





IDN Program Update

IDN Program | ICANN 58 | 15 March 2017

Overview of Session Presentations

- IDN Program Overview and Progress
- Sarmad Hussain

Update by Integration Panel

- Marc Blanchet

Update on IDN Impl. Guidelines

- Mats Dufberg

- Community Updates
 - Cyrillic GP
 - Ethiopic GP
 - Greek GP
 - Latin GP

- Dusan Stojicevic
- Dessalegn Mequanint
- Panagiotis Papaspiliopoulos
- Mirjana Tasic

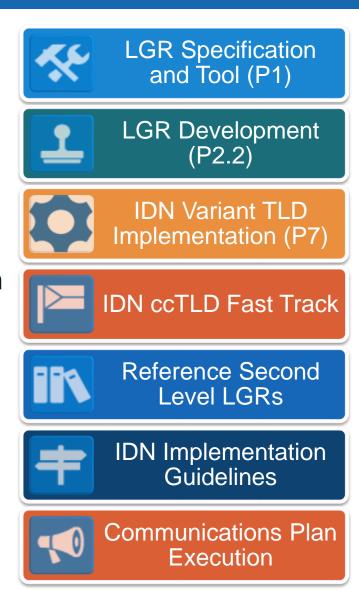
 \odot Q/A





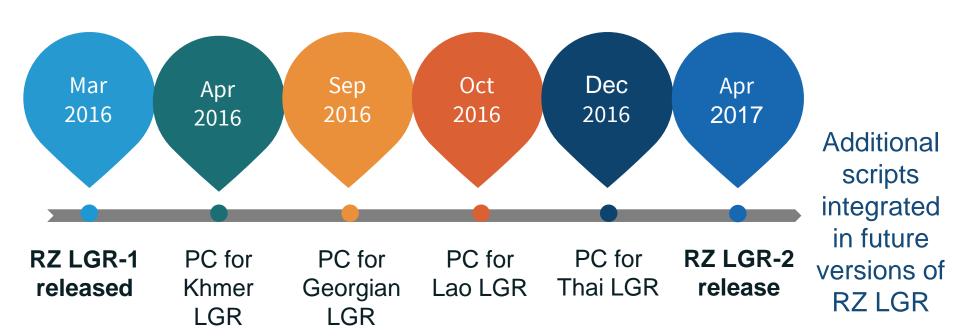
Overview of Presentation

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





Root Zone Label Generation Rules (RZ LGR)

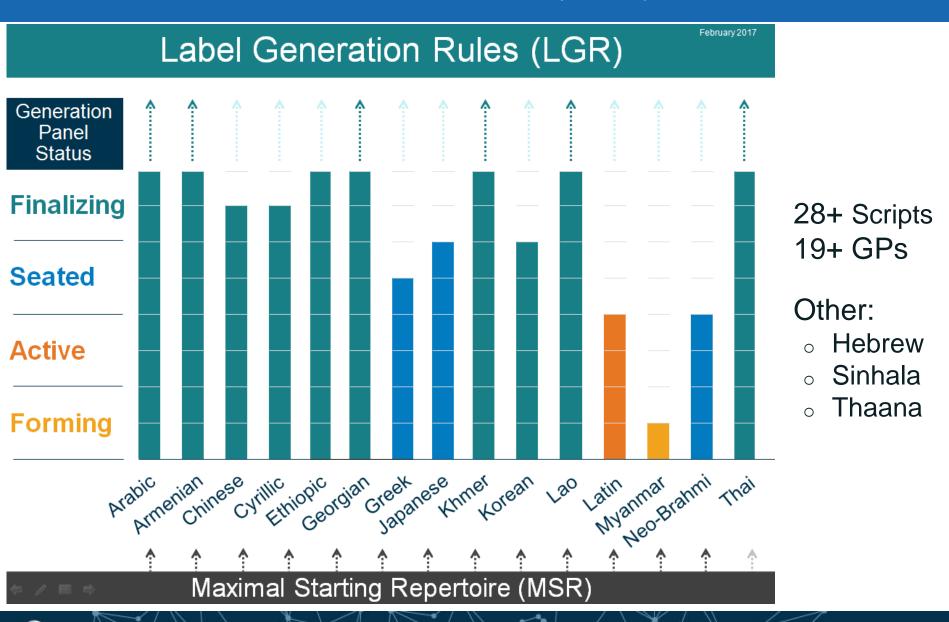


LGR-2 Summary

- Arabic, Georgian, Khmer, Lao and Thai scripts integrated
- Armenian finalized integrated after LGRs of related scripts received



