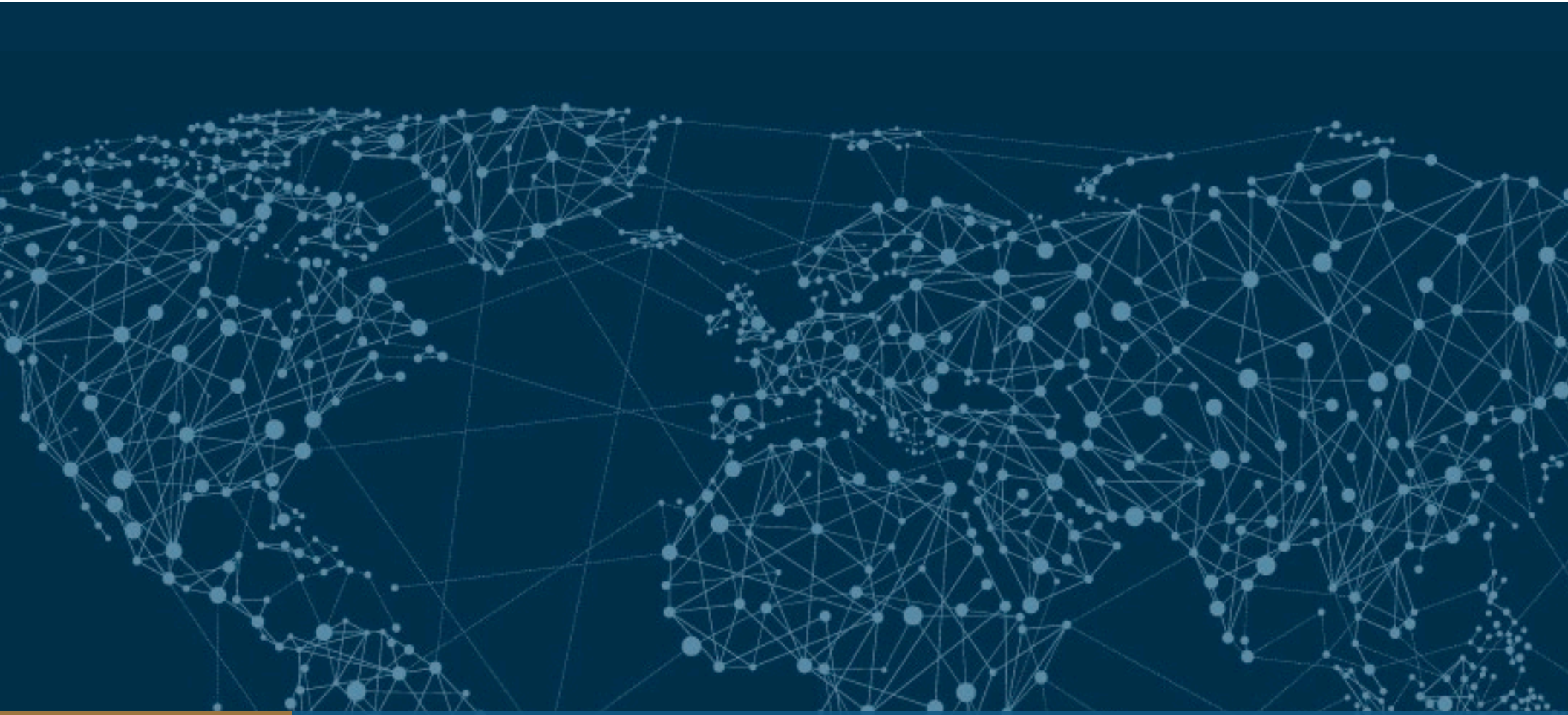

ICANN
ICANN | 52
Singapore
8-12 FEBRUARY 2015



IDN Root Zone LGR Workshop

ICANN 52 | 11 January 2015

Agenda

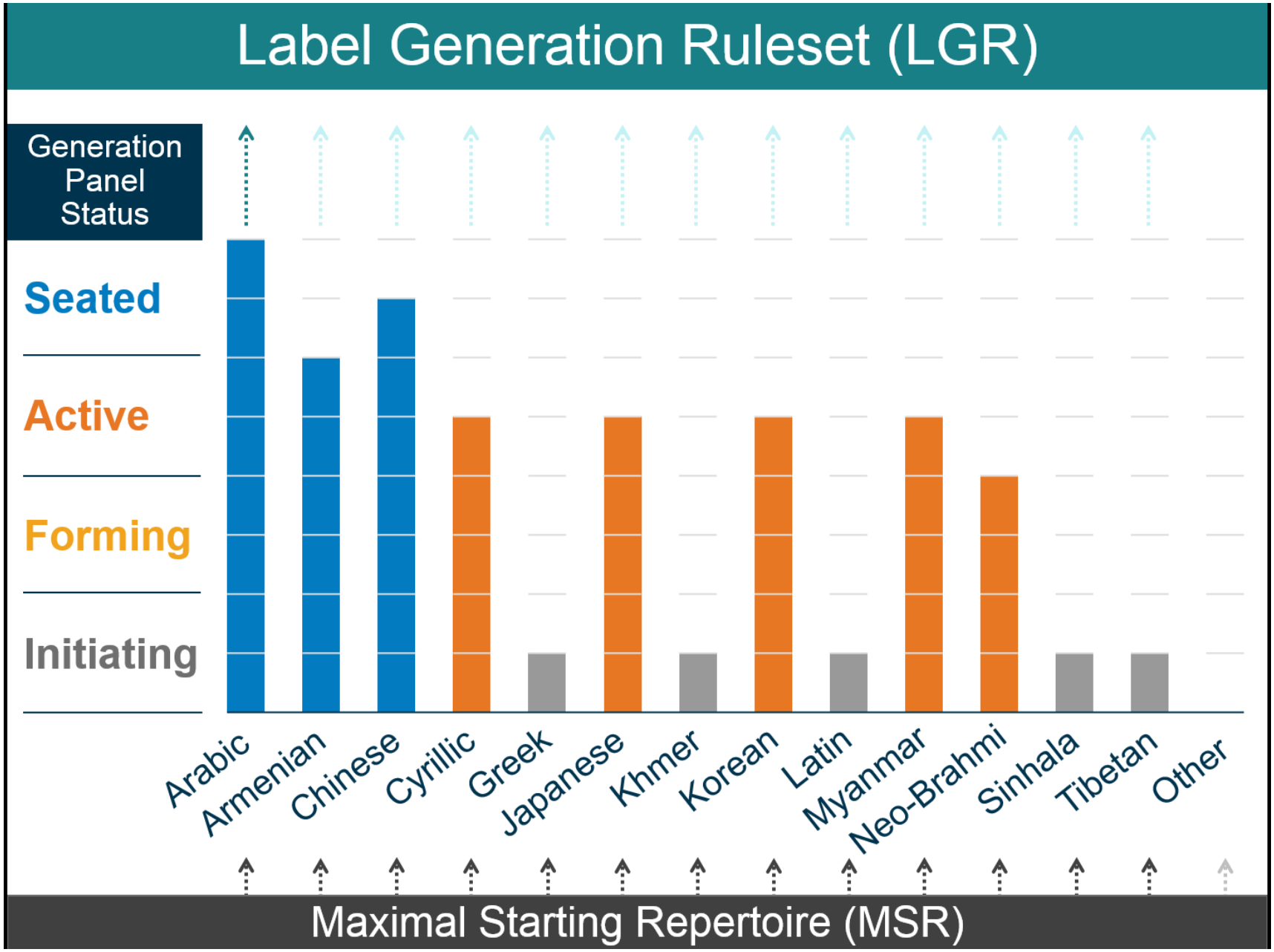
- **Introduction** – Sarmad Hussain
- **Integration Panel Discussion**
  - Guidelines for LGR Development – Wil Tan
  - How to Design Variants and WLE Rules – Michel Suignard
- **Community Updates**
  - Armenian GP Update – Igor Mkrtumyan
  - Cyrillic GP Update – Dusan Stojičević and Yuriy Kargapolov
  - Beyond the Root Zone - Applications of LGR – Philippe Collin
- **Q&A**

IDN Root Zone LGR Workshop

Introduction

Sarmad Hussain
IDN Program Senior Manager

Introduction



Integration Panel Discussion

Guidelines for LGR Development

Wil Tan
Integration Panel Member

LGR Development Process

- Guidelines for Developing Script-Specific LGRs for Integration into the Root Zone LGR document is out for public comment
- This presentation highlights some of its points
- Other documents are available to provide guidance on the Root Zone LGR Project Document Repository

Summary of Tasks

- Start with the MSR
- Select code points (define the LGR repertoire)
- Determine variants
- Determine if WLEs are needed
- Prepare LGR Proposal Submission

