

The image features a promotional graphic for the ICANN 60 Annual General meeting in Abu Dhabi. The graphic is overlaid on a photograph of a city at dusk. On the left, several tall, modern skyscrapers with glass facades are illuminated. In the center, a large, ornate fountain structure is visible. On the right, a large, domed building with a prominent archway is lit up. The foreground is filled with vibrant flower beds in shades of red, pink, and white, with a paved walkway leading through them. The sky is a deep orange and red, suggesting sunset or sunrise. The text is presented in a clean, white, sans-serif font within a white-bordered box.

ICANN
ANNUAL GENERAL

60

ABU DHABI

28 October–3 November 2017

Cyrillic Generation Panel Update

Душан Стојичевић
Dusan Stojicevic
Chair, Cyrillic GP



ICANN 60
30 October 2017

Agenda – Proposal for Cyrillic Script RZ-LGR

1

Introduction
and
Background

2

Methodology

3

Results

4

Issues in
MSR-2

5

Next steps
Discussion?
Gratitude

Introduction – Script of LGR

- ⦿ ISO 15924 Code: Cyril
- ⦿ ISO 15924 English Name: Cyrillic
- ⦿ Latin transliteration of native script name: Cyrillic
- ⦿ Maximal Starting Repertoire (MSR) version: MSR-2

Background on Script and Principle 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**

Background on Script and Principle Languages

- ⦿ Derived from Uncial script, augmented by letters from older Glagolitic alphabet, including some ligatures. Additional letters used for Old Church Slavonic sounds not found in Greek
- ⦿ Named in honor of Byzantine brothers, Saints Cyril and Methodius, who created the Glagolitic alphabet
- ⦿ Believed to be developed by disciples of Cyril and Methodius
- ⦿ Individual languages and groups using Cyrillic script
 - Indo-European Caucasian
 - Sino-Tibetan Chukchi and Kamchatka
 - Mongolian Tungus
 - Turkic Ural
 - Individual - Aleutian, Nivkhs, Ket, Eskimos, Yukaghir languages

Background on Script and Principle Languages

- ◉ South Eastern 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 November 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

Development Process – Inclusion Principle

- ⦿ Any code point which is a letter in established contemporary use in a language

Development Process – Exclusion Principles

1. Any code point **DISALLOWED** by IDNA 2008 protocol
2. Any code point representing a security or stability issue, which cannot be resolved at any other stage of the analysis (i.e., stage of determining code points, variants or whole label rules)
3. Any code point not listed in the MSR or listed in the MSR and deprecated or not recommended for use in Unicode Standard
4. Any code point representing technical signs only or that does not meet the inclusion criterion

Development Process – Exclusion Principles

5. The generation panel lacked sufficient information on the usage
6. The generation panel could only ascertain the use for such languages that had an EGIDS rating higher than five (6 or above), as per the “Guidelines for Developing Script-Specific Label Generation Rules for Integration into the Root Zone LGR”
7. The generation panel had data on the use of code points, but where Integration Panel explicitly expressed disagreement on the validity and relevance of such data in separate communications

Code Point Repertoire

86 code points recommended for inclusion

7 code points recommended for exclusion (shown in the table)

#	Unico de CP	Glyph	Unico de Name	Lang. using CP	EGIDS value	Ref.
1	04EB	ӧ	CYRILLIC SMALL LETTER BARRED O WITH DIAERESIS	Khanty	Khanty 6b	Rule 6 http://www.omniglot.com/writing/khanty.htm
2	04ED	ӓ	CYRILLIC SMALL LETTER E WITH DIAERESIS	Sami	Sami 8b	Rule 6
3	04DB	ӓ	CYRILLIC SMALL LETTER SCHWA WITH DIAERESIS	Khanty	Khanty 6b	Rule 6 http://www.omniglot.com/writing/khanty.htm
4	04C2	ӕ	CYRILLIC SMALL LETTER ZHE WITH BREVE	Gagauz	Gagauz 5	Rule 5 http://www.omniglot.com/writing/gagauz.htm Gagauz alphabet not in Cyrillic from 1996
5	04CC	ӕ	CYRILLIC SMALL LETTER KHAKASSIAN CHE	Khakas	Khakas 5	Rule 5 http://www.omniglot.com/writing/khakas.htm
6	045D	ӗ	CYRILLIC SMALL LETTER I WITH GRAVE	Historical sign		Rule 6
7	0450	Ӑ	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”)
- ⦿ Cyrillic GP found cross-script variants with
 - Armenian
 - Greek
 - 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 two (2) homoglyphic variant
- ⦿ Other one (1) not identical, so included in confusables table

Armenian glyph	Armenian code point	Name	Cyrillic glyph	Cyrillic code point
հ	0570	ARMENIAN SMALL LETTER HO	h	04BB
օ	0585	ARMENIAN SMALL LETTER OH	o	043E

Cross-Script Variants - with Greek Script

- Cyrillic has three (3) homoglyphic variants with Greek script

Greek glyph	Greek code point	Name	Cyrillic glyph	Cyrillic code point
κ	03BA	GREEK SMALL LETTER KAPPA	к	043A
ο	03BF	GREEK SMALL LETTER OMICRON	о	043E
φ	03C6	GREEK SMALL LETTER PHI	φ	0444

