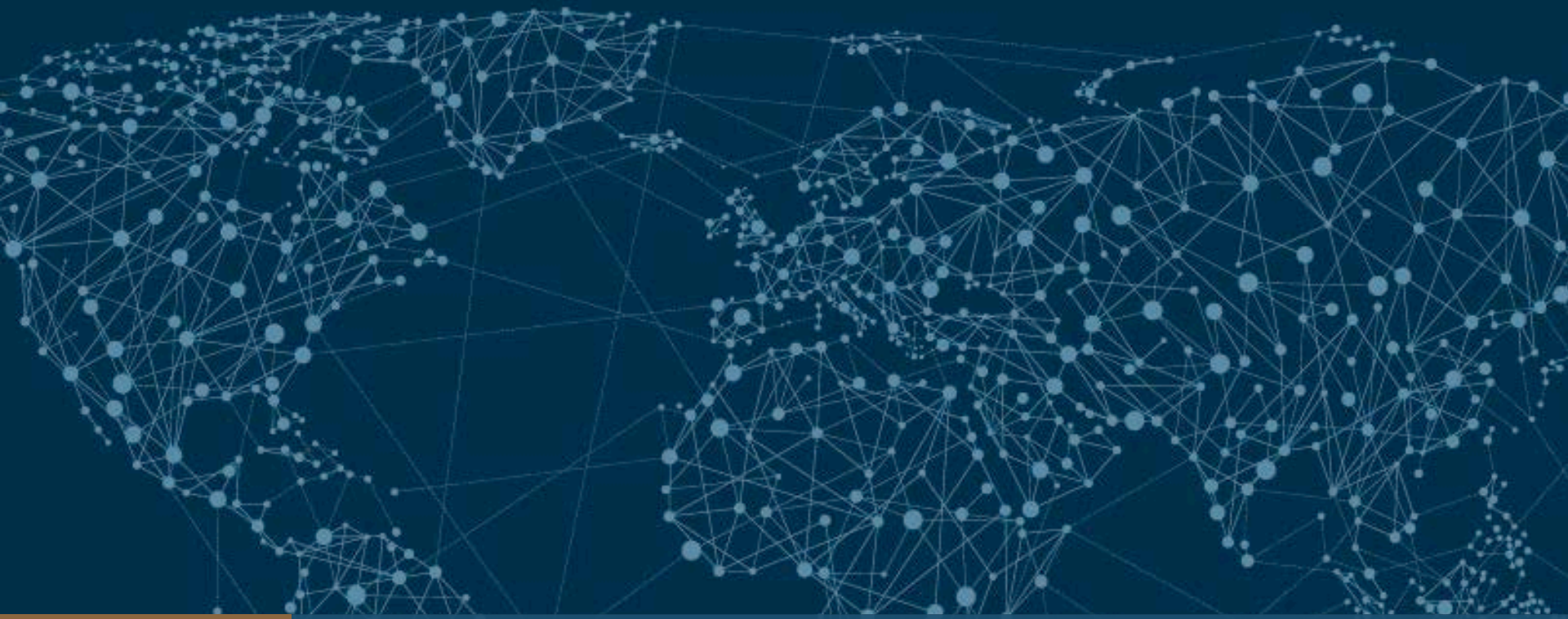

ICANN
ICANN | 53
Buenos Aires
21-25 JUNE 2015



IDN Program Update

Sarmad Hussain | IDN Program Sr. Manager | 24 June 2015

Agenda

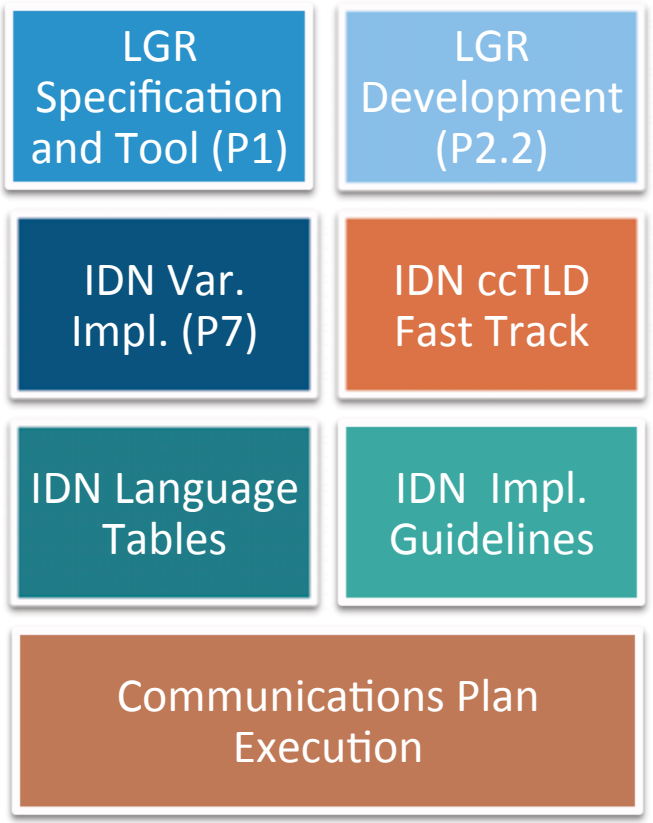
- **IDN Program - Overview and Progress** – Sarmad Hussain
- **Integration Panel Update** – Integration Panel
- **Community Updates:**
  - Arabic Generation Panel – Meikal Mumin
  - Armenian Generation Panel – Igor Mkrtumyan
  - Japanese Generation Panel - Hiro Hotta
  - Korean Generation Panel - Kyongsok Kim
- **Q&A**

IDN Program – Overview and Progress

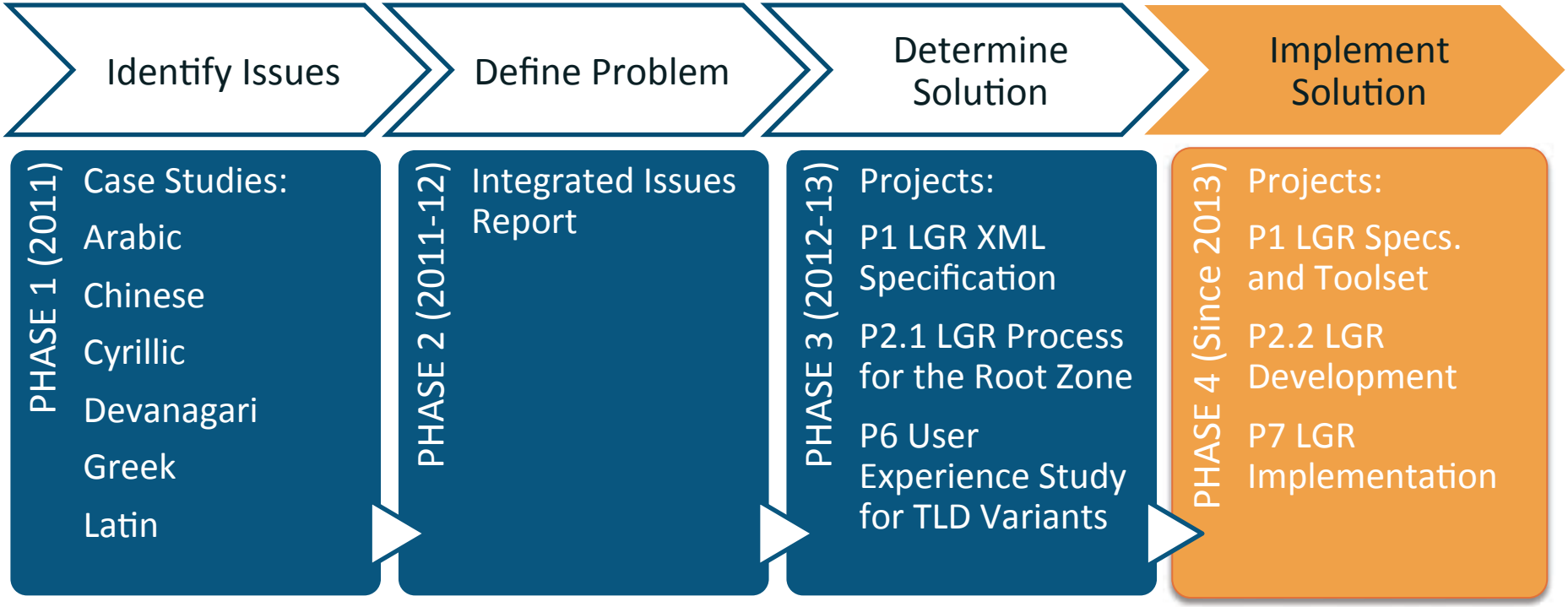
Sarmad Hussain
IDN Program Senior Manager
ICANN

Overview of Presentation

- IDNs at Top Level
  - IDN TLD Program
    - Label Generation Ruleset (LGR)
    - LGR Toolset
    - IDN Variant Implementation
  - IDN ccTLD Fast Track Process Implementation
- IDNs at Second Level for gTLDs
  - IDN Implementation Guidelines
  - IDN Tables
- Community Outreach and Involvement



IDN TLD Program



Community agreed to define Label Generation Rules (LGR).

Reports and documentation of all completed projects available at: https://www.icann.org/resources/pages/reports-2013-04-03-en.