Status of Generation Panels (GPs)



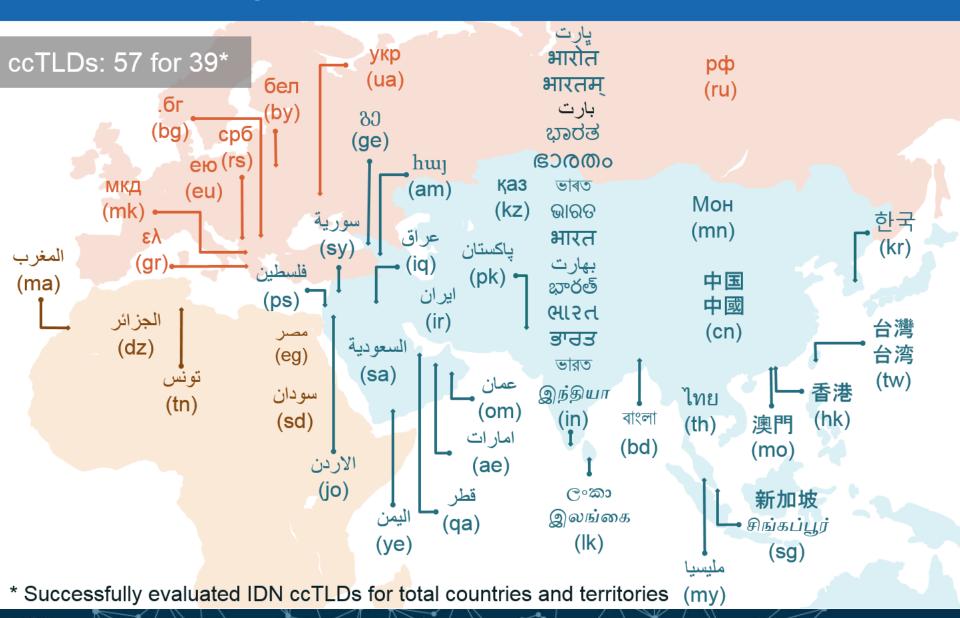


LGR Toolset (beta)

- Label Generation Rulesets (LGRs) used to generate domain name labels, as specified in <u>RFC 7940</u>
- Adding support for "merged" and "element" LGRs for the root zone
- LGR Toolset currently allows for the following:
 - Create a LGR
 - Use a LGR to validate a label and determine its variants
 - Manage LGRs, by comparing or combining them
 - Review impact of a new or a revised LGR on existing labels
- Online beta deployment at: https://lgrtool.icann.org/
- Open source package(s) released with BSD license
 - Released at github: <u>lgr-core</u>, <u>lgr-django</u>, <u>munidata</u>
- User guide available for further details



IDN Country Code Top-Level Domains





IDN ccTLD Fast Track Process

- 57 IDN ccTLDs evaluated representing 39 countries/territories
- 48 IDN ccTLDs delegated representing 38 countries/territories
- Requests cover 21 scripts for 35 languages

Currently under review

- <u>Public comment</u> in Jan 2015 raised issues with second similarity review process (EPSRP)
- Board resolution in June 2015 for ccNSO to review EPSRP (with input from GAC and SSAC)
- ccNSO formed Working Group (WG) on EPSRP
- Public comment in July 2016 on updated EPSRP guidelines
- <u>Final report</u> published, incorporating feedback and discussion
- ccNSO adopted the final report by WG on EPSRP



Reference Second Level LGRs / IDN Tables

- Reference IDN tables in LGR format in RFC 7940
- Developed using <u>Guidelines</u>, vetted for security and stability
- Finalized after a public comment process
- Can be referred to, as registries enhance IDN implementations
- Currently 28 LGRs <u>published</u>
 - Latin: Bosnian, Danish, English, Finnish, French, German,
 Hungarian, Icelandic, Italian, Latvian, Lithuanian, Norwegian,
 Polish, Portuguese, Spanish, Swedish
 - Cyrillic: Belarusian, Bosnian, Bulgarian, Macedonian,
 Montenegrin, Russian, Serbian, Ukrainian
 - Mixed scripts: Korean, Japanese
 - Others: Chinese, Hebrew



