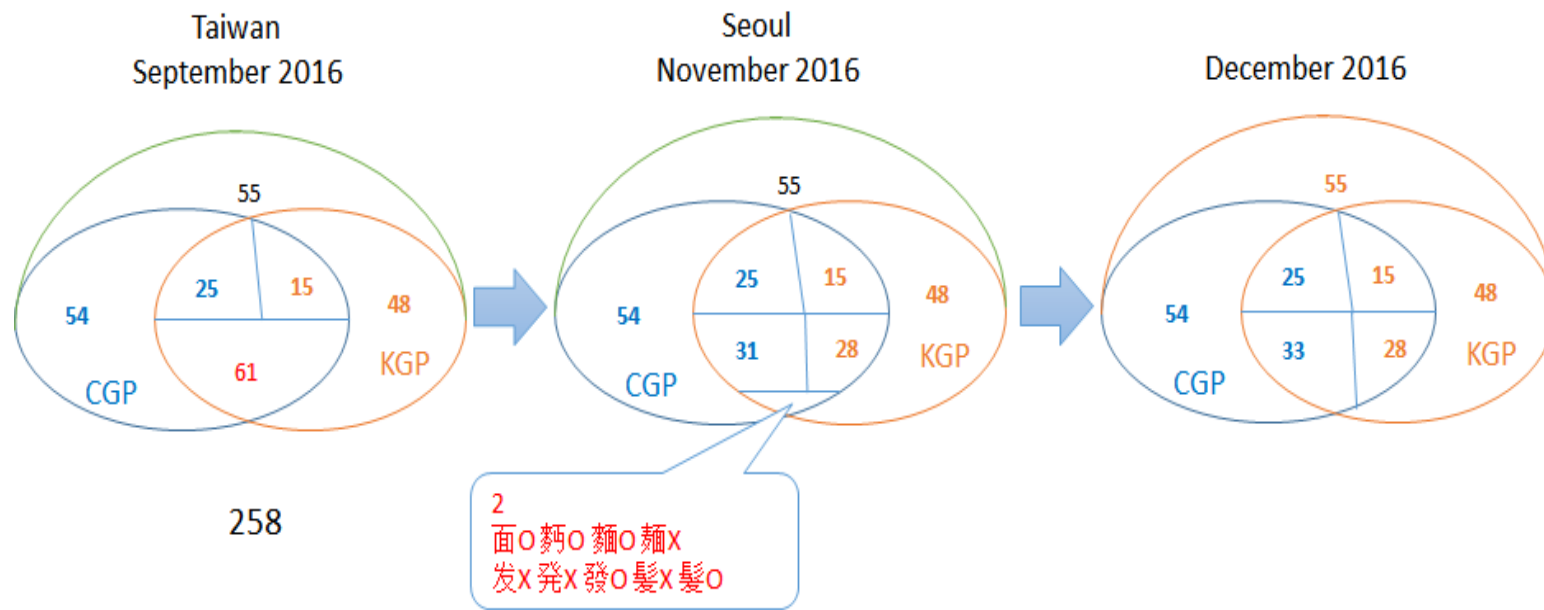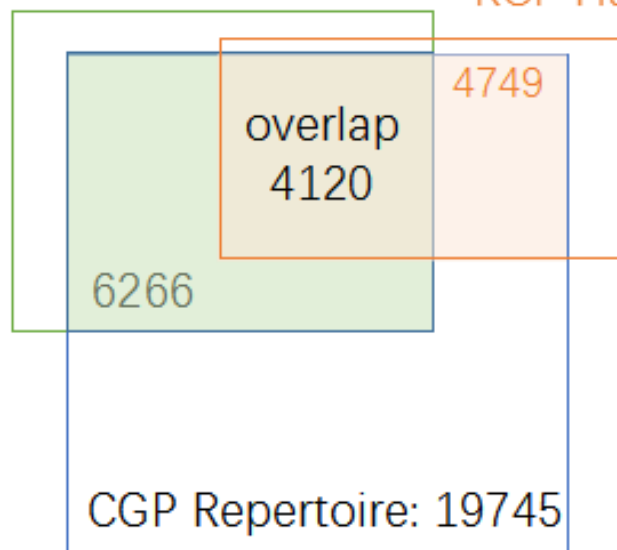dotAsia 19683

