

I C A N N

COMMUNITY FORUM

58

COPENHAGEN

11-16 March 2017





Latin Generation Panel

Mirjana Tasić | ICANN 58 | 14 March 2017

Agenda

- ⦿ Generation Panel background
- ⦿ Latin script
- ⦿ Prevalence of Latin script usage
- ⦿ Composition of the Generation Panel
- ⦿ Work plan
- ⦿ Some preparatory materials finalized

Generation Panel Background

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

Latin Script - Subject of the GP Work

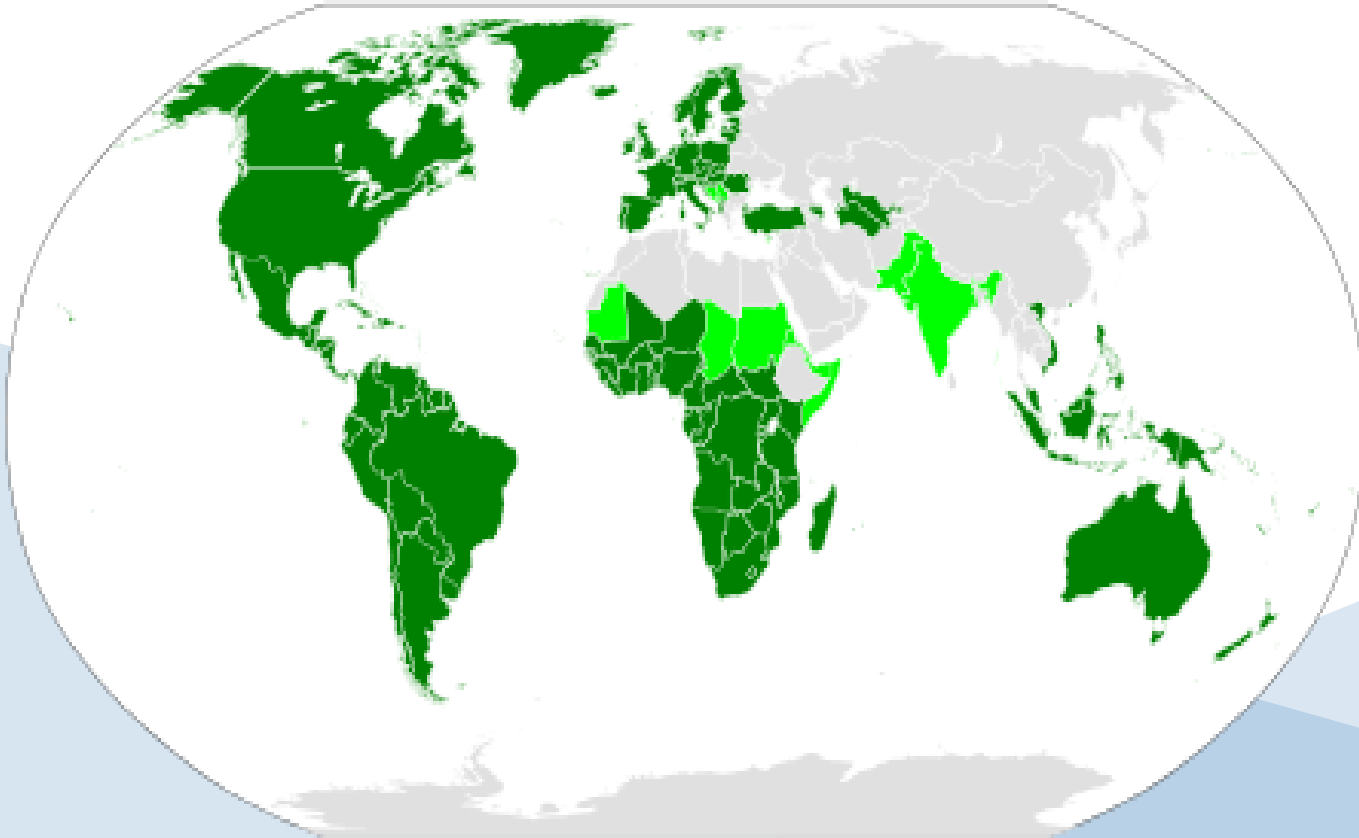
- ⊙ Maximal Starting repertoire V2 (MSR-2)
- ⊙ Lowercase letters
- ⊙ UNICODE ranges
 - Controls and Basic Latin
 - Controls and Latin-1 Supplement
 - Latin Extended-A
 - Latin Extended-B
 - IPA Extensions
 - Combining Diacritical Marks
 - Combining Diacritical Marks Supplement
 - Latin Extended Additional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

Latin Script - Subject of the GP Work

- ◉ Same origin as
 - Cyrillic
 - 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 shows countries where Latin co-exists with other scripts.

Grey areas - Latin-script 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

⦿ 11 members, 4 observers

⦿ Language expertise

○ Arabic	2
○ African languages	1
○ English	10
○ Finish	1
○ French	3
○ German	4
○ Italian	1
○ Japanese	1
○ Middle Eastern Languages	1
○ Spanish	1
○ Swedish	1
○ Serbian, Croatian, Bosnian and Montenegrin	1
○ Urdu	1
○ Yoruba	1

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
 - Community Representatives 2
 - Linguistic Expert 2
 - Registry/Registrar Expert 3
 - Policy Expert 3
 - Technical community, DNS 7
 - 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.