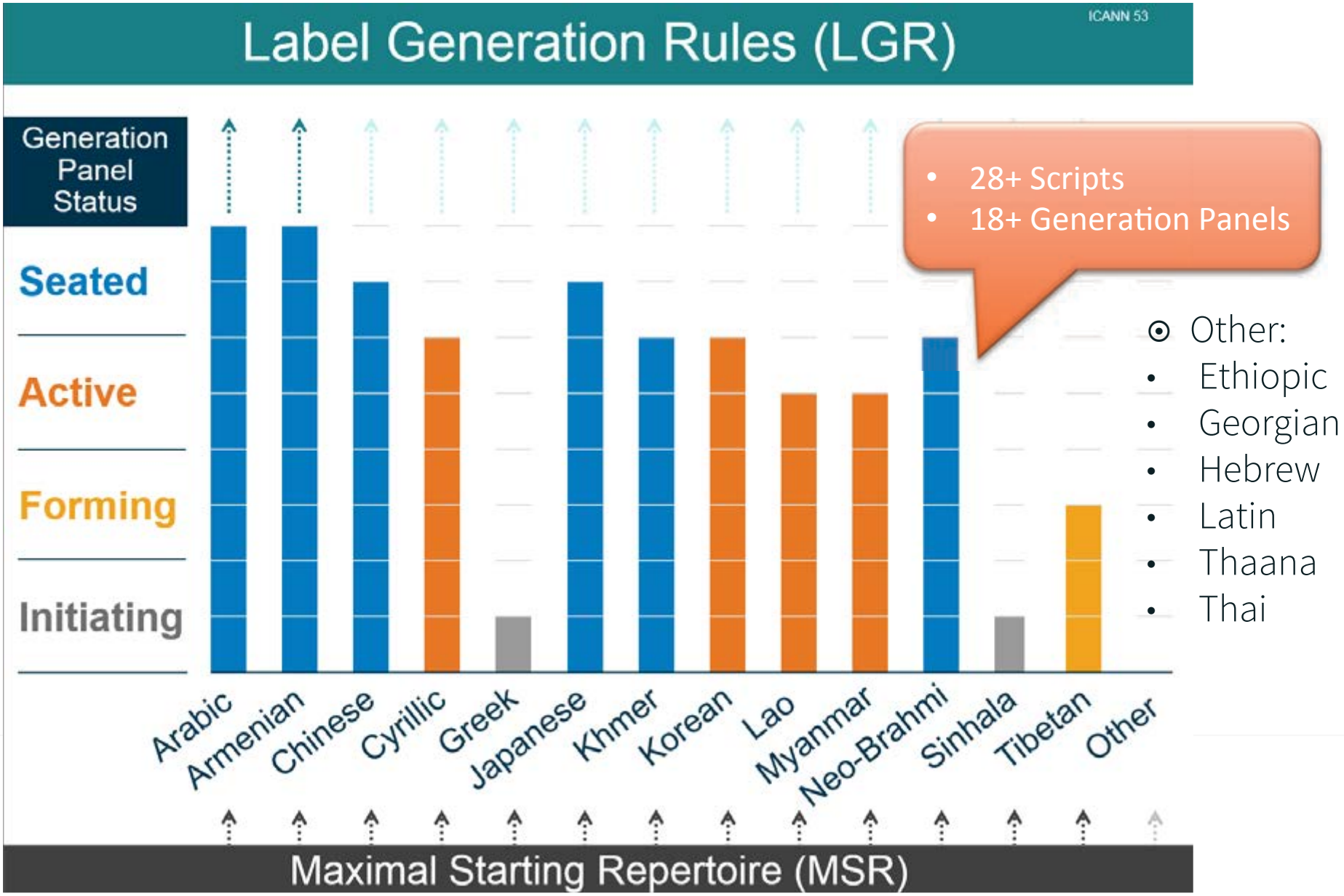
Maximal Starting Repertoire (MSR-2)

- MSR-2 released on 27 April 2015
- Total 28 scripts, including six additional scripts
  - Armenian, Ethiopic, Khmer, Myanmar, Thaana, Tibetan
- Total 33,490 code points shortlisted from 97,973 candidates
  - Additional 700 from 936 candidate code points in the six scripts
- Based on Unicode 6.3
- Upwardly compatible with MSR-1

Guidelines for Generation Panels

- Guidelines released for GPs on 27 April 2015
- https://www.icann.org/public-comments/lgr-guidelines-2015-01-15-en
  - Guidelines for Developing Script-Specific Label Generation Rules for Integration into the Root Zone LGR
  - Considerations for Designing a Label Generation Ruleset for the Root Zone
  - Requirements for LGR Proposals
- Technical documents:
  - Variant Rules
  - Whole Label Evaluation (WLE) Rules
  - Representing Label Generation Rulesets using XML

Status of LGR Development

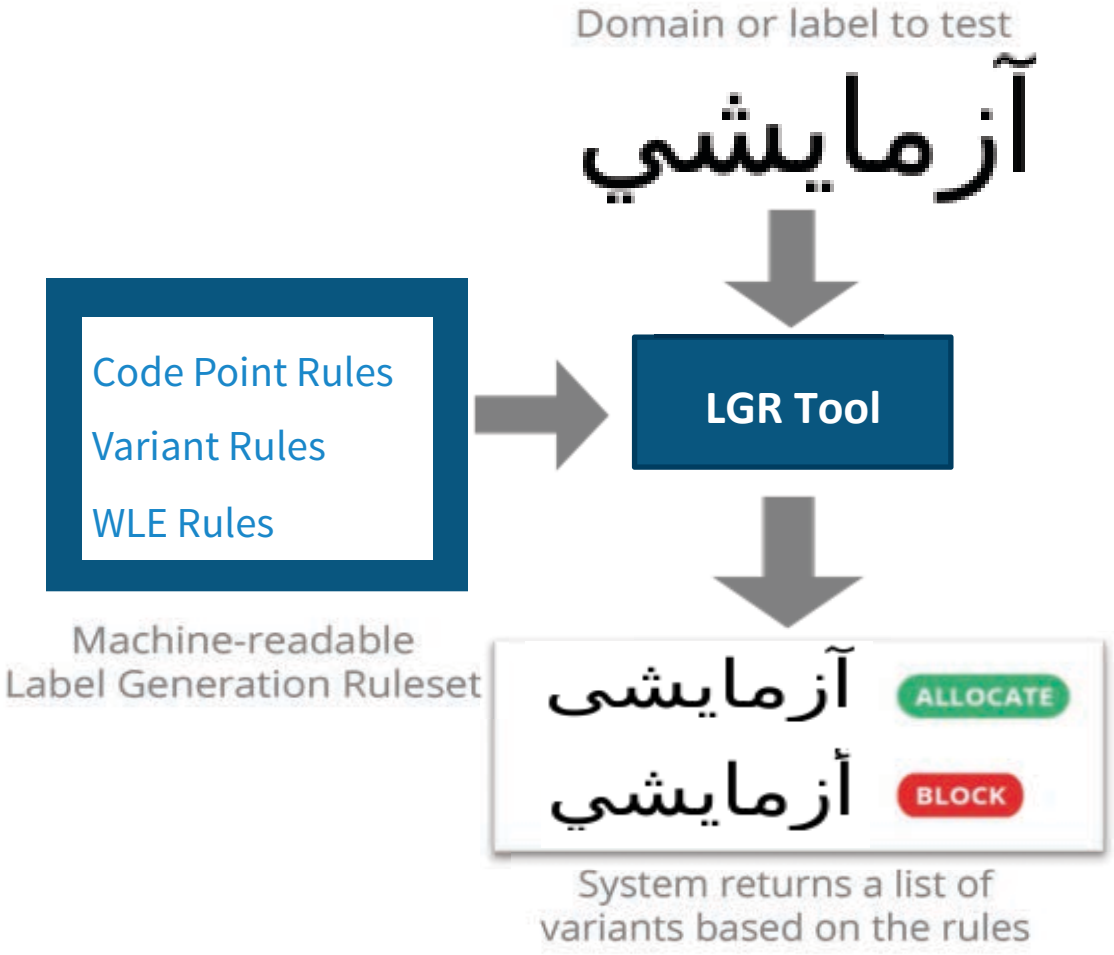


LGR Specification and Toolset

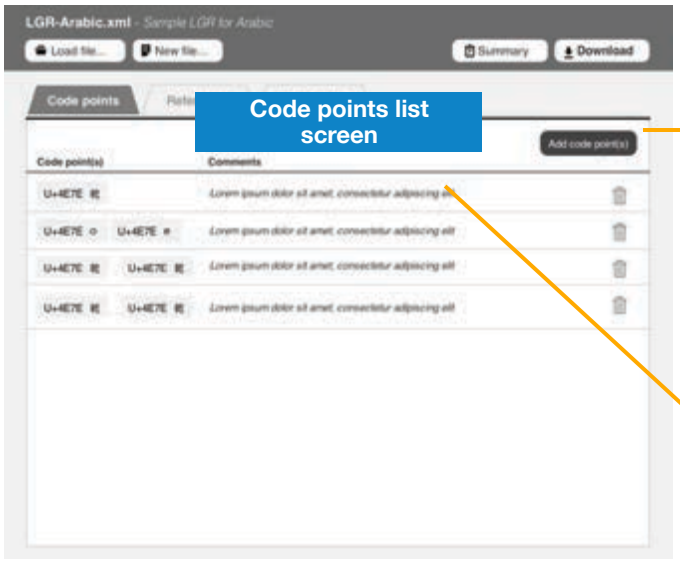
- LGR machine-readable specifications at:
  https://datatracker.ietf.org/doc/draft-davies-idntables

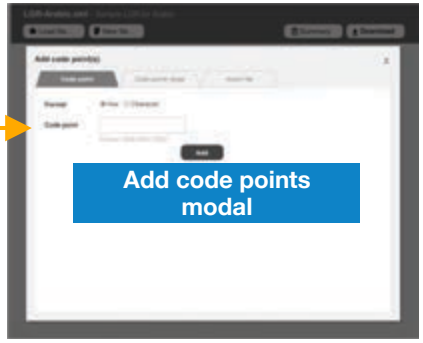
- Public Comment on tool requirement: 3 Dec. 2014

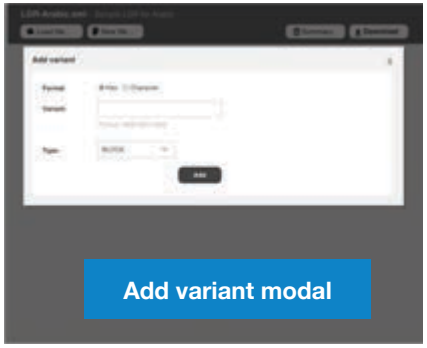
- Toolset tentative timeline:
  - Create LGR – 09/15
  - Use LGR – 12/15
  - Manage LGRs – 3/16
  - Open source