19559

overlap 10959

CDNC 19561
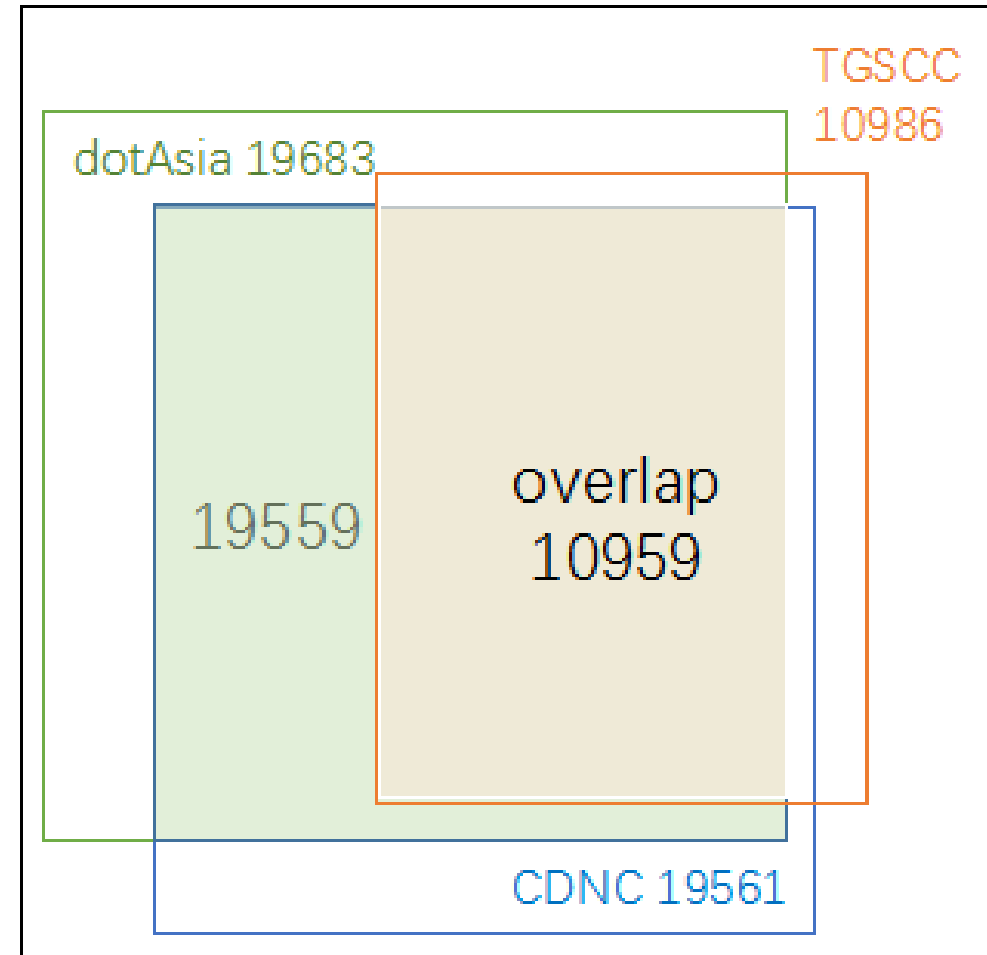
CGP Repertoire 19745

2. CJK Coordination

Coordination between C, J and K

⊙ 445 variant mappings (146/258 unacceptable variant groups)

3. CGP Proposal Draft 201712

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

Origin	Simp	Trad	All Variant
愛(611B)	爱(7231)	愛(611B)	愛(611B)爱(7231)

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

Code Point

Sub-Type	Type	Comment
out-of-repertoire-var	Invalid	Non-CGP chars imported from other GPs