Communication and Outreach Efforts

- ⊙ IDN web pages at <u>icann.org/idn</u>
- IDN Program sessions at ICANN meetings
- IDN Program updates to SOs/ACs at ICANN meetings
- Direct outreach
 - SANOG, January 2017, Pakistan
 - IGF by GSE team, December 2016, Mexico
 - UbuntuNet-Connect by GSE team, Nov., 2016, Uganda
- IDN community wiki pages
- IDN mailing lists
 - {vip, Igr, ArabicGP, ArmenianGP, ChineseGP, ...}@icann.org

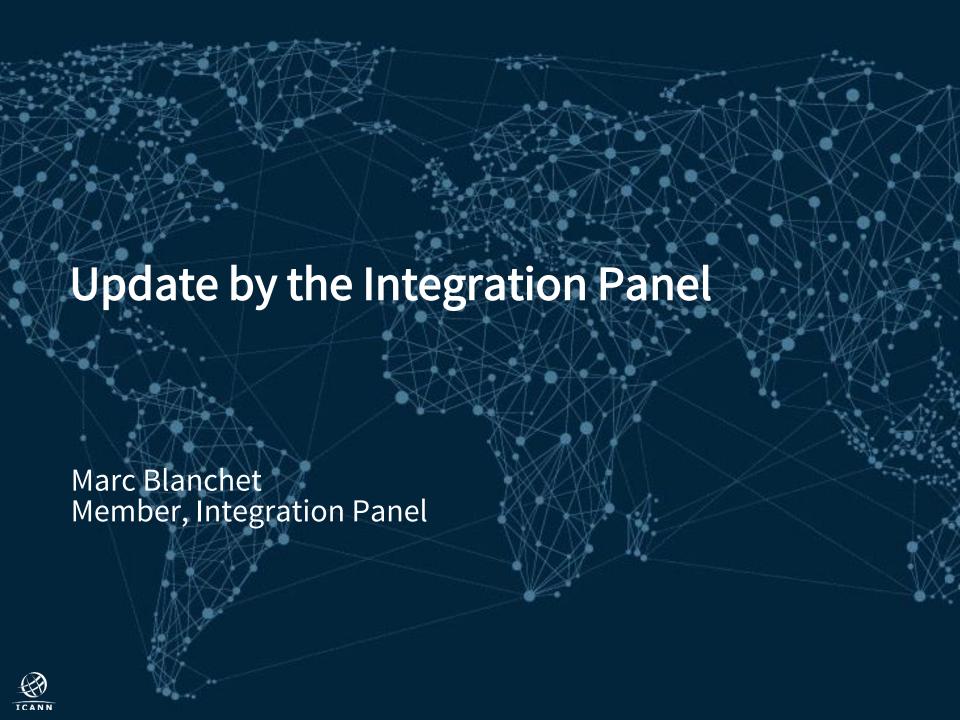


Thank You

For information on IDN Program projects, please visit:
 www.icann.org/idn

For queries regarding the IDN Program, please email:
 IDNProgram@icann.org





Integration Panel

"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



Activities since ICANN 57

Scripts:

- o Ethiopic:
 - Reviewing final drafts in preparation for public comments
- Georgian:
 - Went to public comments. Being integrated into Root Zone LGR-2
- o Korean:
 - Multiple interactions and early drafts
- o Chinese:
 - Multiple interactions and early drafts
- Cyrillic:
 - Reviewing final drafts in prep for public comments
- o Thai:
 - Went to public comments. Being integrated into Root Zone LGR-2
- Lao:
 - Went to public comments. Being integrated into Root Zone LGR-2
- o Neo-Brahmi:
 - Reviewed initial proposal for Devanagari



Activities since ICANN57 (Continued)

- Maximal Starting Repertoire (MSR)
 - No update planned yet, but two (2) Han candidates being discussed as possible additions, waiting for GP input
- Variant Rules document
 - Being updated
 - Also target for IETF RFC
- LGR-2 Integration
 - New scripts: Georgian, Khmer, Thai, Lao
 - Arabic carried from LGR-1
 - Armenian accepted (for LGR-1) but deferred, so that related scripts can be considered together
 - LGR-2: is being finalized in preparation for public comment
 - Details presented at the workshop



Update on IDN Implementation Guidelines

Mats Dufberg Co-Chair, IDN Guidelines Working Group



Background and Purpose

Purpose

- For second-level IDN registration policies and practices
- To minimize the risk of cybersquatting and consumer confusion