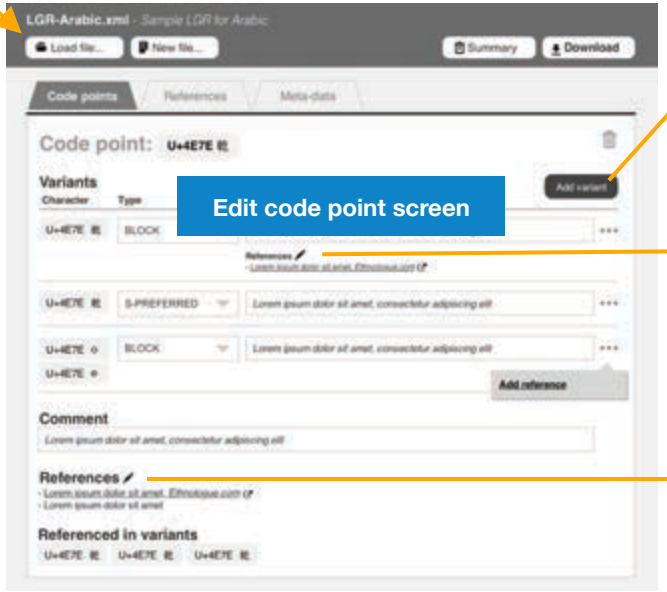


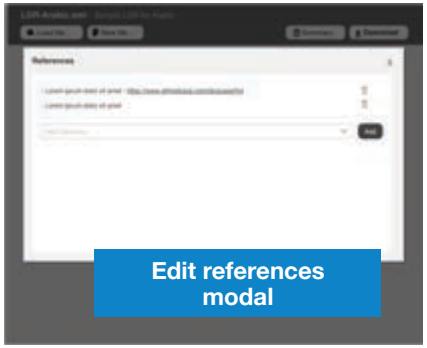
Create Use Case: Code Point Screens Flow



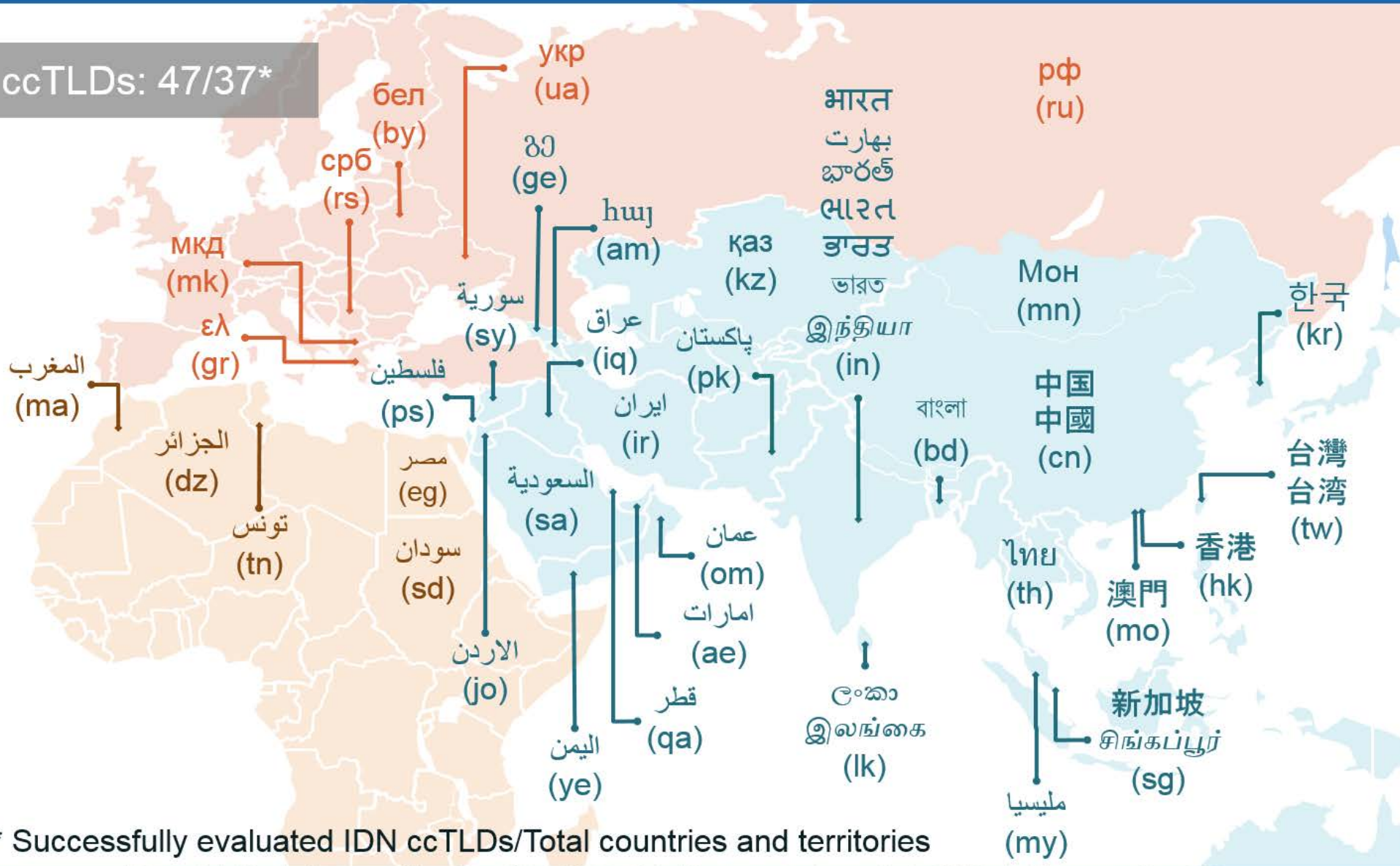








IDN Country Code Top-Level Domains



* Successfully evaluated IDN ccTLDs/Total countries and territories

IDN ccTLD Fast Track Process

- Completed five years of operation
  - Requests in 18 scripts for 27 languages
  - 37 IDN ccTLDs delegated in the Root Zone representing 29 countries/territories

- Currently under annual review
  - Public comment announced on 15 Jan. 2015
    - Second similarity review and process
  - Public comment closed on 17 March 2015
  - Next steps based on direction from the Board

IDN Tables

- **Submitted by new gTLD registries intending to offer IDNs:**
  - Varied character repertoire and contextual rules
  - Reference Label Generation Rulesets (LGRs) for facilitation and consistency in Pre-Delegation Testing (PDT) and the Registry Service Evaluation Process (RSEP)
  - Reuse for better end-user experience
- **Current Status**
  - Development through vendor
    - Guidelines and documentation of authoritative sources
    - IDN table in LGR format
    - Expert review: linguistic and security and stability
    - Public comments
    - Publication after incorporating public feedback
  - Public comment – update process to allow review; coordinate with GPs; wait till LGR for Root Zone; why LGR format
  - RFP released on 5 June 2015 – responses by 7 July 2015

IDN Implementation Guidelines

- **Background and motivation**
  - To promote IDN registration policies and practices and to minimize consumer risk and confusion
  - Last updated in 2011; GNSO community requested for review

- **Next steps**
  - Formulate a WG to undertake the review

Communication and Outreach Efforts

- IDN Program sessions at ICANN meetings
- IDN Program updates to SOs/ACs at ICANN meetings
- Presentations
  - APRALO-APAC Hub Webinar on IDNs on 9 April 2015
  - Support IDN related outreach (APRICOT, MENOG, Russian IGF, DNS Forum, Africa Internet Summit)
  - Direct outreach (Cambodia, Laos, Myanmar, Pakistan)
- IDN pages at ICANN Community Wiki and ICANN website
- IDN mailing lists
  - {vip, lgr, ArabicGP, ArmenianGP, ChineseGP, …}@icann.org

Get Involved

Volunteer

Volunteer for your script Generation Panel (GP)

To contribute expertise, contribute to the GP for your script. You can get involved by simply emailing your CV and a brief statement of interest to idntlds@icann.org

Review

Review work through public comments

Sign up for the IDN mailing list vip@icann.org (to sign up, visit https://mm.icann.org/listinfo/vip) and participated in the review of IDN work being done at ICANN through the public comments