Start With the MSR

- At formation, GP selects an ISO-15924 script code as its scope
- This implicitly restricts the possible code points to:
  - MSR-2 code points tagged with the script code
  - (If applicable) MSR-2 code points tagged “Zinh”
- GPs may research a wider set of code points, for example:
  - To identify interactions with related scripts
  - In order to review and comment on MSR-2
- MSR-2 is out for public comment
  - Six new scripts: Armenian, Ethiopic, Khmer, Myanmar, Thaana, Tibetan
  - Existing scripts in MSR-1 unchanged

Selecting Code Points

- Start with the set of code points defined in scope for GP
  - MSR-2 is tagged with scripts

Script	XML
Armenian	<range first-cp="0561" last-cp="0586" tag="**sc:Armn**" … />
Greek	<range first-cp="03AC" last-cp="03CE" tag="**sc:Grek**" … />
Han	<char cp="4E03" tag="**sc:Hani**" … />
Multiple scripts	<char cp="3006" tag="**sc:Hani sc:Hira sc:Kana**" … />

- Review code points for inclusion
  - GP must positively affirm each inclusion and give a rationale based on its research / alignment with principles in the [Procedure]
  - See Considerations document

Repertoire Considerations

- Many GPs may benefit from existing IDN tables

- However, the Root Zone is a shared resource
  - Broad context – “the entire Internet population” (RFC6912)
  - Necessitates a more restrictive LGR for the Root Zone

- Root Zone LGRs are different from 2nd Level IDN Tables
  - Script-level focus vs. language-level focus
  - No ASCII mixing – even though many IDN tables allow it
  - Variants and dispositions may differ from 2nd level

Determine Variants

- Decide whether there are any code point variants
- Determine their types and how they resolve into dispositions for variant labels
- Per the [Procedure], the goal is to:
  - *Clear the table of all the straightforward, non-subjective cases, mainly by returning a "blocked" disposition"*
- Considerations:
  - Minimize use of "allocatable" variants
- See Variant Rules document

Determine WLE Rules

- Decide if the use of any WLE rule is required
- WLE rules should balance security and simplicity
- A simple rule that lets through a small percentage of false negatives may be a good trade-off
- In many cases, instead of defining syntax for the entire label, it may be simpler to define the necessary contexts for code points (X must precede A, and follow B)
- See WLE Rules document

Coordination Between GPs

- When scripts are related, coordination between GPs is needed to ensure consistency between LGRs before submitting to IP

- In the interest of clarity, GPs with related scripts might produce two versions of its LGR
  - GP Script LGR containing only repertoire and variants relevant to the GP's script
  - Integrated LGR with other related-script GPs – incorporating their variant mappings (to make it symmetric and transitive)
    - Useful for community to understand how the LGR would affect them

Proposal Deliverables

- Formal XML definition of the LGR containing:
  - Code point repertoire
  - Variants (if applicable)
  - WLE rules (if applicable)
- Documented rationale
  - Choice of repertoire, coverage and contents
  - Necessity, choice and type of variants
  - Necessity and design of WLEs
  - Review in light of Process Goals and Principles in Procedure
- Plus: Examples of labels, variant labels and labels blocked by WLEs
  - Only needed if the LGR contains variants or WLEs
- Optional: Informative charts of the LGR repertoire
  - For example, like the annotated PDF files in the MSR
- See Requirements for LGR Proposals document

Throughout the Process

- Keep the Integration Panel in the loop
  - IP can only approve or reject the LGR proposal as a whole
  - Early discussions reduce the chance that some detail will lead to rejection

- Follow the Procedure
  - It is the authoritative prescription
  - The LGR Proposal must be compatible with its principles

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
- Overview documents (links in Document Repository)
  - Guidelines for developing script-specific Label Generation Rules for integration into the Root Zone LGR
  - Considerations for designing a Label Generation Ruleset for Root Zone
  - Requirements for LGR Proposals
- Background technical documents (links in Document Repository)
  - Variant rules
  - Whole Label Evaluation (WLE) rules
  - Representing Label Generation Rulesets using XML
- Foundation documents (links in Document Repository)
  - Procedure to Develop and Maintain the Label Generation Rules for the Root Zone in Respect of IDNA Labels
  - MSR-2

Integration Panel Discussion

How to Design Variants and WLE Rules

Michel Suignard
Integration Panel Member

Variant Basics

- Variants only exist for some scripts, many LGRs won't need them
- Variants must deal with a root zone which is language-neutral, script-based and shared
- Despite apparent restriction due to 'blocked' variants, number of permissible IDN root labels remains huge
- Variant code points only affect labels which otherwise would be identical

Variant Requirements

- Variant mappings must be
  - *Symmetric:* **A → B ⇒ B → A**
  - *Transitive:* **A → B** and **B → C ⇒ A → C**
- Variants that intersect scripts must be defined in each of these scripts
  - Example: ‘o’ in Latin, Greek and Cyrillic

Variant Categories and Types

- In-repertoire, within a single script
  - Variants within the scope determined by a GP
- Out-of-repertoire or across scripts:
  - Variants related to interaction with other GPs
  - For example: homoglyphs across scripts
- Types assigned to variants drive disposition for labels containing these variants
- Two default types:
  - Blocked
  - Allocatable

On the Use of Allocatable Variants

- Best for cases when all of these conditions apply:
  - In-repertoire
  - Variants are inherently the ‘same’ character, examples:
    - Medial form Arabic Yeh ـيـ versus Persian Yeh ـيـ
    - CJK Traditional 鍛 and simplified 锻
  - No easy way for some target users to input correct alternative
- Some cases best treated without using variants at all
  - Arabic/Latin characters with similar marks (handle confusables via String Review)
- Allocatable variants are hard to implement
  - Use to be minimized for all LGRs (blocked or no-variant are preferred options)

Blocked Variants Example: Greek

- In-repertoire
  - Sigma ‘ σ ’ versus final sigma ‘ ς’
- Variants with Latin (out-of-repertoire):
  - o, dotless i, ε , … alone or with additional diacritical marks
- Variants with Cyrillic (out-of-repertoire):
  - o, γ , …

Variants by Integration: Japanese

- Japanese LGR not expected to have its own variants
- Shared variant mappings:
  - Introduced because Root Zone is shared resource that also supports Chinese LGR
  - Can have variant types and disposition unique to the Japanese LGR (expected to be blocked)
  - May result in many distinct Japanese Kanjis blocking each other (in labels otherwise the same)
    - Example: 4E00 一, 58F1 壱, 58F9 壹, and 5F0C 弐 may block each other

Strategy for Creating Repertoire and Variants

1. Create a repertoire consistent with the scope and how the script is used (no out-of-repertoire code points)
2. Determine in-repertoire variants required by the GP (if any)
3. This results in a preliminary LGR corresponding to the need of the community, before integration with other LGRs
4. Through collaboration with GPs for related repertoires, add out-of-repertoire variants as blocked
5. Ensure consistency with mappings from related LGRs (dispositions on variants may be different)

Use of WLE Rules Use in Root Zone LGRs

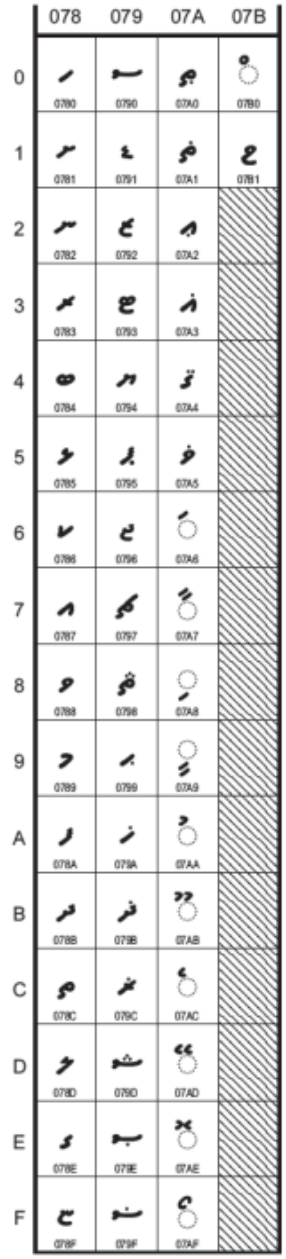
- No need for WLE Rules in many LGRs (complexity versus risk reduction)
- Intended for enforcing fundamental script rules to:
  - Determine required or prohibited context
  - Restrict combining sequences in alphabets
  - Enforce simple composition rules in alphasyllabaries (abugida)
- Not for enforcing spelling rules

WLE Example: Combining Macron Below

- Code point U+0331 COMBINING MACRON BELOW 
  - Rarely used in Latin repertoire for IDN because sequences are normalized out through the IDNA2008 process
  - However, it is used for some African letters that have no pre-composed forms
- A WLE Rule might be created to restrict usage to sequences where it follows ‘c’, ‘q’, ‘s’, and ‘x’
  - Only sequences where U+0331 is allowed are: <0063 0331>, <0071 0331>, <0073, 0331>, and <0078 0331>