Cross-Script Variants - with Latin Script

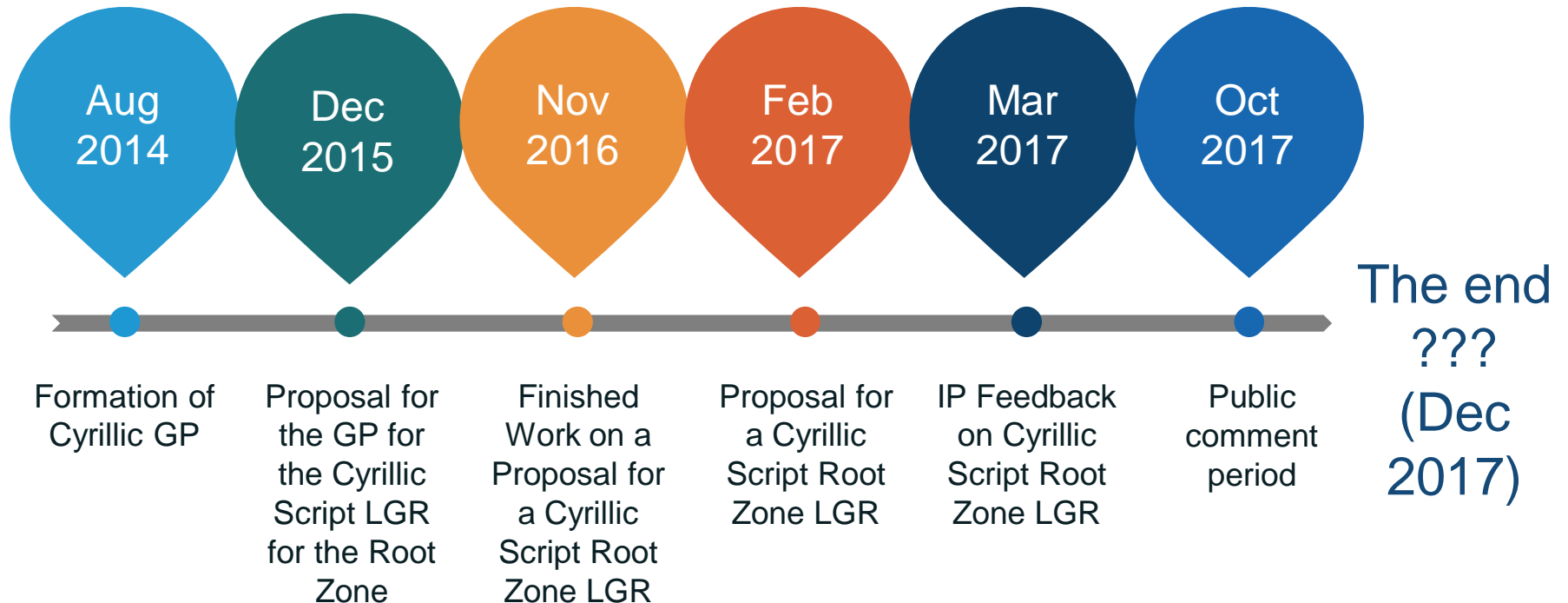
- Cyrillic has following homoglyphic variants with Latin from MSR-2
 - Confusables listed separately

Latin glyph	Latin code point	Cyrillic glyph	Cyrillic code point
a	0061	а	0430
c	0063	с	0441
e	0065	е	0435
o	006F	о	043E
h	0068	һ	04BB
i	0069	і	0456
j	006A	ј	0458
l	006C	л	04CF
p	0070	р	0440
s	0073	ѕ	0455
x	0078	х	0445
y	0079	у	0443
ä	00E4	ӓ	04D3
ë	00EB	ӛ	0451
æ	00E6	ӕ	04D5
ñ	0127	ӧ	045B
ə	01DD	ӕ	04D9
е	0259		

Discussion of Issues in MSR-2

- ⊙ Ukrainian and Belarusian languages have “apostrophe” (U+02BC) as a letter – not punctuation sign
 - *IP response*
- ⊙ Montenegro has two new chars in national scripts (Latin and Cyrillic); ccTLD to implement them at second level
 - *Not yet in Unicode*
- ⊙ Old Church Cyrillic script
 - *Not in second level use, it should not be used for root zone*
- ⊙ Using upper case and lower case Unicode code points in Cyrillic as problem of the confusion during visualization
 - *Need to address that*

Timeline



To Summarize

It took Cyrillic GP more than three years to finalize proposal. But, the work has been done according to the dates defined in original Working plan, with some small delays.

Next Steps

1. **Very short term**
Celebrate actual end of the work with one dinner together

2. **After public comment phase**
Finalize the LGR proposal to include community feedback

3. **Long term phase**
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)
Enkhbold Gombo (Mongolia)	Tattu Mambetalieva (Kyrgyzstan)
Iliya Bazlyankov (Bulgaria)	Yashar Hajiyev (Azerbaijan)
Kadamjon Safiev (Tajikistan)	Yuliya Morenets (Russia)
Mirjana Tasić (Serbia)	Yuriy Kargapolov (Ukraine)
Nazgul Kurmanalieva (Kyrgyzstan)	Yuriy Honcharuk (Ukraine)