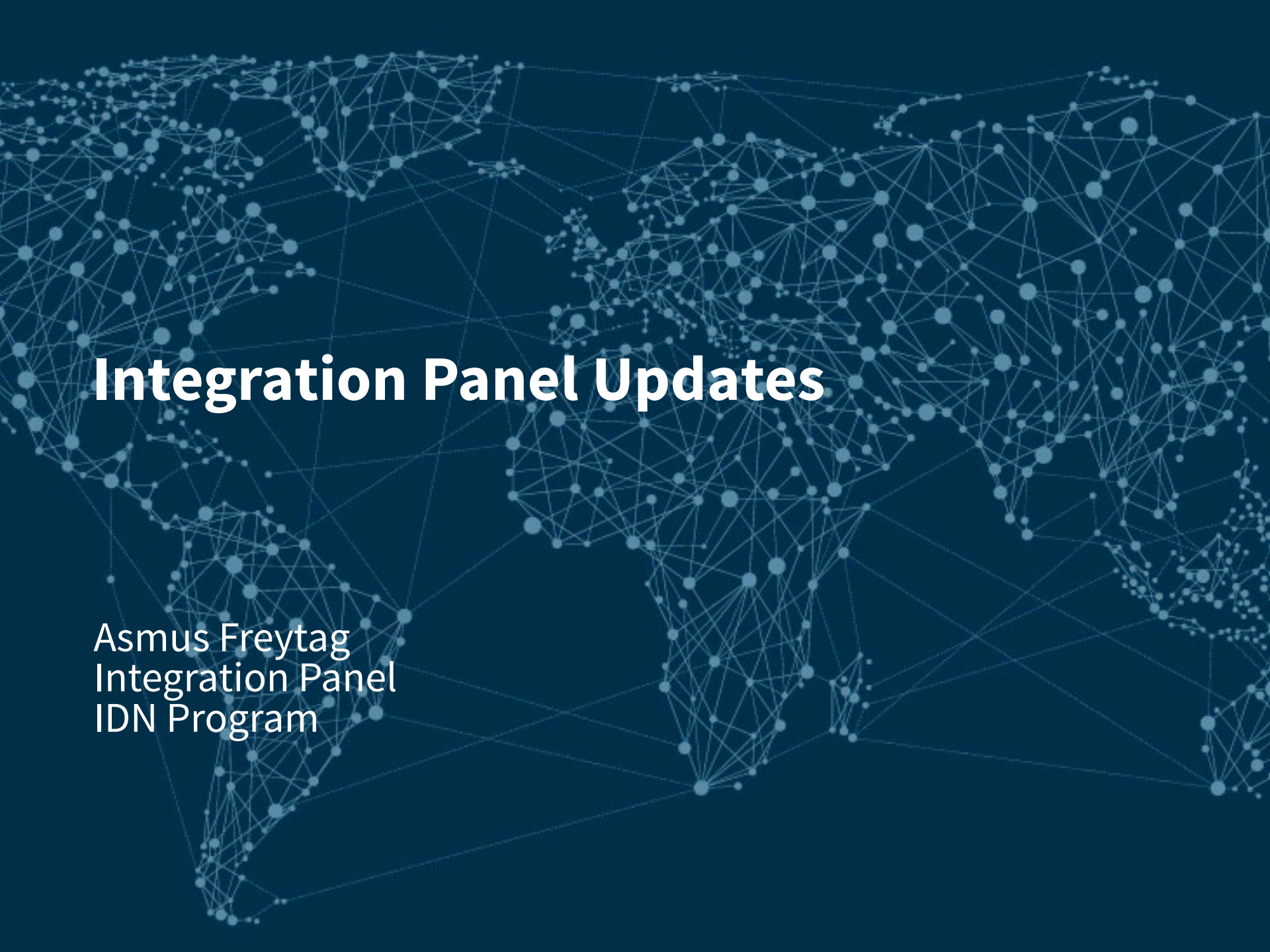
Listen

Stay up to date

Attend regular IDN Program Update sessions at ICANN meetings and sign up on the IDN mailing list vip@icann.org to get updates on the IDN Program at ICANN

Useful Links for IDN Program

- To join a Generation Panel for your language, submit CV and statement of interest at: idntlds@icann.org; call for Generation Panels: http://www.icann.org/en/news/announcements/announcement-11jul13-en.htm
- LGR Document Repository: https://community.icann.org/display/croscomlgrprocedure/Document+Repository
- Community Wiki for LGR Project: https://community.icann.org/display/croscomlgrprocedure/Root+Zone+LGR+Project
- IDN ccTLD Fast Track Page: https://www.icann.org/resources/pages/string-evaluation-completion-2014-02-19-en
- IDN Implementation Guidelines: https://www.icann.org/resources/pages/implementation-guidelines-2012-02-25-en



Integration Panel Updates

Asmus Freytag
Integration Panel
IDN Program

Integration Panel Activities Since ICANN 52

1. Maximal Starting Repertoire – version 2
2. Guidelines for Generation Panels
3. Support to Generation Panels
4. Planning Communications and Outreach
5. Preparation for Integration of LGR Proposals

Maximal Starting Repertoire – v.2 [MSR-2]

- Includes six new scripts:
  - Armenian, Ethiopic, Khmer, Myanmar, Thaana, Tibetan
  - Existing 22 scripts in MSR-1 unchanged
- Now the basis for current round of LGR proposals
- Other scripts explored, not yet ready for inclusion
  - Cherokee
    - Undergoing revision – addition of lower-case register
  - Canadian Aboriginal Syllabic (CAS)
    - Doubts on level of use, community-preparedness
  - Others still (e.g. Ol Chiki ...)

MSR-2 Statistics

MSR-2 Scripts

Script	PVALID	MSR-2
Armenian	39	38
Ethiopic	456	364
Khmer	95	78
Myanmar	177	90
Thaana	50	50
Tibetan	119	80
TOTAL	**936**	**700**

MSR-1 vs. MSR-2

	MSR-1	MSR-2
Scripts	22	28
Code Points	32,790	33,490

IDNA2008

	PVALID
Unicode 6.3	97,946

MSR-2 – Keeping Pace with Unicode

- IETF has identified a problem in Unicode 7.0, and therefore stopped updating IDNA2008 at 6.3
  - Unicode 8.0 just released
  - Unicode 9.0 to be delivered next year
- Result: When IETF resumes update of IDNA2008, MSR may need to include newly coded characters

Guidelines for Generation Panels

- Overview Documents
  - Guidelines for Developing Script-Specific Label Generation Rules for Integration into the Root Zone LGR
  - Considerations for Designing a Label Generation Ruleset for the Root Zone
  - Requirements for LGR Proposals
  - Summary Guidelines
- Background Technical Documents
  - Variant Rules
  - Whole Label Evaluation (WLE) Rules
  - Representing Label Generation Rulesets using XML
  - Packaging the MSR and LGR
- Foundation Documents
  - Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels
  - MSR-2

Resources

- Root Zone LGR Project Wiki
  - https://community.icann.org/display/croscomlgrprocedure/Root+Zone+LGR+Project
- Root Zone LGR Project Document Repository
  - https://community.icann.org/display/croscomlgrprocedure/Document+Repository

Generation Panels – State of Play

1

Draft LGRs received

Arabic, Armenian

2

Detailed discussions

Chinese, Japanese, Korean

Best strategies to co-ordinate use of Han (Chinese) characters across CJK, esp. variants

(e.g., Seoul meeting 15-16 May)

3

Proposals to form GPs received and under review

Neo-Brahmi (Indian), Cyrillic, Khmer

Note: Progress in understanding co-ordination to aid future GPs.

Panels have requested ICANN to appoint technical advisor for co-ordination.

Planning, Communications and Outreach

- Prepare work-plan for FY 2015-2016:
  - Review first LGRs (label generation rule-sets) received from script GPs (generation panels)
  - Produce first combined sets of LGRs
  - Evaluate need for MSR-3 (with/without extra scripts)
  - Support new GPs as they form, and begin work on LGRs

- Note critical dependence on:
  - Speed of formation of GPs
  - Number of GPs concurrently active
  - Time-scale to LGR submission

Preparation for Integration of LGR Proposals

- Review plans to package Integrated Root Zone LGR
- Process:
  - Get familiar with linguistic and technical issues of each script
  - Develop/refine tools to validate and correlate LGRs as received
  - Integrate effects of rules across scripts, e.g.
    - Contrasting variants, as among different languages using Han
    - Confusability among characters of different scripts
- Ultimate goals:
  - A set of mechanical verification steps
  - Applicable to preliminary and final LGRs

From the Integration Panel

Thank you for your attention!

Wil Tan, Marc Blanchet, Asmus Freytag, Michel Suignard, Nicholas Ostler

Update on Arabic GP

Meikal Mumin
Task Force on Arabic Script IDNs (TF-AIDN) /
Arabic Generation Panel

Membership

- TF-AIDN has open membership
  - 33 members
  - 18 countries

- Current members speak 9 languages using Arabic script, with further expertise in use of Arabic script from East Asia, South Asia, Middle East, Maghreb countries and Africa

- Current members come from diverse disciplines – academia (linguistics and technical), registries, registrars, national and regional policy bodies, community based organizations, technical community

Arabic Script TLDs Assigned or Delegated

تونس	بھارت	الجزائر
مصر	المغرب	عمان
قطر	السعودية	ايران
فلسطين	سودان	امارات
عراق	مليسيا	بازار
موقع	شبكة	پاكستان
	سورية	الاردن

Progress - Work accomplished - 1

- Formation of Arabic Script Generation Panel
- Code point analysis
  - Establishing principles for inclusion, exclusion, and deferral of Unicode code points
  - Analysis of Maximum Starting Repertoire (MSR) and feedback to Integration Panel (IP)
  - Analysis of code points for Label Generation Rules (LGR)
- Code point variants
  - Establishing principles for code point variants
  - Analysis of code points for code points variants
  - Analysis of code point variant dispositions (allocatable vs. blocked)