MSR2

I I CORE ∩ JGP **42**

TGSCC 18

dotAsia 124

CDNC
19559+2
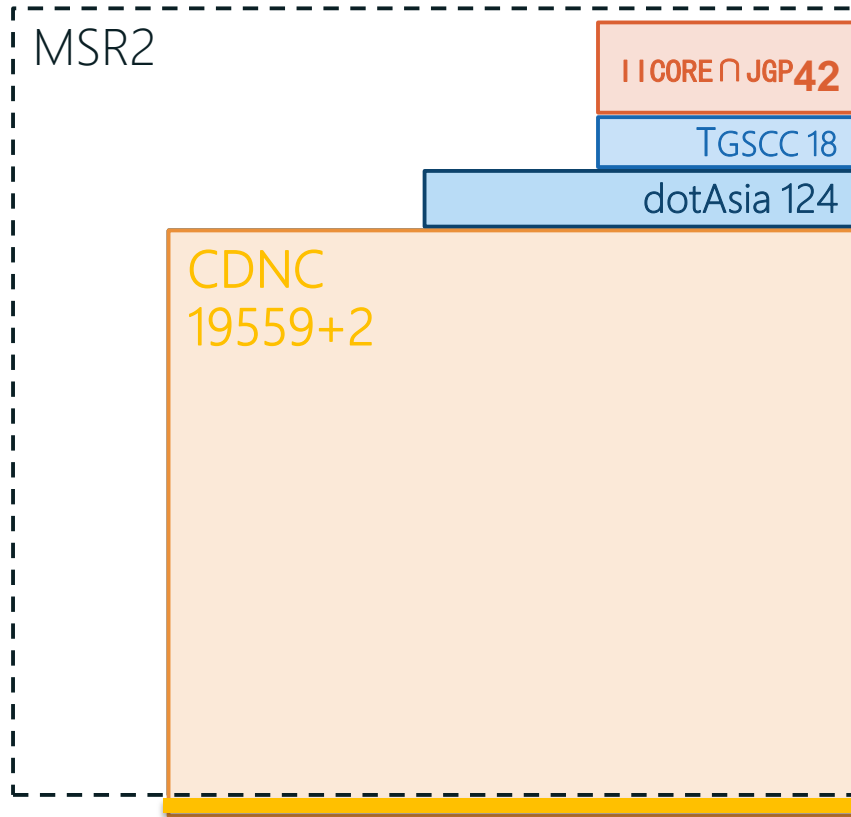
19745 Code Points

Two characters were added into MSR-3

3. CGP Proposal Draft 201712

Original Char	Preferred Simplified	Preferred Traditional
夥 (5925)	夥(5925)伙(4F19)	夥(5925)
升 (5347)	升(5347)	升(5347)昇(6607)陞(965E)
台 (53F0)	台(53F0)	台(53F0)檯(6AAF)臺(81FA)

The 136 multiple allocatable variant mappings are divided into 7 categories:

#	Number	Original	Allocatable Simp	Allocatable Trad
1	2	A	A (muting, reflexive), B	A
2	1	A	A (muting, reflexive), B	C
3	103	A	A	A (muting, reflexive), B
4	23	A	A	B (trad-1, no mixed with trad-2), C (trad-2, no mixed with trad-1)
5	5	A	A	A (muting, reflexive) B (trad-1, no mixed with trad-2) C (trad-2, no mixed with trad-1)
6	1	A	B	C (trad-1, no mixed with trad-2) D (trad-2, no mixed with trad-1)
7	1	A	A	A (muting, reflexive) B (muting, not in Modern Chinese Common Used Table in China mainland, nor Common used Chinese standard table in Taiwan) B (trad-1) C (trad-2)

3. CGP Proposal Draft 201712

Sub-Type	Type	Comment
"r-both-ms"	Allocatable	Reflexive preferred simp and trad variant chars are the same Reflexive preferred simp is one of multiple simp variant mappings Set reflexive preferred simp as "r-both-ms"
"r-both-mt"	Allocatable	Reflexive preferred simp and trad variant chars are the same Reflexive preferred trad is one of multiple trad variant mappings Set reflexive preferred trad as "r-both-mt"
"r-simp-m"	Allocatable	Reflexive preferred simp is one of multiple simp variant mappings Set reflexive preferred simp as "r-simp-m"
"trad-m"	Allocatable	Preferred trad is rare used, not in Modern Chinese Common Used Table in China mainland, nor Common used Chinese standard table in Taiwan Set the preferred trad as "trad-m"
"trad-1"	Allocatable	Among the multiple preferred traditional variants Set the preferred trad with the smallest hex-code as "trad-1"
"trad-2"	Allocatable	Among the multiple preferred traditional variants Set the preferred trad with the largest hex-code as "trad-2"