Relevance

- For gTLD registries and registrars offering IDNs at second level
- For IDN ccTLDs

Status

- Final draft for Public Comment released on 3 March 2017
- Interim draft presented at ICANN 57
- Issues list presented at ICANN 55
- Working Group formed in October 2015
- Call for Community Experts in July 2015



IDN Guidelines WG Members (IDNGWG)

	Name	Organization	SO/AC
1	Satish Babu	ISOC-TRV	ALAC
2	Wael Nasr	TLDVILLA LLC	ALAC
3	Mats Dufberg	IIS	ccNSO
4	Pablo Rodríguez	Puerto Rico TLD	ccNSO
5	Edmon Chung		GNSO
6	Christian Dawson	i2Coalition	GNSO
7	Chris Dillon		GNSO
8	Kal Feher	Neustar	GNSO
9	Dennis Tan	Verisign	GNSO
10	Jian Zhang	KNET	GNSO
11	Ram Mohan	Afilias	SSAC
12	Patrik Fältström (will only review)		SSAC



Topics and Recommendations

- Total of 6 topics and 18 recommendations:
 - 1. Transition (5)
 - Format of IDN Tables (2)
 - 3. Consistency of IDN Tables and Practices (4)
 - 4. IDN Variant Labels (2)
 - 5. Similarity and Confusability of IDN Labels (4)
 - 6. Terminology (1)
 - 7. Registration Data (none?)
 - 8. EPP (none?)

Appendix B: Glossary of Relevant Terms



Thank You

- For detailed guidelines on the topics, please visit:
 - Public Comment on IDN Implementation Guidelines : https://www.icann.org/public-comments/idn-guidelines-2017-03-03-en
 - IDN Guidelines WG Wiki page: <u>https://community.icann.org/display/IDN/IDN+implementation+Guidelines</u>
- ⊙ For feedback, email at:
 - o idngwg@icann.org or IDNProgram@icann.org





Introduction – Script of LGR

- Latin transliteration of native script name: Cyrillic
- Maximal Starting Repertoire (MSR) version: MSR-2



Background on Script and Principal Languages

- Based on Early Cyrillic, from First Bulgarian Empire in 9th century AD
- Used for languages across Eastern Europe and north and central Asia
- Basis of alphabets in languages, past and present, especially those of Slavic origin, and non-Slavic languages influenced by Russian
- Used by more than 250 million people as the official script for their languages, about half from Russia
- With the accession of Bulgaria to the European Union in 2007, Cyrillic became the third official script of the European Union, in addition to the Latin and Greek scripts



Geographic Territories Spread of Cyrillic

- Southeastern part of Europe (Serbia, Montenegro, Macedonia, Bulgaria, Bosnia and Herzegovina)
- Eastern Europe (Belorussia, Ukraine, Russia)
- Central Asia (Kazakhstan, Turkmenistan, Uzbekistan, Kyrgyzstan, Tajikistan, Mongolia)

is the only official orthography

is the only official orthography, but others are recognized for national or regional languages





Methodology

- According to work plan in the proposal for Cyrillic script GP
- Initially language based repertoire compiled, based on second-level IDN tables used by different ccTLDs, including the .su ccTLD which contained inventory for languages currently spoken in Russia
- Language repertoires collated in a face-to-face meeting in Istanbul on 25-26 Nov. 2016
- Continued to use the mailing list to share and finalize documents
- Consulted with Integration Panel (IP), including on crucial query regarding inclusion of U+02BC MSR for Ukrainian and Belarusian



Code Point Repertoire

	#	Unicode CP	Glyph	Unicode Name	Lang. using CF	EGIDS	Ref.
84 code points	1	04EB	Ö	O WITH DIAERESIS	Khanty	Khanty 6b	Rule 6 http://www.omniglot.co m/writing/khanty.htm
recommended for inclusion	2	04ED	ë	CYRILLIC SMALL LETTER E WITH DIAERESIS	Sami	Sami 8b	Rule 6
	3	04DB	ë	CYRILLIC SMALL LETTER SCHWA WITH DIAERESIS	Khanty	Khanty 6b	Rule 6 http://www.omniglot.co m/writing/khanty.htm
8 code points recommended for exclusion	4	04C2	ж	CYRILLIC SMALL LETTER ZHE WITH BREVE	Gagauz	Gagau z 5	Rule 5 http://www.omniglot.co m/writing/gagauz.htm Gagauz alphabet not in Cyrillic from 1996
(shown in the table)	5	04CC	Ą	CYRILLIC SMALL LETTER KHAKASSIAN CHE	Khakas	Khakas 5	Rule 5 http://www.omniglot.co m/writing/khakas.htm
	6	04CF	I	CYRILLIC SMALL LETTER PALOCHKA	Khakas	Khakas 5	Rule 5 http://www.omniglot.co m/writing/khakas.htm
	7	045D	Й	CYRILLIC SMALL LETTER I WITH GRAVE	Historic al sign		Rule 6
TCANN TCANN	8	0450	è	CYRILLIC SMALL LETTER IE WITH GRAVE	Stressed sign		Rule 6