WLE Example: Thaana

- Thaana script written in syllables, but encoded as an alphabet
- Set of rules to enforce that every syllable is well-formed
- Simple rules focused on immediate context for each code point
- All consonants (with one exception) must be followed by a vowel sign
- Only one vowel sign can follow a consonant



Basic consonants

0780 THAANA LETTER HAA
0781 THAANA LETTER SHAVIYANI
0782 THAANA LETTER NOONU
0783 THAANA LETTER RAA
0784 THAANA LETTER BAA
0785 THAANA LETTER LHAVIYANI
0786 THAANA LETTER KAAFU
0787 THAANA LETTER ALIFU
0788 THAANA LETTER VAAVU
0789 THAANA LETTER MEEMU
078A THAANA LETTER FAAFU
078B THAANA LETTER DHAALU
078C THAANA LETTER THAA
078D THAANA LETTER LAAMU
078E THAANA LETTER GAAFU
078F THAANA LETTER GNAVIYANI
0790 THAANA LETTER SEENU
0791 THAANA LETTER DAVIYANI
0792 THAANA LETTER ZAVIYANI
0793 THAANA LETTER TAVIYANI
0794 THAANA LETTER YAA
0795 THAANA LETTER PAVIYANI
0796 THAANA LETTER JAVIYANI
0797 THAANA LETTER CHAVIYANI

Extensions for Arabic

0798 THAANA LETTER TTAA
0799 THAANA LETTER HHAA
079A THAANA LETTER KHAA
079B THAANA LETTER THAALU
079C THAANA LETTER ZAA
079D THAANA LETTER SHEENU
079E THAANA LETTER SAADHU
079F THAANA LETTER DAADHU
07A0 THAANA LETTER TO
07A1 THAANA LETTER ZO
07A2 THAANA LETTER AINU
07A3 THAANA LETTER GHAINU
07A4 THAANA LETTER QAAFU
07A5 THAANA LETTER WAAVU

Vowels

07A6 THAANA ABAFILI
07A7 THAANA AABAAFILI
07A8 THAANA IBIFILI
07A9 THAANA EEBEEFILI
07AA THAANA UBUFILI
07AB THAANA OOBOOFILI
07AC THAANA EBEFILI
07AD THAANA EYBEYFILI
07AE THAANA OBOFILI
07AF THAANA OABOAFILI
07B0 THAANA SUKUN

Consonant for Addu dialect

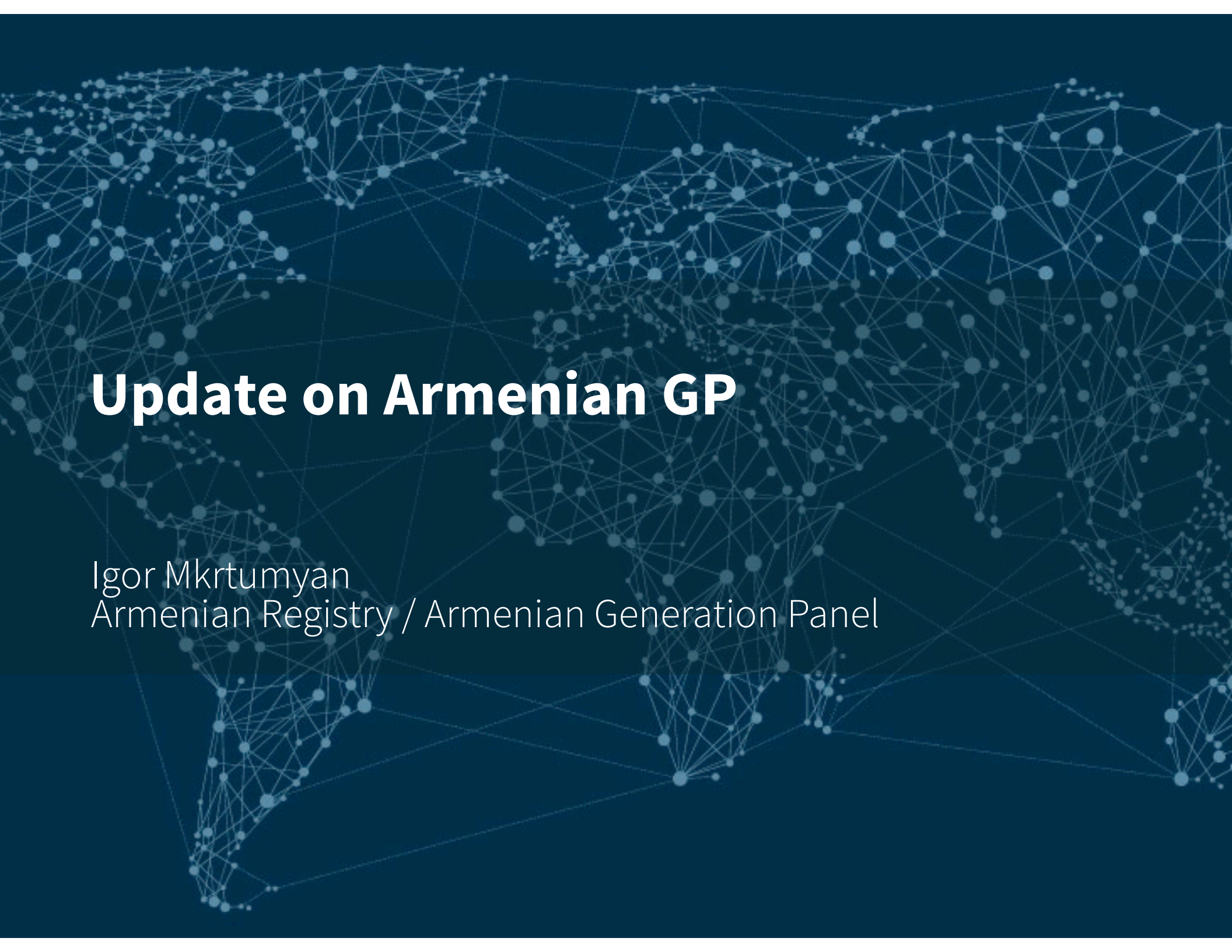
07B1 THAANA LETTER NAA

Conclusion

- Variants and WLE Rules are complex features that should be used sparingly
- Chance of acceptance of a LGR is greatly improved by:
  - Coordination and collaboration between GPs (when appropriate)
  - Interaction with the Integration Panel before formal submission

Resources

- Guidelines for Developing Script-Specific Label Generation Rules for Integration into the Root Zone LGR
  https://community.icann.org/download/attachments/43989034/Guidelines-for-LGR-2014-12-02.pdf
- Variants rules
  https://community.icann.org/download/attachments/43989034/Variant%20Rules.pdf
- Whole Label Evaluation (WLE) rules
  https://community.icann.org/download/attachments/43989034/WLE-Rules.pdf
- Requirements for LGR Proposals
  https://community.icann.org/download/attachments/43989034/Requirements%20for%20LGR%20Proposals.pdf
- Thaana LGR example
  https://github.com/kjd/lgr/blob/master/resources/Sample-LGR-Thaana.xml
- Greek LGR example
  https://github.com/kjd/lgr/blob/master/resources/Sample-LGR-Greek.xml



Update on Armenian GP

Igor Mkrtumyan
Armenian Registry / Armenian Generation Panel

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

Commonality

There are some commonality (visual similarity) with Latin, Greek and Cyrillic.

Composition of the Armenian General Panel

Name	Role
Igor Mkrtumyan	Chair
Grigori Saghyan	Expert
Lianna Galstyan	Expert
Vladimir Sahakyan	Expert
Anna Karakhanyan	Expert
Ruben Hakobyan	Expert
Kristina Babajanyan	Expert
Hrant Dadivanyan	Expert

Work Plan

- Creation of the Armenian GP mailing list
- Acceptance of MSR-2 for Armenian script
- Analysis of visually similar codes in lowercase Armenian scripts
- Analysis of visually similar codes in scripts having commonality with Armenian
- Development of presentation on Armenian GP proposal for IDN Program Update workshop at ICANN 52
- Collecting community opinion and remarks
- Development of a final report to IP
- Final decision on LGRs for the Armenian script

Proposed Schedule of Meeting and Teleconferences

Date	Name	Agenda
Dec 15, 2014	1st meeting of the GP	Setting the goals and time schedule. Distribution of tasks. Formation of small group according to tasks.
Jan 15, 2015	2nd meeting of the GP	Report of groups on the fulfilled jobs. Setting additional tasks.
Jan 30, 2015	3rd meeting of the GP	Combining reports to a presentation for IDN Program Update workshop at ICANN 52.
Feb 15, 2015	4th meeting of the GP	Processing opinions and remarks from ICANN 52 workshop.
Feb 27, 2015	5th meeting of the GP	Discussion of the draft report to the IP. Collecting final opinions.
Mar 15, 2015	6th meeting of the GP	Presentation of the final report to the GP.
Mar 31, 2015	7th meeting of the GP	Submission of the final report to the IP