<action disp="allocatable" only-variants="simp r-simp both r-both r-both-mt" comment="all simplified label" />
<action disp="allocatable" only-variants="trad r-trad both r-both r-both-ms trad-1" comment="all traditional label category one"/>
<action disp="allocatable" only-variants="trad r-trad both r-both r-both-ms trad-2" comment="all traditional label category two"/>
<action disp="blocked" any-variant="simp trad both trad-m trad-1 trad-2" comment="block any other mixed labels" />

<action disp="allocatable" all-variants="r-neither r-trad r-simp r-both r-both-mt r-both-ms" comment="original label" />

4. Feedback from the IP

- ⊙ Further clarify the need to include all 18 characters of the TGSCC set by showing usage in Chinese context
- ⊙ Provide references to all variant mappings using available sources such as Unihan, dotAsia, and any other relevant sources, using the 'ref' attribute on the 'var' element
- ⊙ Review variant sets that differ from second-level practice and provide a rationale for any differences, particularly in cases not forced by alignment with KLGR

| Source	Glyph	Target	Glyph		Type(s)	Ref	Comment
3A52	撎	3A52	撎	≡	r-trad		identity
3A52	撎	64D2	擒	→	simp		
				←	blocked		
64D2	擒	64D2	擒	≡	r-both		identity

The code point U+3A52 was included in CLGR10 because of its membership in IICORE HKSCS but is treated differently between CLGR10 (table follows) and dotAsia (where it is a singleton reflexive variant of "r-both").

The simplified mapping between U+3A52 and U+64D2 in CLGR10 is not supported by Unihan and looks doubtful. Unihan Definition field for U+3A52 indicates that this is a variant of U+64D2 but without simplified mapping. In Unihan U+64D2 has itself a semantic variant relationship with U+6366 㕦, not supported by either CLGR10 or dotAsia.

50+ variant groups might need further investigation and review

ICANN

5. Next Step

- ⊙ To keep communicating with IP on the remaining issues

- ⊙ To include Unihan as reference source for code points and variant mappings

- ⊙ To evaluate whether the 18 TGSCC characters and 42 imported JGP characters could be removed, to make CGP a compact union set of CDNC and dotAsia only

- ⊙ To generate a reduced repertoire and variant mappings table after removing the above 60 characters

- ⊙ KGP LGR Proposal might be affected

- ⊙ To start a review process on the variant groups which have different variant mapping relationships in CDNC and dotAsia. The review work might be done by the CDNC & CGP joint meeting late April.