Cyrillic Script Variants

- No variants in Cyrillic script
 - Some code points visually confusable
 - not considered as variants by the Cyrillic community
 - provide table of confusable code points, so organizations can use as needed



Cross-Script Variants

- Decided to limit these to homoglyphs
- Included code points which are homoglyphs in the lower case but not homoglyphs in the upper case
 - Only lower case because upper case disallowed in IDNA 2008
 - Decision made in consultation with IP ("the IP, at this point, does not require that upper case homoglyphs are included")
- Ocyrillic GP found cross-script variants with:
 - Armenian
 - o Greek
 - o Latin
- Cyrillic GP did not find cross-script variants with Georgian



Cross-Script Variants – with Armenian Script

- Armenian GP indicates three (3) variants with Cyrillic script
- Opinion of Cyrillic GP that only one (1) homoglyphic variant
- Other two (2) not identical, so included in confusables table

	Armenian glyph	Armenian code point	Cyrillic glyph	Cyrillic code point
(0	0585	0	043E



Cross-Script Variants – with Greek Script

Cyrillic has three (3) homoglyphic variants with Greek script

Greek glyph	Greek code point	Cyrillic glyph	Cyrillic code point
К	03BA	К	043A
О	03BF	0	043E
ф	03C6	ф	0444



Cross-Script Variants – with Latin Script

- Cyrillic has following homoglyphic variants with Latin from MSR-2
 - Confusables listed separately

Latin glyph	Latin code point	Cyrillic glyph	Cyrillic code point
а	0061	а	0430
С	0063	С	0441
е	0065	е	0435
0	006F	0	043E
i	0069	i	0456
j	006A	j	0458
1	006C	I	04CF
р	0070	р	0440
S	0073	S	0455
У	0078	У	0443
Х	0079	Х	0445
ä	00E4	ä	04D3
ë	00EB	ë	0451
æ	00E6	æ	04D5
Э	01DD	ə	04D9



Next Steps

- 1. Short phase (planned for this session)
 - i. Finalize proposal, based on IP feedback
 - ii. Finalize XML and test data files
 - iii. Issue the LGR for public comment
- 2. After public comment phase
 - iv. Finalize the LGR proposal to include community feedback
- 3. Long-term phase
 - v. Address new code points included in the MSR in the future
 - if needed in the root zone LGR, GP to re-convene and create additional proposal



Members of Cyrillic GP

Alex Khmyl (Belarus)	Nelly Stoyanova (Bulgaria)
Alexei Sozonov (Russia)	Nodir Mirzoev (Tajikistan)
Almaz Bakenov (Kyrgyz Republic)	Oleksandr Tsaruk (Ukraine)
Daniel Kalchev (Bulgaria)	Pavel Gusev (Kazakhstan)
Dmitry Belyavskiy (Russia)	Predrag Lesic (Montenegro)
Dmitry Kohmanyuk (Ukraine)	Sanja Simonova (Macedonia)
Dušan Stojičević (Serbia, chair)	Sergey Povalishev (Belarus)
Dušan Stojičević (Serbia, chair) Enkhbold Gombo (Mongolia)	Sergey Povalishev (Belarus) Tattu Mambetalieva (Kyrgyzstan)
Enkhbold Gombo (Mongolia)	Tattu Mambetalieva (Kyrgyzstan)
Enkhbold Gombo (Mongolia) Iliya Bazlyankov (Bulgaria)	Tattu Mambetalieva (Kyrgyzstan) Yashar Hajiyev (Azerbaijan)



Thanks!