Armenian GP Mailing List

- 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

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

Visual Similarity Evaluation (Armenian and Latin)

Armenian Script	Latin Script	Visual similarity
ԛ	q	ԛ - Armenian
		q - Latin
ԟ	q	ԟ - Armenian
		q - Latin
ԁ	d	ԁ - Armenian
		d - English
հ	h	հ - Armenian
		h - Latin
յ	j	յ - Armenian
		j - Latin

Armenian Script	Latin Script	Visual similarity
ո	n	ո - Armenian
		n - Latin
ս	u	ս - Armenian
		u - Latin
ց	g	ց - Armenian
	9 number)	ց - Latin
		9 - number
օ	o	օ - Armenian
		o - Latin

Visual Similarity Evaluation (Armenian and Greek)

Armenian	Description	Code Point	Greek	Description	Greek Code Point	Visual Similarity
ղ	ghad	572	η	eta	03b7	Similar
ւ	yiwn	582	ι	iota	03b9	Similar
օ	oh	585	ο	omicron	03bf	Identical

Visual Similarity Evaluation (Armenian and Cyrillic)

Armenian Script	Code Point	Cyrillic Script	Visual similarity
ш	561	ш (школа)	ш - Armenian
			ш - Cyrillic
п	578	п (пирог)	п - Armenian
			п - Cyrillic
о	585	о (окно)	о - Armenian
			о - Cyrillic
щ	057A	щ (щенок)	щ- Armenian
			щ - Cyrillic

Visual Similarity Evaluation

(within Armenian)

String	Script		Script	Script
ւո	տ		ռ	ո
ւի	փ		ֆ	դ
ււ	ս		շ	ջ
ււււ	ա		ե	է
ւս	ա		բ	ր
իւ	խ			

Conclusions

- *There are two standardized mutually intelligible modern literary forms, Eastern Armenian and Western Armenian, with different orthographies. But as a set of scripts, (MSR-2) is the same and not concluded to any LGR. As a result, the Armenian GP will not address in the LGR document issues arising from the different orthographies and the use of Armenian in the Diaspora in developing the LGR.*

Conclusions (continued)

- *Visual similarities will not be reflected in the LGR for the Root Zone. They will rather be solved by mechanisms beyond the application of the LGR that are expected to be part of the overall registration process. The problem will be solved by limiting Armenian domain names strictly to the Armenian MSR table, Latin dash (codepoint '2d') and Latin numbers (codepoints '30' - '39').*

Conclusions (continued)

- *Armenian GP anticipates that the relationships with the related scripts (Cyrillic, Greek, and Latin) would not affect the content of the Armenian LGR. 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 visually similar code points of related scripts. We are not sure that the same blocking mechanism will be implemented in other IDN domain registration procedures but it can be recommended to corresponding IDNs.*

Conclusions (continued)

- *The visual similarity of strings and scripts within Armenian IDN can be used by domain registrants for phishing or registering a domain similar to a brand domain. However we can not set any rule forbidding the visual similarity of domain names as there is no way to distinguish whether it is normal or intentional because we can't analyze thousands of brand names, trademarks and company names.*

Conclusions (continued)

The necessity of LGR should be evaluated yet after collecting the community opinion and remarks.

Thank you!

Update on Cyrillic Generation Panel

Yuriy Kargapolov
.УКР IDN ccTLD

Dusan Stojičević
.RS ccTLD / .СРБ IDN ccTLD

Cyrillic Generation Panel

General information

1. Script for which the panel is to be established – List the ISO 15924 script code (from http://www.unicode.org/iso15924/iso15924-codes.html)

2.

Code	N°	English Name	Nom français	Property Value Alias	Date
Cyrl	220	Cyrillic	cyrillique	Cyrillic	2004-05-01
Cyrs	221	Cyrillic (Old Church Slavonic variant)	cyrillique (variante slavonne)		2004-05-01

13 countries

108 languages

General information

3. Language groups that use the Cyrillic alphabet

1) Indo-European languages

Slavic group: **(1) Belarusian, (2) Bulgarian, (3) Macedonian, (4) Montenegrin, (5) Russian, (6) Serbian, (7) Ukrainian**

Iranian group: **Kurdish, Ossetian, (8) Tajik**

2) Sino-Tibetan languages: **Dungan**
3) Mongolian languages: **(9) Mongolian, Buryat, Khalkha, Kalmyk**
4) Turkic languages: **Bashkir, Chuvash, (10) Kazakh, Tatar, (11) Uzbek, (12) Kyrgyz, (13) Turkmen**
5) Uralic languages: **Komi-Permyak, Meadow Mari, Hill Mary, Kildin Sami**
6) Tungusic languages
7) Chukchi and Kamchatka languages
8) Individual languages - Aleutian, Nivkhs, Ket, Eskimos, Yukaghir languages

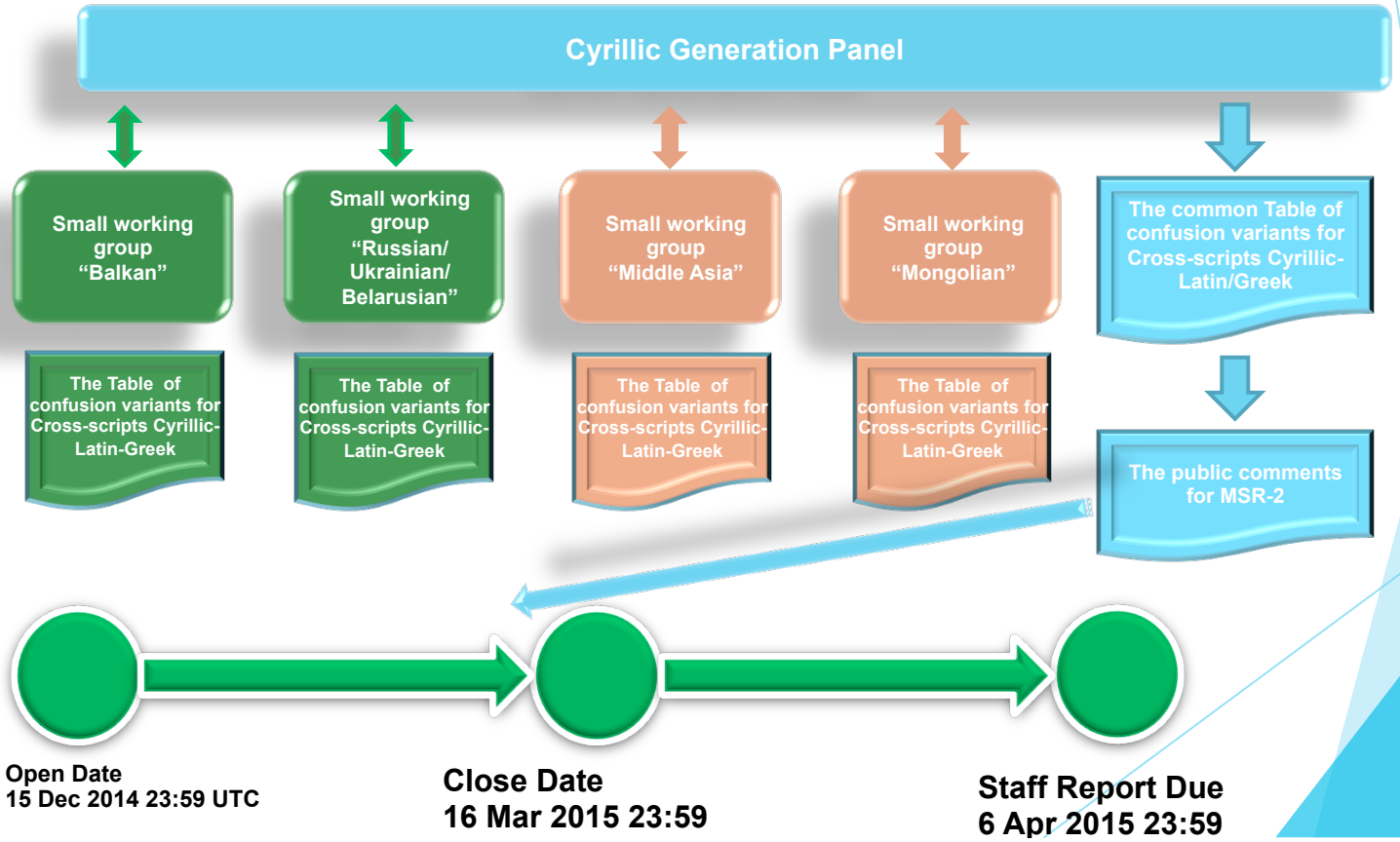
Researches on 95 ethnic minority languages in Russia weren't conducted

Cyrillic **maybe and structurally and historically was related with *Latin* and *Greek* but more detail should examined during work of Panel.**