MSR2

~~JGP Lffect 42~~

~~TGSCC 18~~

dotAsia 124

CDNC 19559+2
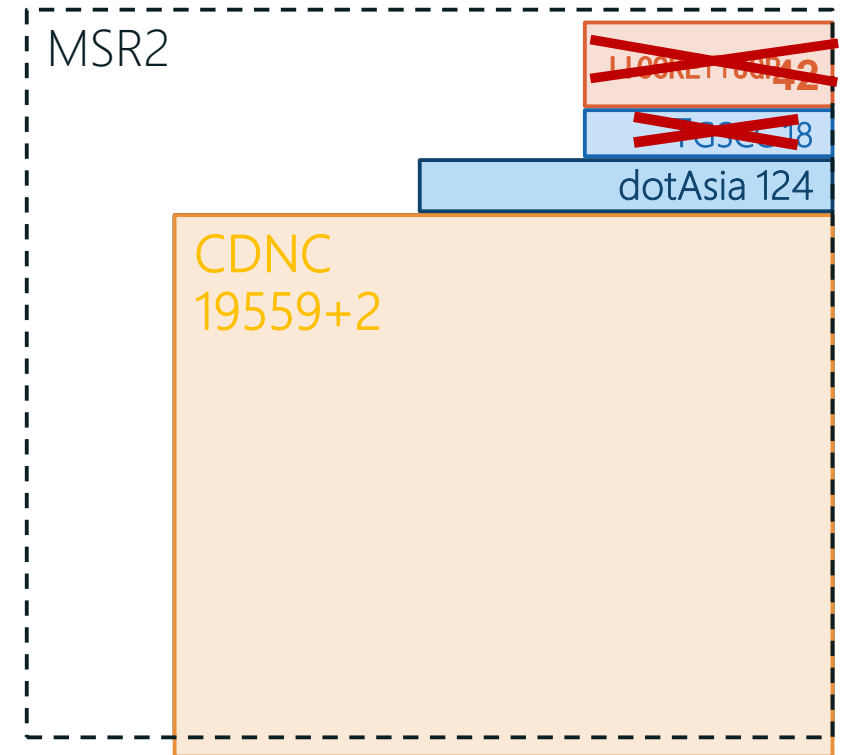
THANK YOU

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

Japanese Generation Panel (JGP) Update

Hiro Hotta
Chair of Japanese GP

What Japanese Generation Panel (JGP) 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

we are here!

Development from November 2017 is written in red

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	Member	Internet in general
⊙	Kazunori Konishi	Member	Internet in general
⊙	Tsugizo Kubo	Member	Trademarks and domain names
⊙	Yoshitaka Murakami	Member	Trademarks and gTLD markets from registry/ registrar perspective
⊙	Shuichi Tashiro	Member	Character codes
⊙	Yoshiro Yoneya	Member	Technical aspects of IDN, LGR

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

- ⊙ Whole Label Evaluation (WLE) rule
  - ○ Japanese LGR may have no or very limited number of tiny rules for the usage of characters even if defined

CJK Coordination
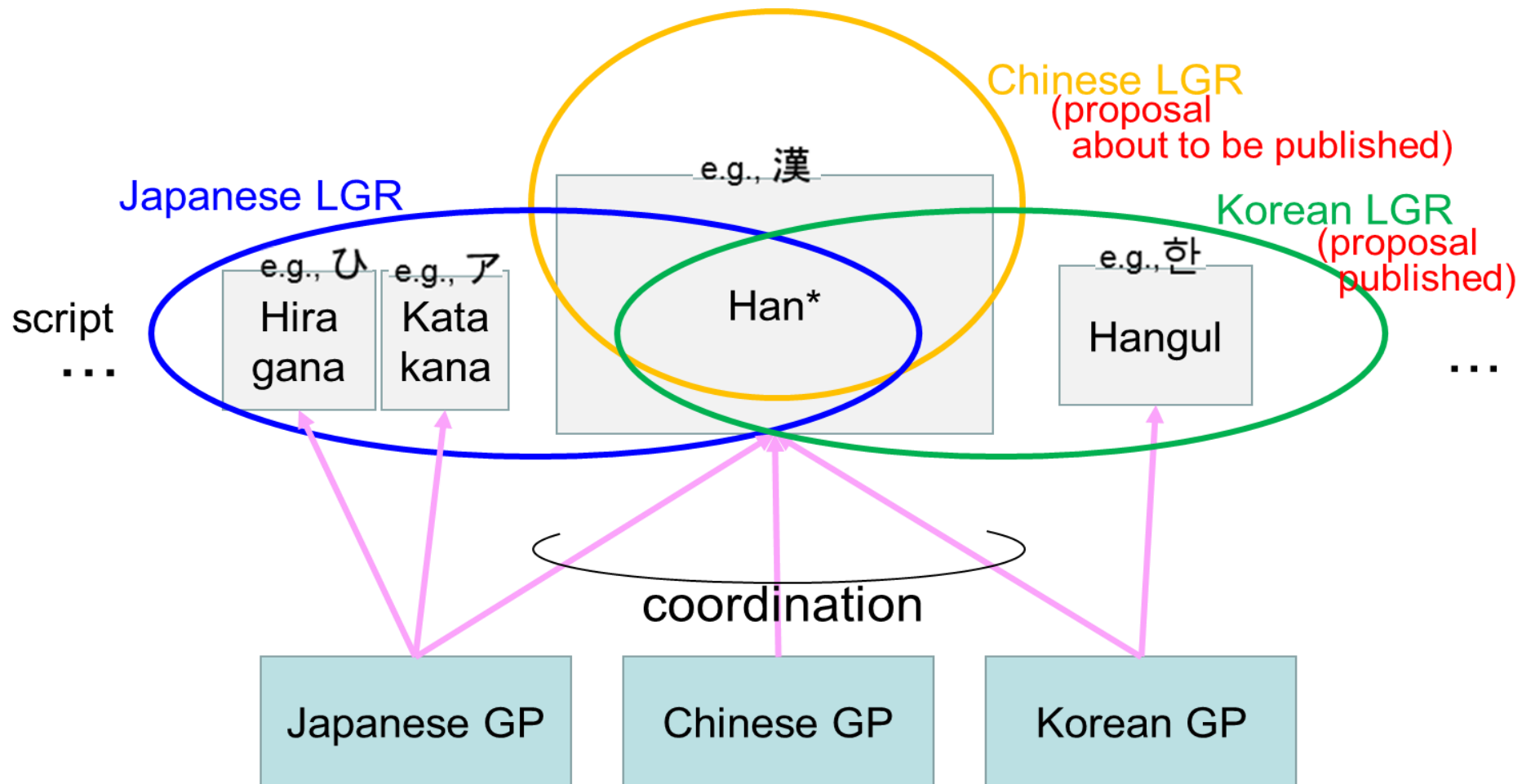
Coordinated definition of variants has been completed