Dušan Stojičević Душан Стојичевић

dusan@dukes.in.rs stojicevic@gransy.com





Agenda

- Introduction to Ethiopic Script
- Overview of GP Members
- Challenges
- Progresses and Timeline for Remaining Activities



Introduction to Ethiopic Script

1 1600+ old "Syllabary"

Used for the National Languages of Ethiopia and Eritrea

Languages or Writing Systems Using the Language:

Several languages use the script in their writing systems. However, only eight are considered for the LGR proposal:

Amharic, Argobba, Awngi, Harari, Hamt'agna,
 Silt'e, Tigrè, Tigrigna



3

Introduction to Ethiopic Script

Users Community

Eritrea, Djibouti, Egypt, South Africa, South Sudan, Kenya, United States, Canada, Israel, Europe, Australia

5

Each Letter of the "Syllabary" has both a consonant and a vowel

- o h = ke o h = ki
- ο h = kuο h = ka

6

Code Point Redundancy or Phonemic Decay

There are about 98 code points with sound and meaning alike in Amharic language writing system

- ሐልዮ Song
- *ጎ*ልዮ Theory

Ethiopic GP Members

- Dessalegn Mequanint Yehuala (Chair), Lecturer and researcher in computer science
- Mulugeta Seyoum (Secretary), Assistant Professor in linguistics
- Gezahegn Tadesse, Director IT projects, Awash International Bank
- Balcha Reba, Director standards at MCIT
- Assefa Kore, IT expert at Ethio-telecom
- Kinfe Michael Yilma Desta. Ph.D. student in Law
- Tigabu Dagne Akal, Lecturer at AIT in computer engineering
- Halefom Hailu Abraha, expert in cyber law, INSA
- Teshome Yehualashet, Assistant Professor in linguistics
- Kassahun Lemlemu, Journalist at MK
- Destaye Alemayehu, M.Sc. student in computer science

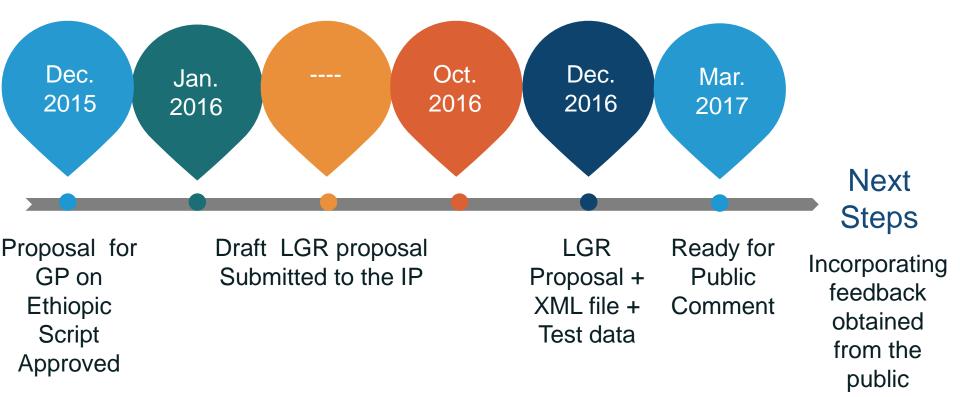


Challenges

- Languages are under-resourced
 - Getting pertinent information both in paper and digital forms for the six languages has been challenging
 - We carried out corpus analysis only for two of the languages
- Complexity in treating variants
 - Amharic language is the official working language of the Federal Democratic Republic of Ethiopia
 - It is also the official or working language of several of the states within the federal system
 - Security versus impacts of variants in the remaining seven languages
 - Striking the balance between security versus impacts of variants
- Deviation from the original timeline
 - We have made submission to the IP 5 times 5 levels of iteration



Progress and Timeline for Remaining Activities



To Summarize

Proposal for the Ethiopic Script LGR for the Root Zone has been submitted 5 times- the last two submissions were complete in the sense that unlike in the first two submissions the GP submitted the XML file and test data along with the LGR proposal- a total of five submissions and five level of iterations (received five feedbacks from the IP).

The LGR proposal for Ethiopic Script is ready to be submitted for public comment.