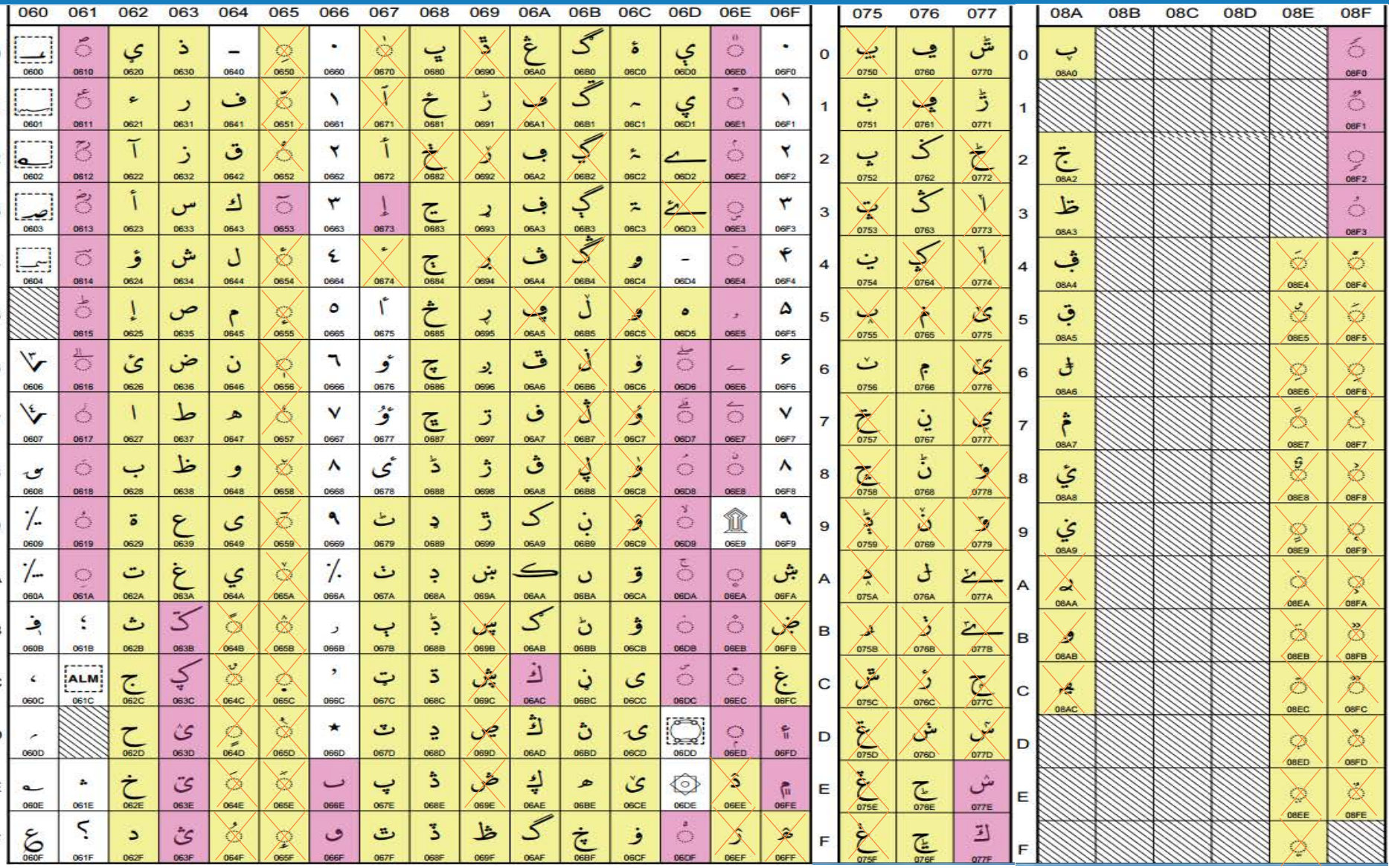
Progress - Work accomplished - 2

- Development of Whole Label Evaluation (WLE) rules
  - Establishing principles for WLE rules
  - Determining the WLE rules
- Documentation of the work
  - Text document summarizing Label Generation Rules for Arabic
  - Label Generation Rules for Arabic in XML format
- Finalization of LGR proposal at F2F meeting in Singapore (ICANN52)
- Discussion with Integration Panel (IP) for feedback

- Currently reviewing IP feedback and finalizing proposal

- Next steps: Publish the final proposal for public comments

Current Code Point Repertoire



Examples of Variants

Partial table for HEH

06C1	06D5	b/b	
06C1	0647	a/a	Labels in Arabic language using 0647 in the final and isolated positions will be typed in other languages using 06C1 (Urdu, Pashto, Saraiki, etc.) which is identical in isolated and has a variant glyph or identical glyph form in final position.
06C1	06BE	b/b	
06C1	0629	b/b	
06C1	06C3	b/b	
06C1	06C0	b/b	
06C1	06C2	b/a	06C1 ->b 06C2 (blocked), 06C2 ->a 06C1 (allocatable)

Examples of WLE Rules

S.No.	Code points cannot co-occur within a label	Notes
1	0643 and 06A9	ك and ک
2	0643 and 06AA	ك and ک
3	06CC and 0649	ی and ی
4	06C1 and 0647	ہ and ہ
5	06C1 and 06D5	ہ and ہ
6	0647 and 06D5	ہ and ہ
7	0647 and 06BE	ہ and ھ
8	0629 and 06C3	ۃ and ۃ

Sample Output

Generating all variants for "پاکستان" (067E 0627 06A9 0633 062A 0627 0646) (IDN ccTLD for Pakistan)

Variant: پاکستاں (067E 0627 0643 0633 062A 0627 06BA)
Variant: پاکستان (067E 0627 0643 0633 062A 0627 0646)
Variant: پاکستاں (067E 0627 06AA 0633 062A 0627 06BA)
Variant: پاکستان (067E 0627 06AA 0633 062A 0627 0646)
Variant: پاکستاں (067E 0627 06A9 0633 062A 0627 06BA)

1200 variants generated for "پاکستان" (067E 0627 06A9 0633 062A 0627 0646)

By disposition: { allocatable=5, blocked=1194, original=1 }

Thank You

- Workspace, news and document archive at https://community.icann.org/display/MES/TF-AIDN+Work+Space

- Publicly archived email discussion at http://lists.meswg.org/mailman/listinfo/tf-aidn

Update on Armenian GP

Igor Mkrtumyan
Armenian Generation Panel

Igor

General information

Armenian IDN

- Code: Armn
- N°: 230
- English Name: Armenian
- English name of the script: Hye
- Native name of the script: հայ

The announcement for the successful completion of Armenia's string evaluation has been posted at https://www.icann.org/news/announcement-2014-11-20-en.

Armenian language

- The **Armenian language** is an Indo-European language spoken by the Armenians. It is the official language of the Republic of Armenia and the self-proclaimed Nagorno-Karabakh Republic. It has historically been spoken throughout the Armenian Highlands and today is widely spoken in the Armenian Diaspora.

Armenian language

- Armenians has its own unique script, the Armenian alphabet, invented in 405 AD by Mesrop Mashtots.
- Linguists classify Armenian as an independent branch of the Indo-European language family.
- There are two standardized modern literary forms, Eastern Armenian and Western Armenian, with which most contemporary dialects are mutually intelligible.
- Total Armenian population in the world is about 10mln.

Geographic territories or countries with significant user communities for the script

Official language in Armenia and Nagorno Karabakh Republic

Big Diaspora using the Armenian language in

Argentina	Lebanon
Brazil	Poland
Cyprus	Romania
France	Syria
Georgia	Turkey
Hungary	Ukraine
Iran	United States
Iraq	Uruguay

Armenian GP

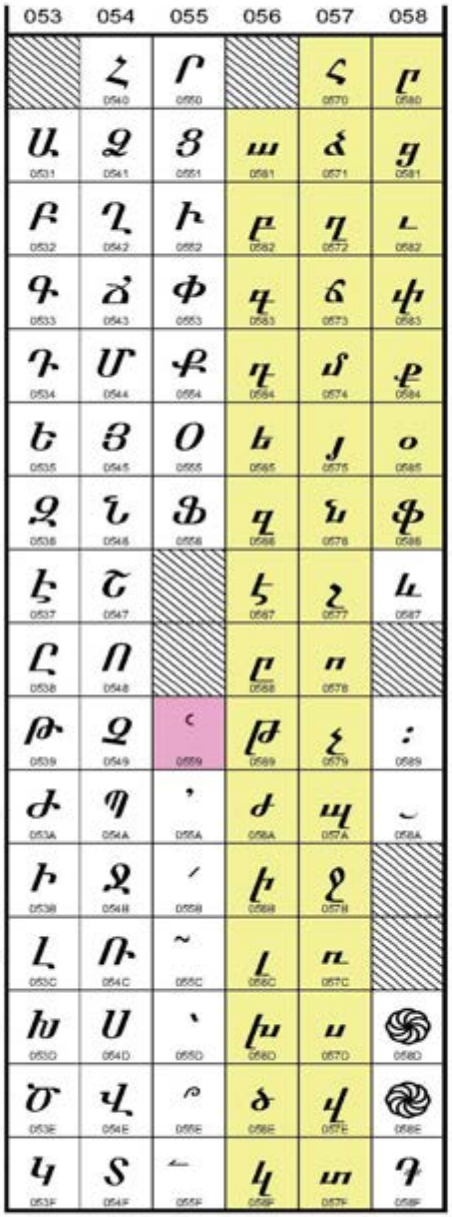
- Armenian GP was created
- Armenian GP mailing list was created Armeniangp@icann.org
- General information about the mailing list is at: https://mm.icann.org/mailman/listinfo/armeniangp

Armenian MSR-2 table

- There are two standardized mutually intelligible modern literary forms, Eastern Armenian and Western Armenian with some difference in orthographies. But as both are written with the same script it doesn't influence the LGR document.
  The following web site contains a full description of the Armenian repertoire for use in the Armenian writing system:
  http://www.omniglot.com/writing/armenian.htm

MSR-2 Armenian Repertoire

0530 Armenian 058F



MSR-2 Armenian Repertoire

Uppercase letters

0531 Ա ARMENIAN CAPITAL LETTER AYB
0532 Բ ARMENIAN CAPITAL LETTER BEN
0533 Գ ARMENIAN CAPITAL LETTER GIM
0534 Դ ARMENIAN CAPITAL LETTER DA
0535 Ե ARMENIAN CAPITAL LETTER ECH
0536 Զ ARMENIAN CAPITAL LETTER ZA
0537 Է ARMENIAN CAPITAL LETTER EH
0538 Ը ARMENIAN CAPITAL LETTER ET
0539 Թ ARMENIAN CAPITAL LETTER TO
053A Ժ ARMENIAN CAPITAL LETTER ZHE
053B Ի ARMENIAN CAPITAL LETTER INI
053C Լ ARMENIAN CAPITAL LETTER LIWN
053D Խ ARMENIAN CAPITAL LETTER XEH
053E Ծ ARMENIAN CAPITAL LETTER CA
053F Կ ARMENIAN CAPITAL LETTER KEN
0540 Հ ARMENIAN CAPITAL LETTER HO
0541 Ձ ARMENIAN CAPITAL LETTER JA
0542 Ղ ARMENIAN CAPITAL LETTER GHAD
0543 Ճ ARMENIAN CAPITAL LETTER CHEH
0544 Մ ARMENIAN CAPITAL LETTER MEN
0545 Յ ARMENIAN CAPITAL LETTER YI
0546 Ն ARMENIAN CAPITAL LETTER NOW
0547 Շ ARMENIAN CAPITAL LETTER SHA
0548 Ո ARMENIAN CAPITAL LETTER VO
0549 Չ ARMENIAN CAPITAL LETTER CHA
054A Պ ARMENIAN CAPITAL LETTER PEH
054B Ջ ARMENIAN CAPITAL LETTER JHEH
054C Ռ ARMENIAN CAPITAL LETTER RA
054D Ս ARMENIAN CAPITAL LETTER SEH
054E Վ ARMENIAN CAPITAL LETTER VEW
054F Տ ARMENIAN CAPITAL LETTER TIWN
0550 Ր ARMENIAN CAPITAL LETTER REH
0551 Ց ARMENIAN CAPITAL LETTER CO
0552 Ւ ARMENIAN CAPITAL LETTER YIWN
0553 Փ ARMENIAN CAPITAL LETTER PIWR
0554 Ք ARMENIAN CAPITAL LETTER KEH
0555 Օ ARMENIAN CAPITAL LETTER OH
0556 Ֆ ARMENIAN CAPITAL LETTER FEH