General information

4. Structure of Cyrillic Generation Panel and organization of work

19 members from 12 countries



The Tables of confusion variants

A. The Greek point codes Table concerning which confusion variants were considered

					A	B	Γ	Δ	E	Z	H	Θ	I	K	Λ	M	N	Ξ	O	Π	P	Σ	Σ	T	Y	Φ	X	Ψ	Ω	Ϊ	Ÿ
					U+0391	U+0392	U+0393	U+0394	U+0395	U+0396	U+0397	U+0398	U+0399	U+039A	U+039B	U+039C	U+039D	U+039E	U+039F	U+03A0	U+03A1	U+03A2	U+03A3	U+03A4	U+03A5	U+03A6	U+03A7	U+03A8	U+03A9	U+03AA	U+03AB
ά	έ	ή	ί	ΰ	α	β	γ	δ	ε	ζ	η	θ	ι	κ	λ	μ	ν	ξ	ο	π	ρ	ς	σ	τ	υ	φ	χ	ψ	ω	ϊ	
U+03AC	U+03AD	U+03AE	U+03AF	U+03B0	U+03B1	U+03B2	U+03B3	U+03B4	U+03B5	U+03B6	U+03B7	U+03B8	U+03B9	U+03BA	U+03BB	U+03BC	U+03BD	U+03BE	U+03BF	U+03C0	U+03C1	U+03C2	U+03C3	U+03C4	U+03C5	U+03C6	U+03C7	U+03C8	U+03C9	U+03CA	

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
U+0041	U+0042	U+0043	U+0044	U+0045	U+0046	U+0047	U+0048	U+0049	U+004A	U+004B	U+004C	U+004D	U+004E	U+004F	U+0050	U+0051	U+0052	U+0053	U+0054	U+0055	U+0056	U+0057	U+0058	U+0059	U+005A
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
U+0061	U+0062	U+0063	U+0064	U+0065	U+0066	U+0067	U+0068	U+0069	U+006A	U+006B	U+006C	U+006D	U+006E	U+006F	U+0070	U+0071	U+0072	U+0073	U+0074	U+0075	U+0076	U+0077	U+0078	U+0079	U+007A

Latin)

	À	Á	Â	Ã	Ä	Å	Æ	Ç	È	É	Ê	Ë	Ì	Í	Î	Ï	Ð	Ñ	Ò	Ó	Ô	Õ	Ö	Ø	Ù	Ú	Û	Ü	Ý	Þ	
	U+00C0	U+00C1	U+00C2	U+00C3	U+00C4	U+00C5	U+00C6	U+00C7	U+00C8	U+00C9	U+00CA	U+00CB	U+00CC	U+00CD	U+00CE	U+00CF	U+00D0	U+00D1	U+00D2	U+00D3	U+00D4	U+00D5	U+00D6	U+00D8	U+00D9	U+00DA	U+00DB	U+00DC	U+00DD	U+00DE	
ß	à	á	â	ã	ä	å	æ	ç	è	é	ê	ë	ì	í	î	ï	ð	ñ	ò	ó	ô	õ	ö	ø	ù	ú	û	ü	ý	þ	ÿ
U+00DF	U+00E0	U+00E1	U+00E2	U+00E3	U+00E4	U+00E5	U+00E6	U+00E7	U+00E8	U+00E9	U+00EA	U+00EB	U+00EC	U+00ED	U+00EE	U+00EF	U+00F0	U+00F1	U+00F2	U+00F3	U+00F4	U+00F5	U+00F6	U+00F8	U+00F9	U+00FA	U+00FB	U+00FC	U+00FD	U+00FE	U+00FF

The Tables of confusion variants

1. Cross-scripts Cyrillic (Russian/Ukrainian/Belarusian/Balkan segment) – Greek

The Table presents only those Greek code points which could be described as confounding

Russian/ Ukrainian/ Belarusian/Balkan	**А**	**В**	**Г**	**Ґ**	**Ѓ**	**Е**	**І**	**К**	**Ќ**	**Л**	**М**	**Н**	**О**	**П**	**Р**	**Т**	**У**	**Ў**	**Ф**	**Х**	**Ш**
Greek Capital	**Α**	**Β**	**Γ**	**Γ**	**Γ**	**Ε**	**Ι**	**Κ**	**Κ**	**Λ**	**Μ**	**Η**	**Ο**	**Π**	**Ρ**	**Τ**	**Υ**	**Υ**	**Φ**	**Χ**	
Russian/ Ukrainian/ Belarusian/Balkan	**а**	**в**	**г**	**ґ**		**е**	**і**	**к**	**ќ**	**л**	**м**	**н**	**о**	**п**	**р**	**т**	**у**	**ў**	**ф**	**х**	**ш**
Greek small	**α**	**β**				**ε**	**ι**	**κ**	**κ**	**λ**			**ο**	**π**	**ρ**	**τ**	**γ**	**γ**		**χ**	**ω**

Were analyzed case for presence of cross-script Homoglyphs

Analysis on the script-internal Homoglyphs case not performed

The **same** visual code points marked **by green**

The **similar** visual code points marked **by blue**

The Tables of confusion variants

2. Cross-scripts Cyrillic (Russian/Ukrainian/Belarusian/Balkan segment) – Latin

The Table presents only those Latin code points which could be described as confounding

Russian/ Ukrainian/ Belarusian/ Balkan	**А**	**В**	**Г**	**Ґ**	**Е**	**S**	**И**	**І**	**Ї**	**J**	**К**	**Ќ**	**М**	**Н**	**О**	**П**	**Р**	**С**	**Т**	**У**	**Ў**	**Х**	**Ь**											
Latin Capital	**A**	**B**			**E**	**S**		**I**	**I**	**J**	**K**	**K**	**M**	**H**	**O**		**P**	**C**	**T**	**Y**	**Y**	**X**												
Russian/ Ukrainian/ Belarusian/ Balkan	**а**	**в**	**г**	**ґ**	**е**	**s**	**и**	**і**	**ї**	**j**	**к**	**ќ**	**м**	**н**	**о**	**п**	**р**	**с**	**т**	**у**	**ў**	**х**	**ь**											
Latin (Expanded - IDN)		**À**	**Á**	**Â**	**Ã**	**Ä**	**Å**	**Æ**	**Ç**	**È**	**É**	**Ê**	**Ë**	**Ì**	**Í**	**Î**	**Ï**	**Ð**	**Ñ**	**Ò**	**Ó**	**Ô**	**Õ**	**Ö**	**Ø**	**Ù**	**Ú**	**Û**	**Ü**	**Ý**	**Þ**	
Cyrillic Capital									**С**	**Ё**	**Ё**	**Ё**	**Ё**	**І**	**І**	**Ї**	**Ї**														**Ў**		
Latin (Extended) small	**ß**	**à**	**á**	**â**	**ã**	**ä**	**å**	**æ**	**ç**	**è**	**é**	**ê**	**ë**	**ì**	**í**	**î**	**ï**	**ð**	**ñ**	**ò**	**ó**	**ô**	**õ**	**ö**	**ø**	**ù**	**ú**	**û**	**ü**	**ý**	**þ**	**ÿ**
Cyrillic small	**в**								**с**	**ё**	**ё**	**ё**	**ё**	**і**	**і**	**ї**	**ї**										**й**	**й**	**й**	**й**	**ў**		**ў**

Were analyzed case for presence of cross-script Homoglyphs

Analysis on the script-internal Homoglyphs case not performed

The **same** visual code points marked **by green**

The **similar** visual code points marked **by blue**

The Tables of confusion variants

Homoglyphs of Punctuation

In current version **MSR-2** the code point **U+02BC (Modify Letter Apostrophe)** indicated as one of *Homoglyphs of Punctuation*

Important note. This is code point in Cyrillic alphabets **not a punctuation sign** – this is **a LETTER** for Ukrainian and Belorussian languages:

a) **In Ukrainian this letter to do same function as Russian letter «ь» (U+044C Cyrillic Small letter Soft Sign);**

b) **in Belarusian this letter to do same function as Russian letter «ъ» (U+044A Cyrillic Small letter Hard Sign)**

The letters «ʼ» and «ъ» ***can't (cannot)*** be the first or last letter of any word, only in the middle.
The letter «ь» can be the last letter or in the middle of word, but ***can't (cannot)*** be the first letter of any word.

can't	**=**	**cannot**	**м'ясо**	**!=**	**мясо**
English	*sign*	*English*	*Ukrainian*	*letter*	*Russian*

Conclusions

Cyrillic Generation Panel:

a) **considered confusion options only for cases of "external" cross-scripts;**

b) **has done work which gave preliminarily results for cases of confusion variants relative 2 regions within Cyrillic scripts: Balkan and Russian/ Ukrainian/ Belarusian;**