Update by Greek Generation Panel Panagiotis Papaspiliopoulos Chair, Greek GP

Script for LGR

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

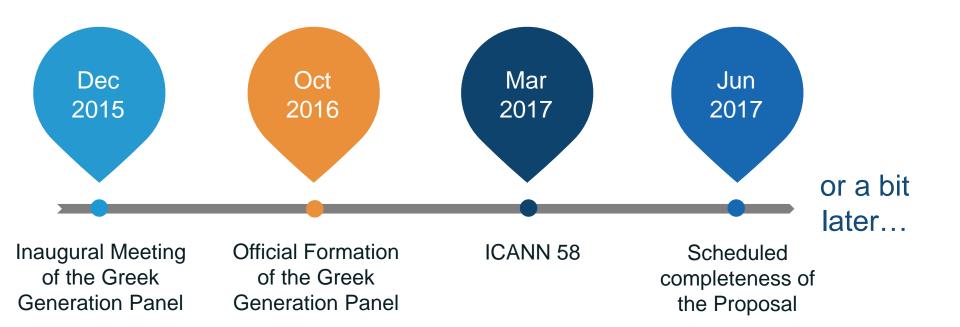


Composition of the Greek Generation Panel

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



Timeline



Moreover...

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



Structure of the Proposal

- Background on script
 - The Greek Language Question
 - The Greek Orthography
- Overall Development Process and Methodology
- Repertoire
- Variants
 - Within Script Variants
 - Cross Script Variants
- Whole Label Evaluation Rules
- Contributors
- References



Useful Key Points

The Greek Language Question

Diglossia: Katharevousa vs. Dimotiki
 (e.g. Πειραιεύς - Πειραιάς)

The Greek Orthography

- Polytonic vs. Monotonic (e.g. the Lord's Prayer)
 - Πάτερ ἡμῶν ὁ ἐν τοῖς οὐρανοῖς, ἁγιασθήτω τὸ ὄνομά σου, ἐλθέτω ἡ βασιλεία σου, γεννηθήτω τὸ θέλημά σου, ὡς ἐν οὐρανῷ καὶ ἐπὶ τῆς γῆς
 - Πάτερ ημών ο εν τοις ουρανοίς, αγιασθήτω το όνομά σου, ελθέτω η βασιλεία σου, γεννηθήτω το θέλημά σου, ως εν ουρανώ και επί της γης



Proposed Characters for Registrations

Only Monotonic characters are to be allowed!

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



Issues of Concern: The Sigma and Final Sigma

The Sigma and Final Sigma

```
σ (small, U+03C3) e.g. ασπίδα (shield) 
Σ (capital, U+03A3) e.g. Σοφία (female name) 
ς (small final, U+03C2) e.g. τέλος (end) 
but... 
IDNA2003: \mathbf{\sigma} <-> \mathbf{\Sigma} <-> \mathbf{\varsigma} (e.g. θησαυρός – treasure) 
θησαυρός \rightarrow ΘΗΣΑΥΡΟΣ \rightarrow θησαυροσ \rightarrow not correct Greek!
```

IDNA2008: **σ** and **ς** are different accepted characters, treated separately – reverse mapping not possible!



Issues of Concern, Continued

- Status: under process
 - 1. Other "Within Script Variants" cases (e.g. vowels with/without diacritics)
 - 2. Cross Script Variants
 - Greek and Latin
 - Greek and Cyrillic
 - Greek and other scripts
 - 3. Whole Label Evaluation Rules



Engage with the Greek Generation Panel



Reach us at:

Email: GreekGP@icann.org

Website:

https://community.icann.org/display/croscomlgrprocedure/

Greek+Script+GP

Thank You and Questions



Panagiotis Papaspiliopoulos





Agenda

- Generation Panel background
- Latin script
- Composition of the Generation Panel
- Work plan



Generation Panel Background

- Summer 2016 GP restarted with new call for volunteers
- Real work started October 2016
- GP draft proposal finalized for IP comment
- Waiting for Integration Panel comments



Latin Script - Subject of the GP Work

- Maximal Starting Repertoire V2 (MSR-2)
- Lowercase letters
- UNICODE ranges
 - Controls and Basic Latin
 - Controls and Latin-1 Supplement
 - Latin Extended-A
 - Latin Extended-B
 - IPA Extensions
 - Combining Diacritical Marks
 - Combining Diacritical Marks Supplement
 - Latin Extended Additional
 - Latin Extended-C