Modifier letters

0559 ՙ ARMENIAN MODIFIER LETTER LEFT HALF RING
• technical use
055A ՚ ARMENIAN APOSTROPHE
055B ՛ ARMENIAN EMPHASIS MARK
055C ՜ ARMENIAN EXCLAMATION MARK
055D ՝ ARMENIAN COMMA
055E ՞ ARMENIAN QUESTION MARK
055F ՟ ARMENIAN ABBREVIATION MARK

Lowercase letters

0561 ա ARMENIAN SMALL LETTER AYB
0562 բ ARMENIAN SMALL LETTER BEN
0563 գ ARMENIAN SMALL LETTER GIM
0564 դ ARMENIAN SMALL LETTER DA
0565 ե ARMENIAN SMALL LETTER ECH
0566 զ ARMENIAN SMALL LETTER ZA
0567 է ARMENIAN SMALL LETTER EH
0568 ը ARMENIAN SMALL LETTER ET
0569 թ ARMENIAN SMALL LETTER TO
056A ժ ARMENIAN SMALL LETTER ZHE
056B ի ARMENIAN SMALL LETTER INI
056C լ ARMENIAN SMALL LETTER LIWN
056D խ ARMENIAN SMALL LETTER XEH
056E ծ ARMENIAN SMALL LETTER CA
056F կ ARMENIAN SMALL LETTER KEN
0570 հ ARMENIAN SMALL LETTER HO
0571 ձ ARMENIAN SMALL LETTER JA
0572 ղ ARMENIAN SMALL LETTER GHAD
0573 ճ ARMENIAN SMALL LETTER CHEH
0574 մ ARMENIAN SMALL LETTER MEN
0575 յ ARMENIAN SMALL LETTER YI
0576 ն ARMENIAN SMALL LETTER NOW
0577 շ ARMENIAN SMALL LETTER SHA
0578 ո ARMENIAN SMALL LETTER VO
0579 չ ARMENIAN SMALL LETTER CHA
057A պ ARMENIAN SMALL LETTER PEH
057B ջ ARMENIAN SMALL LETTER JHEH
057C ռ ARMENIAN SMALL LETTER RA
057D ս ARMENIAN SMALL LETTER SEH
057E վ ARMENIAN SMALL LETTER VEW
057F տ ARMENIAN SMALL LETTER TIWN
0580 ր ARMENIAN SMALL LETTER REH
0581 ց ARMENIAN SMALL LETTER CO
0582 ւ ARMENIAN SMALL LETTER YIWN
0583 փ ARMENIAN SMALL LETTER PIWR
0584 ք ARMENIAN SMALL LETTER KEH
0585 օ ARMENIAN SMALL LETTER OH
0586 ֆ ARMENIAN SMALL LETTER FEH
0587 և ARMENIAN SMALL LIGATURE ECH YIWN

Punctuation

0589 ։ ARMENIAN FULL STOP
058A ֊ ARMENIAN HYPHEN

Religious symbols

058D ֍ RIGHT-FACING ARMENIAN ETERNITY SIGN
058E ֎ LEFT-FACING ARMENIAN ETERNITY SIGN

Currency symbol

058F ֏ ARMENIAN DRAM SIGN

Armenian MSR-2 table

- This defines a formal reference for the justification of the repertoire for this proposal.
- The following subset of code points identified as the Unicode range U+0561..U+0586 will be available for the root zone:

Selected code points

Code	Script	Name	Code	Script	Name
0561	ա	ARMENIAN SMALL LETTER AYB	0574	մ	ARMENIAN SMALL LETTER MEN
0562	բ	ARMENIAN SMALL LETTER BEN	0575	յ	RMENIAN SMALL LETTER YI
0563	գ	ARMENIAN SMALL LETTER GIM	0576	ն	ARMENIAN SMALL LETTER NOW
0564	դ	ARMENIAN SMALL LETTER DA	0577	շ	ARMENIAN SMALL LETTER SHA
0565	ե	ARMENIAN SMALL LETTER ECH	0578	ո	ARMENIAN SMALL LETTER VO
0566	զ	ARMENIAN SMALL LETTER ZA	0579	չ	ARMENIAN SMALL LETTER CHA
0567	է	ARMENIAN SMALL LETTER EH	057A	պ	ARMENIAN SMALL LETTER PEH
0568	ը	ARMENIAN SMALL LETTER ET	057B	ջ	ARMENIAN SMALL LETTER JHEH
0569	թ	ARMENIAN SMALL LETTER TO	057C	ռ	ARMENIAN SMALL LETTER RA
056A	ժ	ARMENIAN SMALL LETTER ZHE	057D	ս	ARMENIAN SMALL LETTER SEH
056B	ի	ARMENIAN SMALL LETTER INI	057E	վ	ARMENIAN SMALL LETTER VEW
056C	լ	ARMENIAN SMALL LETTER LIWN	057F	տ	ARMENIAN SMALL LETTER TIWN
056D	խ	ARMENIAN SMALL LETTER XEH	0580	ր	ARMENIAN SMALL LETTER REH
056E	ծ	ARMENIAN SMALL LETTER CA	0581	ց	ARMENIAN SMALL LETTER CO
056F	կ	ARMENIAN SMALL LETTER KEN	0582	ւ	ARMENIAN SMALL LETTER YIWN
0570	հ	ARMENIAN SMALL LETTER HO	0583	փ	ARMENIAN SMALL LETTER PIWR
0571	ձ	ARMENIAN SMALL LETTER JA	0584	ք	ARMENIAN SMALL LETTER KEH
0572	ղ	ARMENIAN SMALL LETTER GHAD	0585	օ	ARMENIAN SMALL LETTER OH
0573	ճ	ARMENIAN SMALL LETTER CHEH	0586	ֆ	ARMENIAN SMALL LETTER FEH

Variants

There is some similarity between Armenian letters GIM and ZA but we considered that it is not enough to include them into variants. So our conclusion is:

The Armenian repertoire does not contain in-script variants.

Cross-script homoglyphs

- Cyrillic, Greek and Latin scripts are examples of related scripts where some cross-script homoglyphs exist.

Armenian-Cyrillic homoglyphs

Armenian Letter	Cyrillic Letter
հ U+0570 Armenian small letter HO	һ U+04BB Cyrillic small letter SHHA
օ U+0585 Armenian small letter OH	о U+043E Cyrillic small letter O

‘ո’ U+0578 Armenian small letter VO and ‘п’ U+43F Cyrillic small letter PE are usually somewhat distinct, unless both are in italic serif style, like ‘*п*’ vs. ‘*n*’, where they would be homoglyphs. However, they are not included into cross-script variants because italic font styles are insufficient justification for creating a cross-script variant, especially as that is not the most common style for presenting domain names to the user.

Armenian and Latin homoglyphs

Armenian Letter	Latin Letter
ԛ U+0566 Armenian small letter ZA	q U+0071 Latin small letter Q
հ U+0570 Armenian small letter HO	h U+0068 Latin small letter H
ո U+0578 Armenian small letter VO	n U+006E Latin small letter N
ս U+057D Armenian small letter SHE	u U+0075 Latin small letter U
ց U+0581 Armenian small letter CO	g U+0067 Latin small letter G
ւ U+0582 Armenian small letter YIWN	ɩ U+0269 Latin small letter IOTA
օ U+0585 Armenian small letter OH	o U+006F Latin small letter O

Armenian and Latin homoglyphs

The ARMENIAN SMALL LETTER ZA and the LATIN SMALL LETTER Q are not perfect homoglyphs but the difference is may not be perceivable at normal size. A label, such as .quo would be readily accepted by users as “the same” as the label .quo (in Latin).
The ‘g’ homoglyph situation only exists in sans-serif style, which, however, is a very common choice for user interfaces.
Armenian YIWN and Latin IOTA are considered homoglyphs as their visual perception is quite similar.
The ARMENIAN LETTER YI and j U+006A LATIN SMALL LETTER J are not considered homoglyphs as “dot” changes the visual perception of the letter. That is why this case is not included into the cross-script variants.

Armenian and Greek homoglyphs

Armenian Letter	Greek Letter
հ U+0572 Armenian small letter GHAD	η U+03B7 Greek small letter ETA
ւ U+0582 Armenian small letter YIWN	ι U+03B9 Greek small letter IOTA
օ U+0585 Armenian small letter OH	ο U+03BF Greek small letter OMICRON

Armenian YIWN and Greek IOTA as well as Armenian GHAD and Greek ETA are considered homoglyphs as their visual perception is quite similar. They are included in the cross-script variants.
The case of OH/OMICRON/Cyrillic O/Latin O is considered cross-script homoglyphs because of strong similarity.