Reduction of the Number of Allocatable Labels

- Any combination of characters is allowed as Japanese labels
- It may make the number of variant strings very huge, as many variant groups are defined by importing Chinese variants
- Until now, imported variant definitions have become stable
- E.g., 慶応大学 has three 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
  - Case1 : day-use Kanji
  - Case2 : day-use kanji and Kanji designated for personal names
- JGP is considering case1 works fine in reducing the number of allocatable labels
  - It reduces the maximum number of allocatable labels of actually registered Japanese labels under .JP from 486 to 8
- XML of case1 has been developed for inspection by IP

Handling Similar - Looking Characters

- ⊙ Any combinations of Japanese characters are used to express Japanese words
- ⊙ So, the initial intention of the definition of Japanese labels was "there are no variants in Japanese LGR", except those imported variant definitions of C-LGR and K-LGR
- ⊙ However, IP proposes the following two kinds of variants
  - ○ [Kind A] punctuation characters "U+30FC (ー) and U+4E00 (一)" and "U+30FD (ヽ) and U+4E36 (丶)" are proposed to be variants because they have confusable glyphs that are single stroke characters
  - ○ [Kind B] mutually resembling Kanji characters and Kana characters, such as "カ (Katakana) and 力 (Kanji)" and "オ (Katakana) and 才 (Kanji)", are proposed to be variants
- ⊙ Kind A is considered acceptable, but Kind B is not because all Japanese characters are considered very unique

Thank You and Questions

Korean Generation Panel (KGP) Update

Kim Kyongsok
Chair of Korean GP

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
(Coordination between KGP and CGP)

5

Work Achieved To-Date 3
(Brief History of KGP Activities)

6

Future Plan and Schedule

1. Script(s) Covered by K-LGR and Where They are Used

- ⊙ K-LGR covers Korean script (= Hangul + Hanja)

- ⊙ "Korean script" usually means "Hangeul" or "Hangul". However, in the context of the Korean LGR (K-LGR), Korean script is a union of Hangul (한글) and Hanja (한자).

- ⊙ Korean language has a long history, more than 2000 years

- ⊙ Hangul: invented in 1443

- ⊙ Hanja was used before Hangul was invented. Hanja is still used in Rep. of Korea

- ⊙ Korean language is mainly used in 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