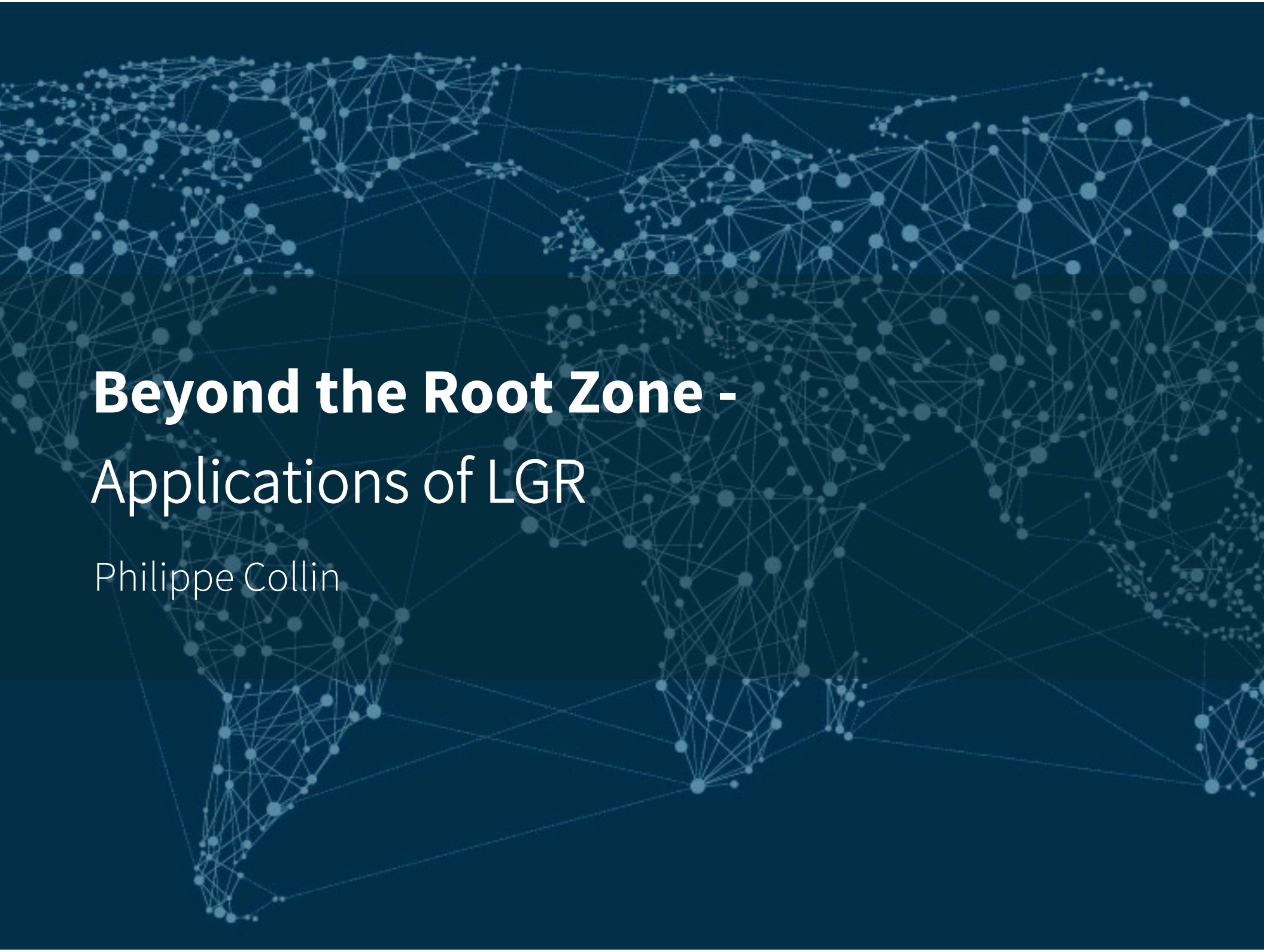
c) **can't form a complete and balanced position to do full public comment version 2 of the Maximal Starting Repertoire (MSR-2) at this moment; but the Cyrillic Generation Panel will make all possible efforts to make its proposals in due time (16 Mar 2015 23:59 UTC);**

d) **has no data on the analysis of possible options for the confusion variants for two regions within Cyrillic scripts: Mongolian and Middle Asia;**

e) **however, the unit will prepare some recommendations based on available data;**

f) **potentially can to form position on develop policy recommendations which can form base for LGR *(develop policy should be evaluated after collecting the Cyrillic community opinion and remarks).***

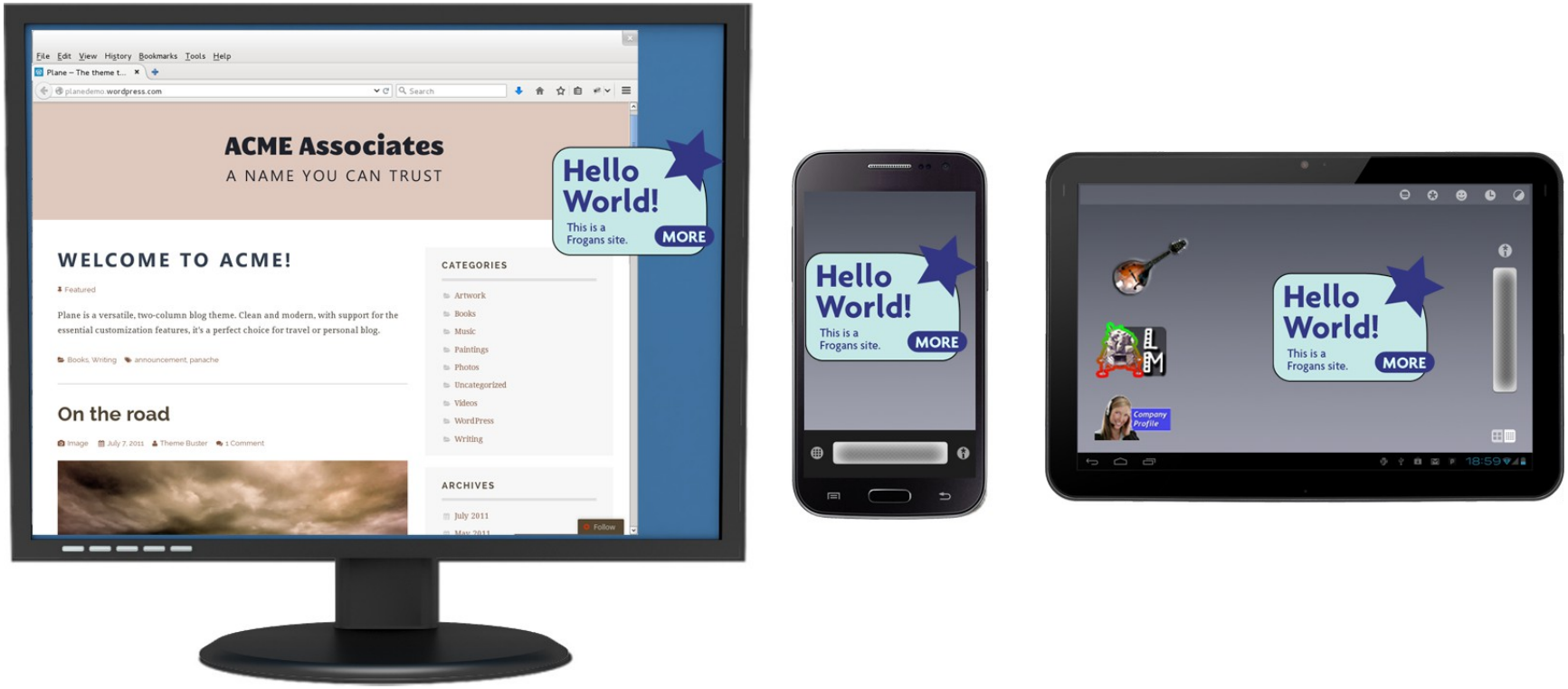
Thanks!