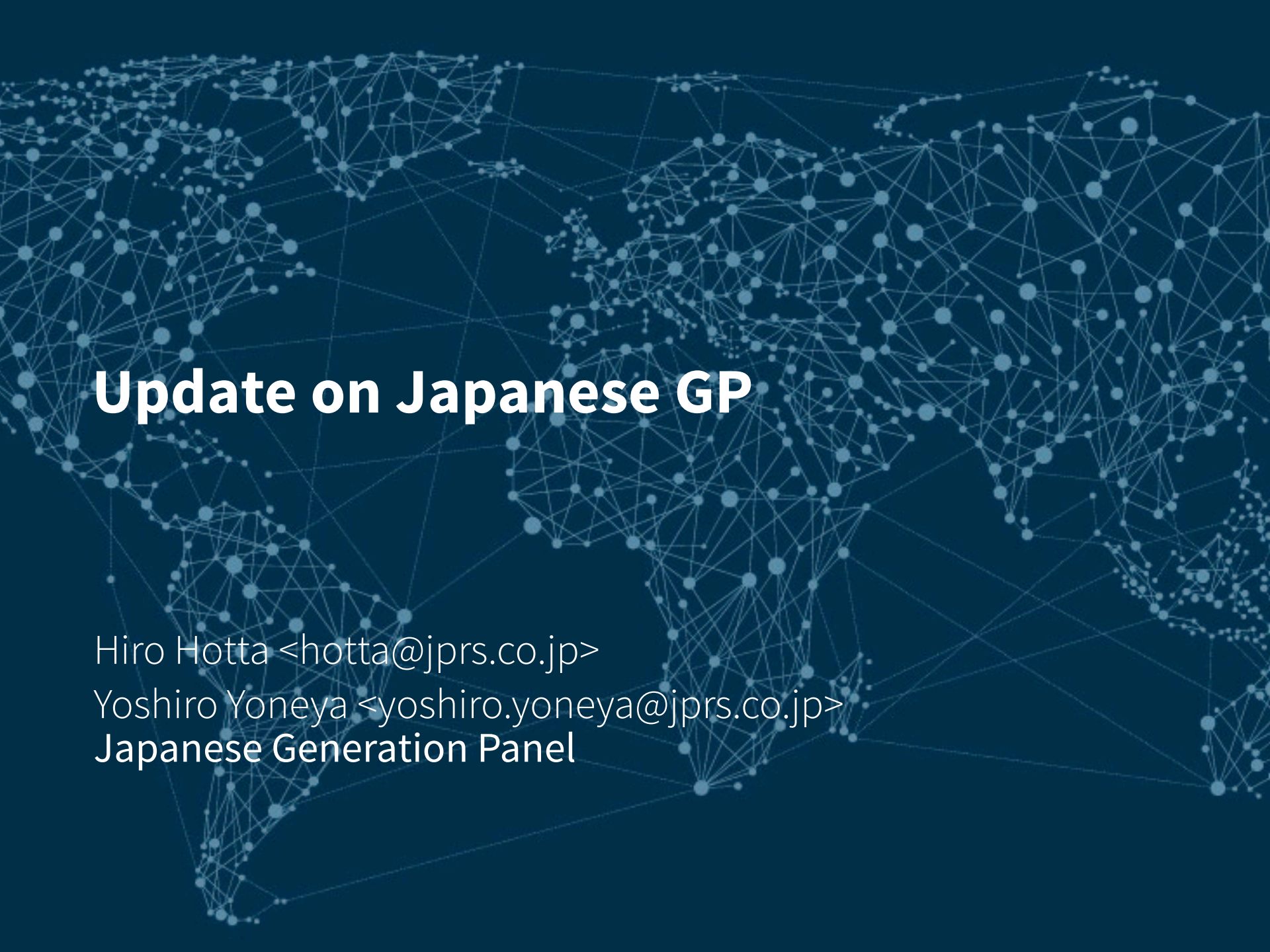
Conclusions

As result of integration, blocked variants can exist between script-specific repertoires.
Even if a label is applied for in one script, the existence of certain other labels in another script may block it from being delegated .
For example, both .uno and .ино would be well-formed TLDs in Latin and Armenian,
If one of these two labels is applied for and allocated, the other would be blocked. Otherwise, both could be independently allocated to different applicants causing user confusion.
Armenian GP is hardly expecting any other application for an Armenian TLD but in cases when there is a homoglyph similarity in related scripts it should be blocked.
The Armenian GP has added out of repertoire blocked variants for all homoglyphs investigated above, except for cases identified as merely "similar".

Second level domains

Visual similarities in the second level domains will be solved by registration rules. There will be restrictions in the policy of domain registration. The problem will be solved by limiting Armenian domain names strictly to the Armenian MSR table, Latin dash (code point '2d') and Latin numbers (code points '30' - '39')".
Visual similarities of related scripts will be blocked by the domain registration program as it will check the scripts for the correspondence to the Armenian MSR table and will not allow domains names with the visually similar code points of related scripts.

Thank you!



Update on Japanese GP

Hiro Hotta <hotta@jprs.co.jp>
Yoshiro Yoneya <yoshiro.yoneya@jprs.co.jp>
Japanese Generation Panel