2. 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, Sinyeong PARK

- ⦿ Registration Agency: Seong-jin PARK, ChangKi JANG, Myungsoo LEE

- ⊙ 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

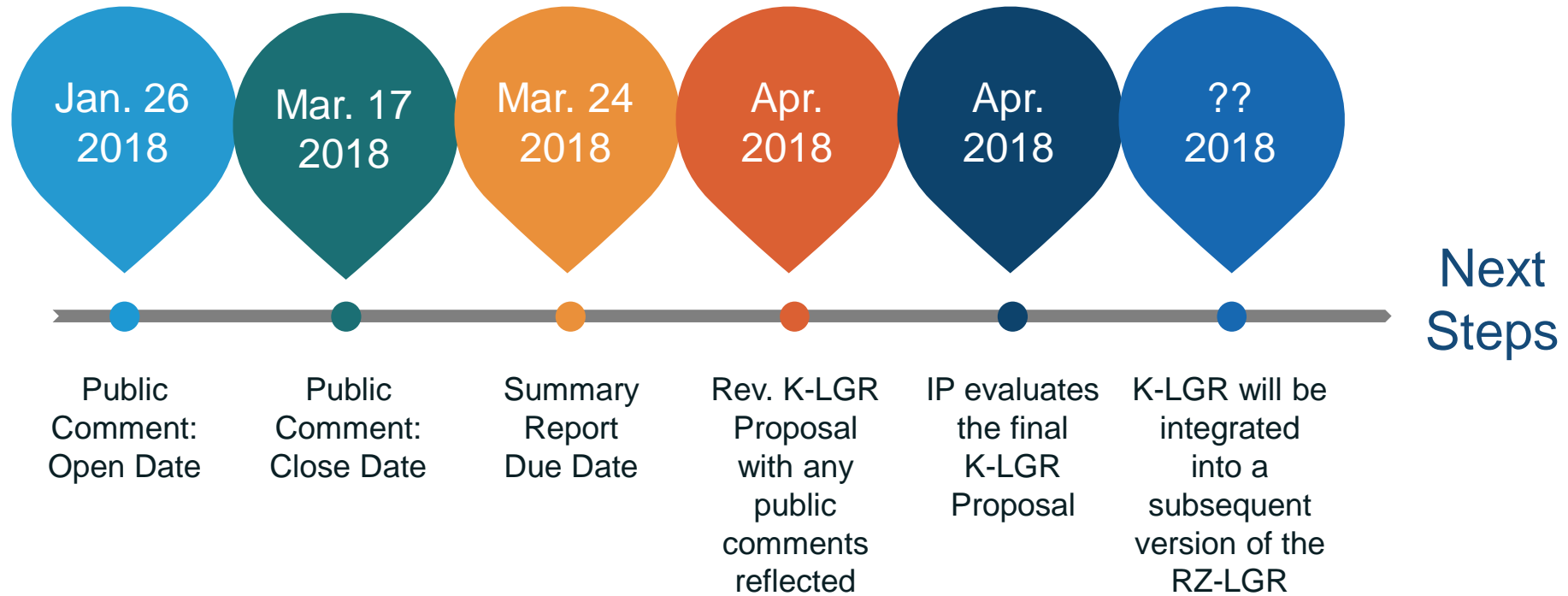| Source of Hanja Character Set	# chars
1) KS X 1001 (268 comptb. chars excluded)	4620
2) IICORE - K column marked	4743
K-LGR v1.0 (2017.12.10.): Hanja List (Union of 1) and 2))	4758

⊙ Unification of variant groups for Hanja between KGP and CGP (Chinese GP)

   ○ 304 Chinese variant groups reviewed by KGP and CGP

   ○ 304 Chinese variant groups: contain 2 or 3 K Hanja chars
     (K Hanja char: Hanja char included in K-LGR Hanja repertoire)

⊙ Result of Unification of variant groups between KGP and CGP

   ○ K-LGR v1.0: 152 variant groups containing 2 or 3 K Hanja chars

   ○ The other Chinese variant groups were split so that no more than 1 K char in Chinese
     variant group

   ○ No conflict in variant groups between K-LGR and C-LGR

5. Work Achieved 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

- 26 KGP meetings

- Several CJK coordination meetings during ICANN meetings 49 ~ 60

- Several CJK coordination meetings in Rep. of Korea, China, and Taiwan

6. Future Plan

Jan. 26 2018	Mar. 17 2018	Mar. 24 2018	Apr. 2018	Apr. 2018	?? 2018
Public Comment: Open Date	Public Comment: Close Date	Summary Report Due Date	Rev. K-LGR Proposal with any public comments reflected	IP evaluates the final K-LGR Proposal	K-LGR will be integrated into a subsequent version of the RZ-LGR

Next Steps

Thank You and Questions

Engage with ICANN and IDN Program

Thank You and Questions

Visit us at **icann.org/idn**
Email: IDNProgram@icann.org

@icann

facebook.com/icannorg

youtube.com/icannnews

flickr.com/icann

linkedin/company/icann

slideshare/icannpresentations

soundcloud/icann