Beyond the Root Zone - Applications of LGR

Philippe Collin

Frogans sites: small, secure, multi-platform, multi-device

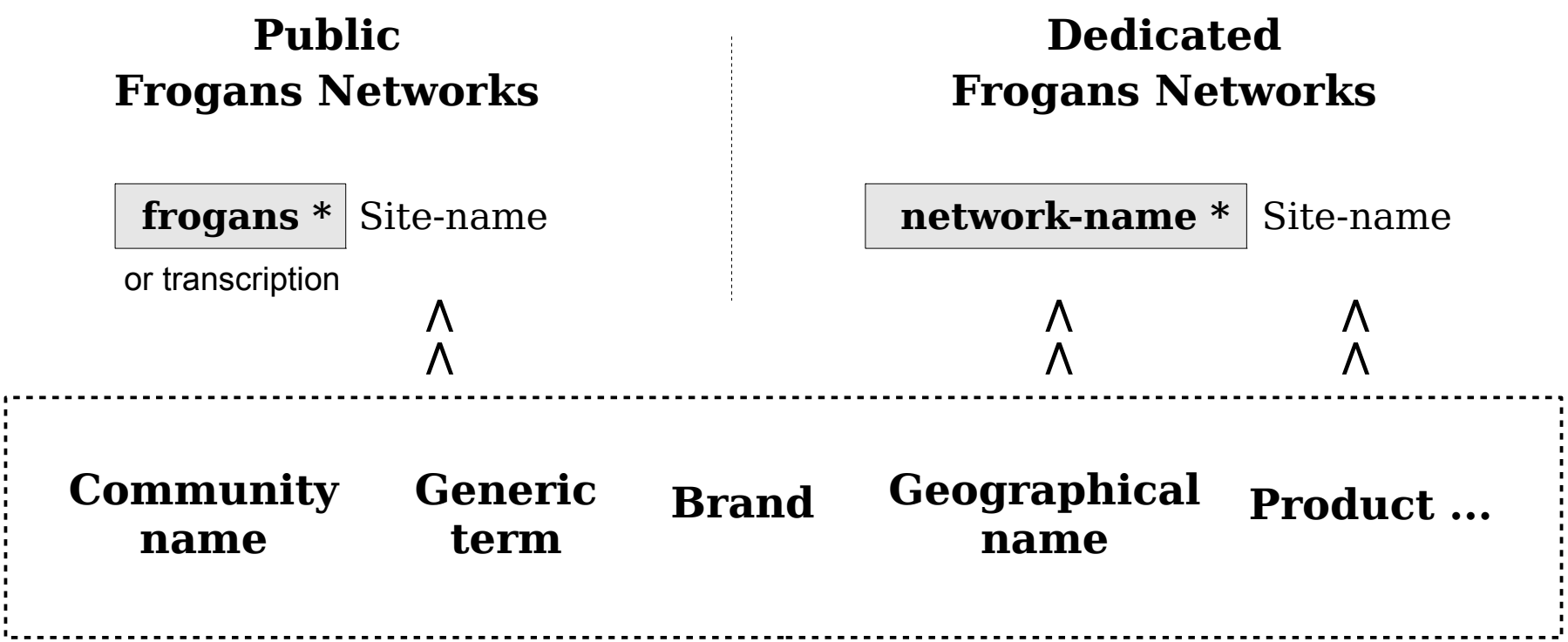


- Publication of a new type of site with a new international addressing system
- Same browsing experience and display across all devices
- Sites viewable via Frogans Player downloadable free of charge from the OP3FT

Frogans addresses within the Internet addressing environment

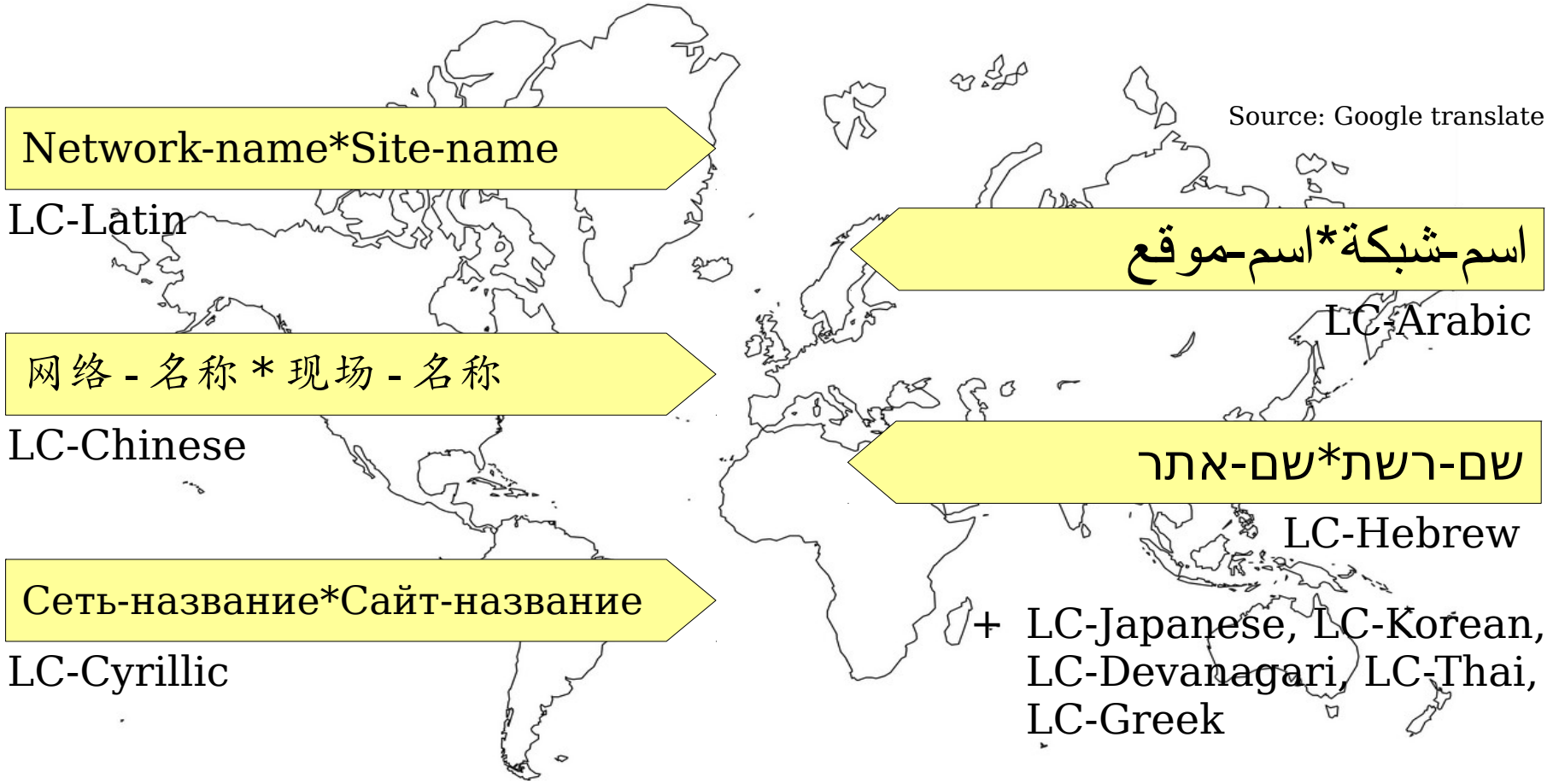
- Frogans addresses represent a new market
- Frogans addresses are used to identify Frogans sites
- Frogans addresses do not replace domain names

Frogans networks on the Internet: sets of Frogans addresses



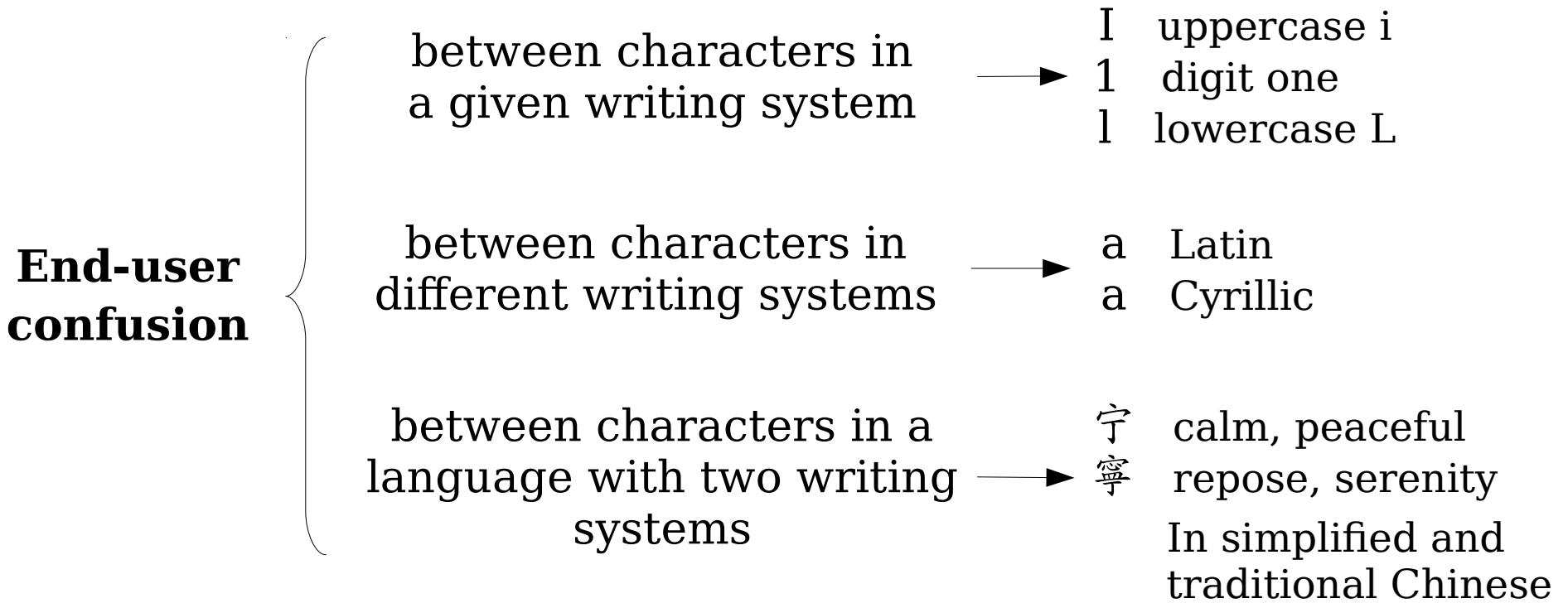
- Two types of Frogans networks on the Internet
- Customizable network names for Dedicated Frogans Networks
- Supports writing systems from all around the world