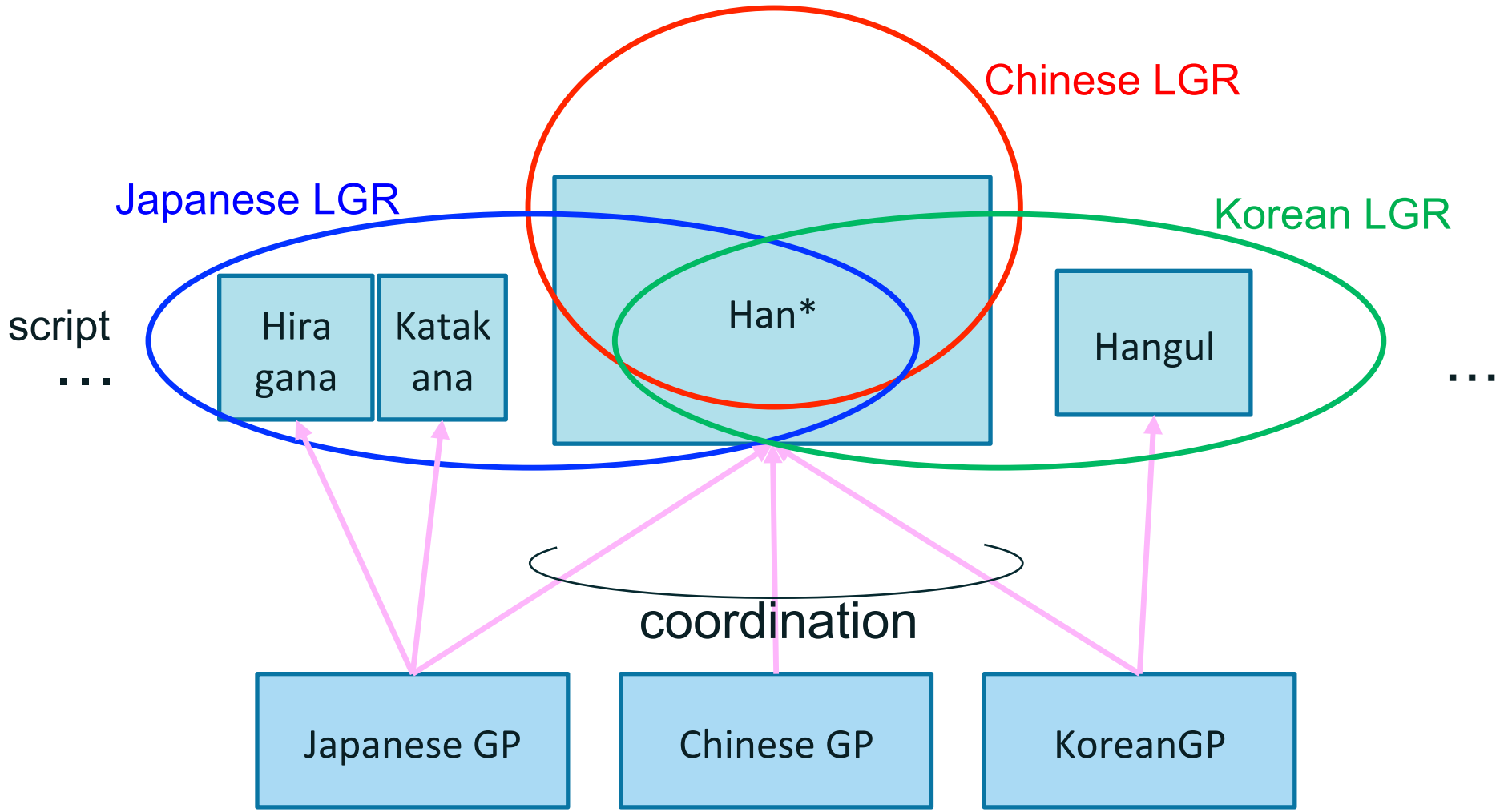
JGP meetings & related events

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

JGP members

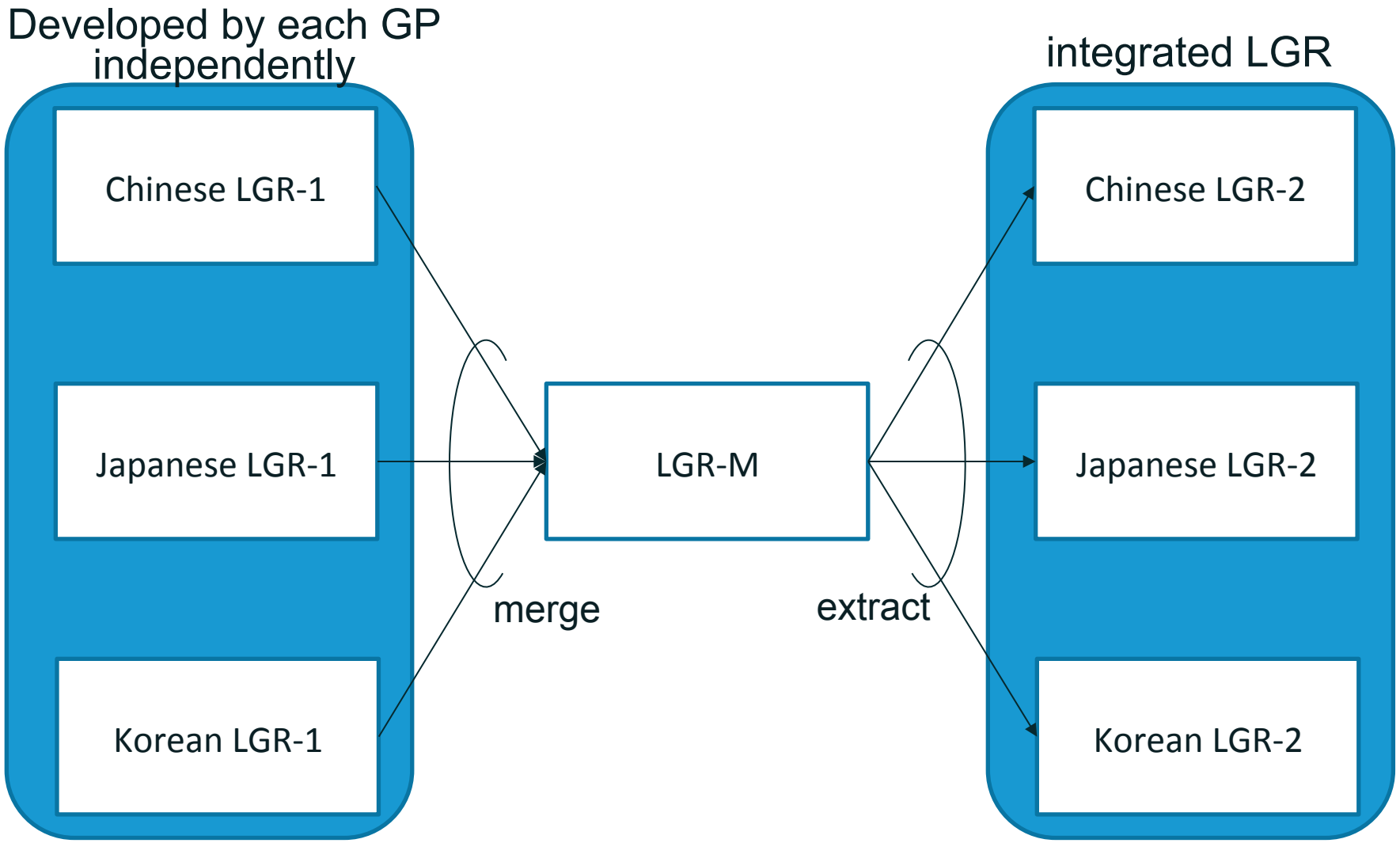
- Members and their expertise
  - Hiro Hotta chair
    - Policy/business aspects of registry/registrar
  - Akinori Maemura vice chair
    - Internet governance and domain name in general
  - Shigeki Goto
    - Internet in general
  - Kazunori Konishi
    - Internet in general
  - Tsugizo Kubo
    - Trademarks and domain names
  - Yoshitaka Murakami (from February 4, 2015)
    - Trademarks and gTLD markets as registry consultant
  - Shuichi Tashiro
    - Character codes
  - Yoshiro Yoneya
    - Technical aspects of IDN, LGR

Relationship among CJK language LGRs



* “Han” is called “Kanji” in Japan, “Hanja” in Korea

Framework of CJK LGR integration for Han characters



Activities

- JGP establishment
  - Proposal submitted to ICANN (February 6, 2015)
  - Establishment approved by ICANN (March 10, 2015)
- Detailed task description of JGP
  - Done
  - Some more tasks or issues may come out or tasks may be modified on the way of LGR development
    - through discussion with ICANN/IP
    - through discussion with CGP and KGP (as well as IP)
    - through investigation inside JGP
- Development of Japanese LGR
  - CJK LGR integration algorithm was developed and agreed by CGP, JGP, and KGP as a framework
  - As the input to the algorithm, preliminary Japanese LGR (which is called LGR-1) was developed

Discussion status for Japanese LGR-1

More detailed explanation is from the next slide

- Scopes of the character codes
  - Kanji, Hiragana, Katakana
  - For Kanji
    - JIS (Japanese Industrial Standard) level-1 and level-2
- Variants & each variant type
  - For Kanji
    - Japanese LGR-1 will define no variants for itself
    - Integrated Japanese LGR (whick is called LGR-2) will import (= passively adopt) variants of Chinese LGR-1 and Korean LGR-1
    - Types of each variant in Japanese LGR will be defined in a systematic way == > need more investigation if reduction of allocatable labels is needed, and how it can be done if needed
- WLE (whole label evaluation)
  - Japanese LGR-1 may have no or very limited number of tiny rules even if defined == > need more investigation

Overview of Japanese LGR-1 (J-LGR-1)

- 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 are under discussion (nor for LGR-1 but for LGR-2 after CJK integration)
- WLE
  - Rules (although not very many) are under discussion

Background of the J-LGR-1 (1/3)

- Selection of repertoire
  - Based on Japanese Industrial Standards (JIS) X 0208:2012 Level-1 and Level-2
  - JIS X 0208 is most widely used character set in Japan (all characters have mapping to Unicode)
    - And also has track record in IDN.jp service
  - JGP couldn't find any rational reason to add / remove characters from JIS X 0208

Background of the J-LGR-1 (2/3)

- Definition of variants
  - JGP compared CGP's draft LGR-1 and JGP's draft LGR-1
    - Tentatively JGP concluded that serious influence of CGP's variants to Japanese IDN TLD might be very limited
  - JGP assessed occurrence of "may be seriously affected" codepoints in currently-registered Japanese .JP domain names
    - Tentatively JGP concluded that such codepoints are mostly used as independent unique characters in Japanese IDNs
  - From the above assessment, tentatively JGP concluded that no variants needed for Japanese LGR-1

Background of the J-LGR-1 (3/3)

- Definition of WLE
  - JGP considered some trivial cases
    - Hira/Kana Prolonged mark as the first character of a label
    - Iteration mark as the first character of a label
  - JGP is now discussing if such cases should be defined in LGR-1 or had better be assessed at the new gTLD application evaluation step by manned panels

Update on Korean GP

Kyongsok Kim
Korea Generation Panel

Agenda

Korean LGP Status Update

- Introduction: kore, issues
- A list of Hanja Characters for KLGR
- A list of Hanja Variants
- A special class of Variants in C-LGR
- Timeline of KGP activities
- Status of KGP
- Result of CJK joint meeting

Introduction: kore, issues

Korean LGP Status Update

01 Characters to be included in “kore” (Korean Label)

Both Hangeul and Hanaja will be included

RE : Hanja (Hanzi, Han) : coordination among CJK.

02 Issues to be considered in KGP RE: CJK coordination

- Hanja repertoire
- Hanja Variants
- Special class of Hanja Variants in C-LGR
- Review of variants in C-LGR
- Timeline of KGP activities

A list of Hanja Characters for KLGR

Korean LGP Status Update

Source of Hanja Character Set	# Chars
1) KS X 1001	4,888
2) KPS 9566	4,653
3) IICORE-K column marked	4,743
4) IICORE-KP column marked	4,653
5) Qualifying Examination of Korean Hanja Proficiency	5,940
Union of 1) ~5) [Hanja list v0.1 (2015. 5. 11)]	6,046
Missing in MSR-1	-29
[Hanja list v0.2 (2015. 6. 12)]	**6,017**

A list of Hanja variants

Korean LGP Status Update

01 66 pairs (66 variant groups) [v0.2 (2015. 6. 12)]

An example : one pair (廄 05ECF 廐 05ED0)

02 # of variant groups of K-LGR is ‘much less’ than # of variant groups of C-LGR

03 KR is reviewing and examining variant groups of C-LGR

A special class of Variants in C-LGR

Korean LGP Status Update

01 About 157 Simp-Trad chars :

Currently, the char is a simplified char in Chins. However, the char was used in China (has been used in Korea) as a Traditional char for a long time.

02 An example of Simp-Trad char : 机

1) In China :

机 : Currently, Simplified char, 'machine'.

机 : simplified from Traditional char 機 (machine).

2) In Korea : the two chars are distinct

机 : desk (reading "gwe")

機 : machine (reading "gi")

Timeline of KGP activities

Korean LGP Status Update

June 20 ~ 25
- KGP Status Update at ICANN #53
- K-LGR-1 v0.2, progress and plan

June 30 ~ July 3
- KGP Status Update at AprIGF
- K-LGR-1 v0.2, progress and plan

July 10
- KGP #9 meeting
- review of meeting result (ICANN53,AprIGF) and variants

August
- KGP #10 meeting
- review of C and K's LGR-1 variants list and Hanja repertoire(v0.3)

September
- Public Hearing
- KGP #11 meeting

October
- KGP #12 meeting
- KGP Status Update at ICANN #54
- K-LGR-1 v1.0 (Hanja repertoire)

Status of KGP

Korean LGP Status Update

01 Member of Korean label Generation Panel

Classification	Name
Technical Expert	Kyongsok Kim(Chair)
Linguistic Expert	Kyungran Kang
	Jungdo Choi
Policy Expert	Youngeum Lee
	Dongman Lee
Community	Yangwoo Ko
	YJ Park

Classification	Name
.kr registry (KISA)	Minjung Park
	Yunmi Choi
	Ryoung Chae
	Minjee Kim
Hanja Expert	Sanghyun Shin
	Sungduk Cho

Status of KGP

Korean LGP Status Update

02 KGP's activities history

2013	2013. 12 : organization of Korean LGP
2014	2014. 3 : Participate CJK joint meeting @ ICANN49 Singapore 2014. 6 : Participate ICANN50 London and KGP status update 2014. 7 : 1st KGP meeting 2014. 8 : 2nd KGP meeting 2014. 10 : Participate ICANN51 LA and KGP status update
2015	2015. 1 : 3rd KGP meeting and re-composition KGP member(Join Hanja expert and Registry) 2015. 2 : Participate ICANN52 Singapore and KGP status update 2015. 4 : 4th , 5th KGP meeting (reorganization of KGP) 2015. 5 : 6th , 7th KGP meeting(K-LGR-1 v0.1) and CJK Joint meeting in Seoul 2015. 6 : 8th KGP meeting(K-LGR-1 v0.2) and participate ICANN53

Result of CJK joint meeting

Korean LGP Status Update

24 participants from 6 countries

24 participants from 6 countries including CGP, JGP, KGP, and other structures took part in the meeting.

Adopted Yoneya algorithm

The CJK GPs agree with the principle framework of Yoneya's proposed algorithm for coordination of the CJK LGRs for Han characters, and agree to work together to refine the algorithm to incorporate respective language requirements.

Action items

5 action items adopted to move forward.

Thank you

Engage with ICANN

Thank You and Questions

Reach us at: idntlds@icann.org
Email: engagement@icann.org
Website: icann.org

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