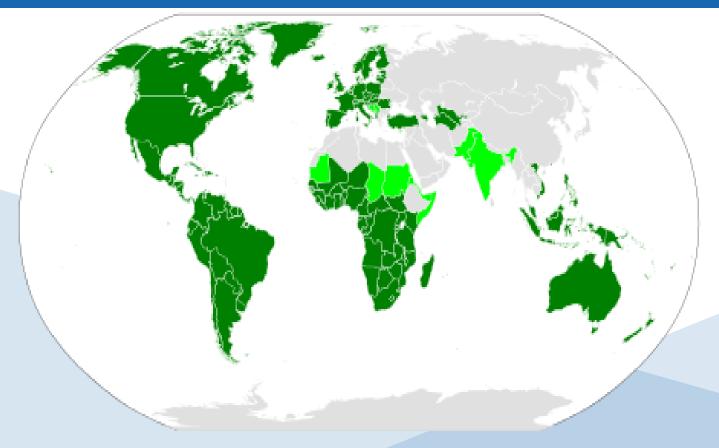
Latin Script - Subject of the GP Work

- Same origin as
 - Cyrillic
 - o Greek
 - Armenian
- Used by missionaries in creating alphabets for languages without scripts
- Non-exhaustive list of 455 languages in scope
- Non-exhaustive list of EGIDS 1-5 languages contains 300 languages
- At a glance, there are 279 MSR-2 code points for attestation

Note: EGIDS stands for the Expanded Graded Intergenerational Disruption Scale. This is a tool that is used to measure the status of a language in terms of endangerment or development.



Prevalence of Latin Script Usage



The dark green areas show the countries where the Latin script is the sole main script.

Light green areas shows countries where Latin co-exists with other scripts.

Grey areas – Latin alphabets are sometimes extensively used in areas colored grey due to the use of unofficial second languages, such as French in Algeria and English in Egypt, and to Latin transliteration of the official script, such as pinyin in China.



Composition of the Generation Panel

- Lack of language expertise
 - Chinese Pinyin
 - Native languages in Oceania and Australia
 - Native languages in North and South America
- Diversity

0	Community Representatives	2
0	Linguistic Expert	2
0	Registry/Registrar Expert	3
0	Policy Expert	3
0	Technical community ,DNS	7
0	IDNA/Unicode Expert	8

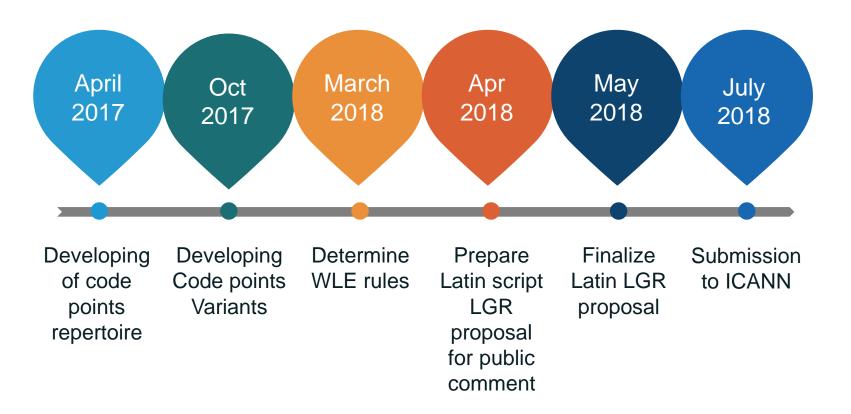


Work Plan

- 1. Developing Code Point Repertoire
- 2. Developing Code Point Variants
- 3. Developing Whole Label Evaluation Rules
- 4. Preparation Latin script LGR for public comment
- 5. Finalization of LGR Documents for Latin Script and submission to ICANN
 - Panel will hold fortnightly the conference call with agenda and materials ready for discussion
 - It is expected to have few face-to-face meetings, probably during Phase 2 and 3
 - All working materials will be shared online



Timeline



Note: This timeline will be updated in a timely manner according to the increasing/decreasing number of action items, and the coordination situation

To Summarize: 15 months is estimated time for project finalization



Thank You and Questions

Website:

https://community.icann.org/display/croscomlgrprocedure/ Latin+GP

Email: Mirjana.Tasic@rnids.rs



Engage with ICANN and IDN Program



Thank You and Questions

Reach us at: IDNProgram@icann.org

Website: icann.org/idn



twitter.com/icann



gplus.to/icann



facebook.com/icannorg



weibo.com/ICANNorg



linkedin.com/company/icann



flickr.com/photos/icann



youtube.com/user/icannnews



slideshare.net/icannpresentations