Frogans addresses: 10 linguistic categories



- Covers at least 179 languages
- Each linguistic category has its own set of rules
- The linguistic category doesn't concern the content of the Frogans site

Frogans addresses: managing confusion



- Raises potential security issues for end users
- The most important issue relates to spoofing
- Currently focused on visual and semantic confusion

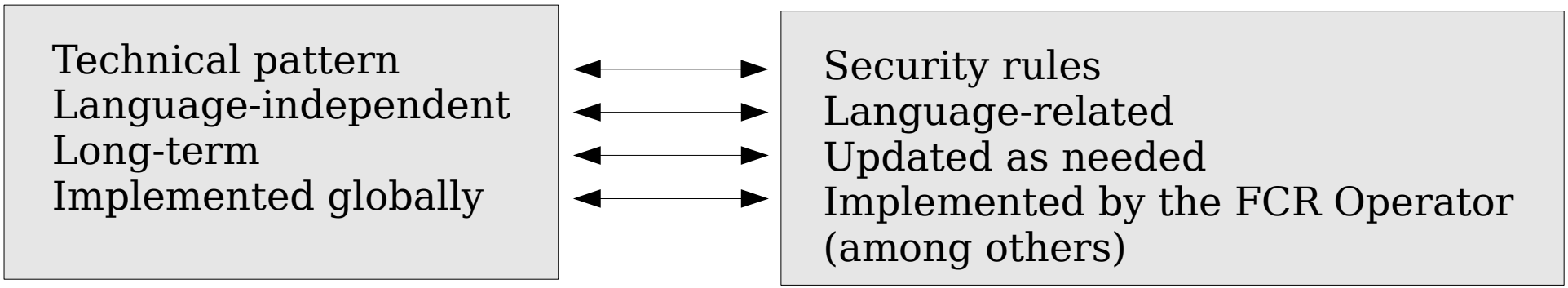
A two-part model for specifying Frogans addresses

IFAP

International
Frogans
Address
Pattern

FACR

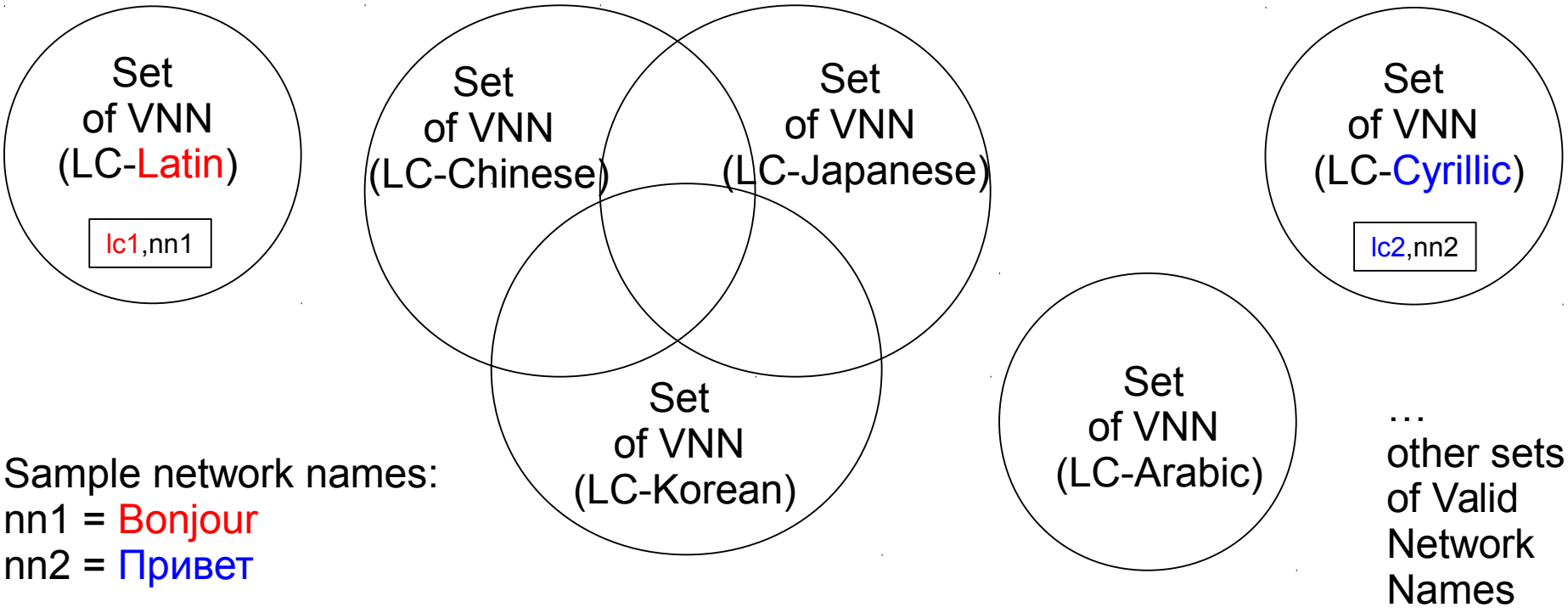
Frogans
Address
Composition
Rules



★ Purely technical approach insufficient →
FACR/IFAP are supported by **FTUP, UDRP-F,** and **end-user awareness**

- Called for by the OP3FT Bylaws
- Provide stability for a widely distributed and installed technology
- Provide flexibility and reactiveness demanded to solve security issues

FACR: Overlapping linguistic categories



- Ten linguistic categories available in FACR 1.0
- Specific rules are defined for each Linguistic Category
- Convergence forms are defined both within each LC and between LCs

A few examples

	Valid network name		Invalid Network name	
Latin	**vidéo** 0076 0069 0064 00E9 006F	Latin	**vidéo** 0076 0069 0064 0065 0301 006F	Network name is not IFAP compliant (not in NFKC form)
Latin	**hello**	Latin	**hello** 0068 0065 0196 0196 006F *0196 LATIN CAPITAL LETTER IOTA*	Network name 2 is not FACR valid (U+ 0196 is not employable)
Latin	**paypal**	Latin	**paypal** 0070 0430 0079 0070 0061 006C *0430 CYRILLIC SMALL LETTER A*	Network name 2 is not FACR valid (U+ 0430 is not employable)
Latin	**HELLO**	Chinese	**HELLO**	Network name 2 is not FACR valid (missing a character of native scripts)
Latin	**hello**	Latin	**Hello**	The 2 network names are identical (IFAP)
Latin	**straße** 0073 0074 0072 0061 00DF 0065	Latin	**strasse** 0073 0074 0072 0061 0073 0073 0065	The 2 network names are identical (IFAP)
Latin	**HELLO**	Latin	**HELL0**	The 2 network names have the same Intra-LC convergence form
Latin	**amis**	Latin	**arnis**	The 2 network names have the same Intra-LC convergence form (Latin-Confusable)
Japanese	へ 3078 (Hiragana)	Japanese	へ 30D8 (Katakana)	The 2 network names have the same Intra-LC convergence form (Japanese Confusable)
Chinese or Japanese	醜 919C	Chinese or Japanese	丑 4E11 (Simplified Chinese variant of 919C)	The 2 network names have the same Intra-LC convergence form (Chinese Variant)
Latin	**scope**	Cyrillic	**scope** 0455 0441 043E 0440 0435	The 2 network names have the same Inter-LC convergence form
Latin	**BEAT**	Greek	**BEAT** 0392 0395 0391 03A4 **βεατ** 03B2 03B5 03B1 03C4	The 2 network names have the same Inter-LC convergence form

Thank you for your attention!

- Welcome to the Frogans project
  https://project.frogans.org/
- The official Web site of the Frogans technology:
  https://www.frogans.org/
- International Frogans Address Pattern (IFAP) technical specification:
  https://www.frogans.org/en/resources/ifap/access.html
- Frogans Address Composition Rules (FACR) technical specification:
  https://www.frogans.org/en/resources/facr/access.html
- The UDRP-F and its Rules of procedure:
  https://www.frogans.org/en/resources/udrpf/access.html
- The Frogans Technology Conference:
  https://conference.frogans.org/
- The Frogans technology mailing lists:
  https://lists.frogans.org/

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