1. Developing Code Point Repertoire

- Determine criteria for including code points
- Analyze each code point in Latin script in MSR-2 to determine its inclusion
- Analyze languages using Latin script to attest whether all letters are included in MSR-2
- Finalize included and excluded code points
- Release for public comment by Latin GP

2. Developing Code Point Variants

- Determine criteria for within-script variants
- Determine criteria for cross-script variants, including identification of relevant scripts
- Analyze each code point to determine within-script variant sets
- Analyze each code point to determine cross-script variant sets
- Review the impact of variant analysis on current delegations of TLDs
- Finalize variant sets, reviewing the symmetry, transitivity, security and stability of the system
- Coordinate with other relevant GPs for finalizing cross-script variants
- Release for public comment by Latin GP

3. Developing Whole Label Evaluation Rules

- Determine relevant WLE rules

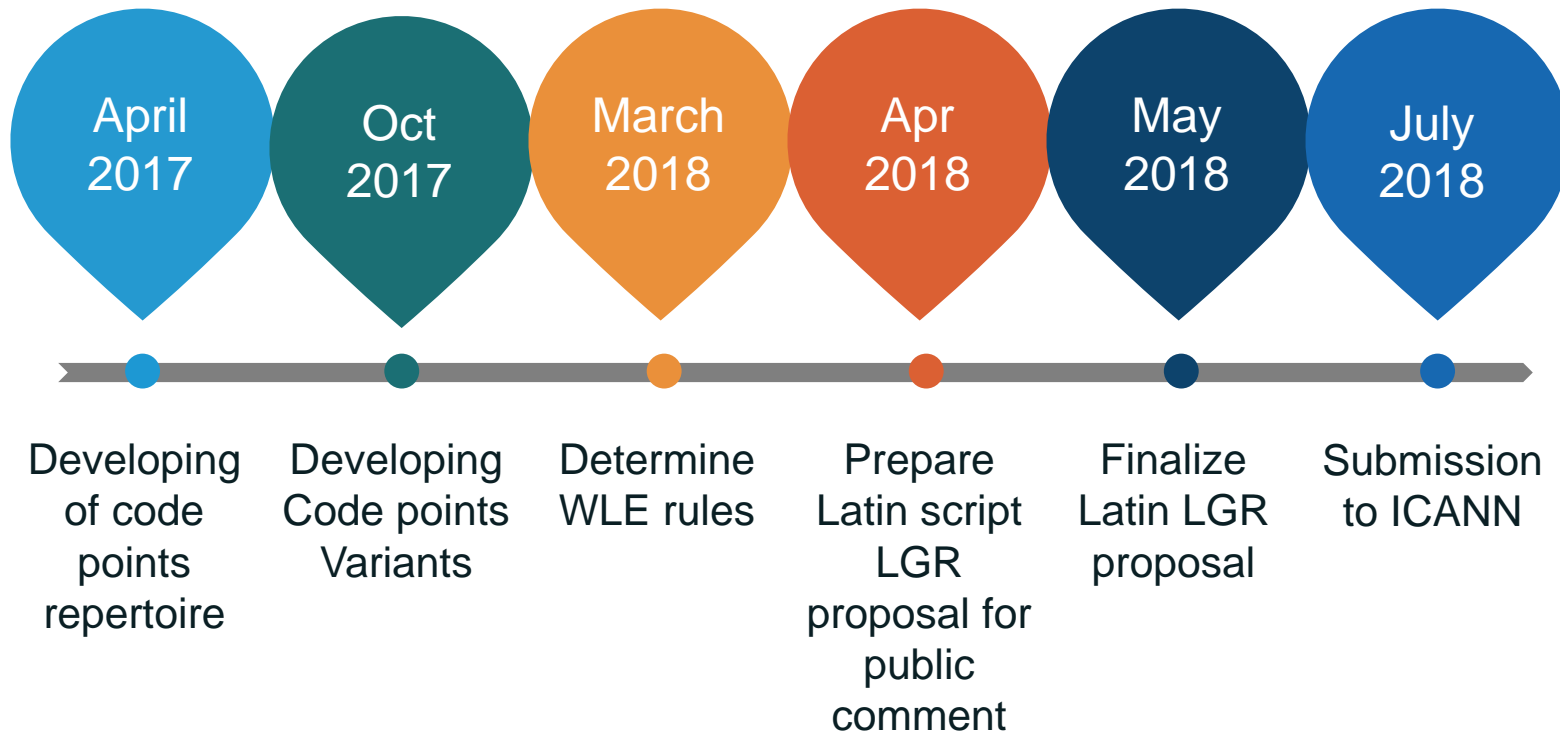
4. Prepare Latin script LGR proposal for public comment

- Finalize the documentation for Latin script LGR
- Finalize the XML formulation of the proposal
- Collect labels to test the proposed LGR

5. Finalize Latin LGR proposal

- Release for formal public comment by ICANN
- Finalize Latin LGR proposal based on public comment

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

Appendix A: World Languages Using Latin Script

Language	ISO 639-3	Classification	Pop.	Lang, status	Language map	Comment
Abenaki Eastern Abnaki	aaq	Algic, Algonquian, Eastern Algonquian, Abenaki	90	9	USA Maine	no known L1 speakers
Acehnese , Achehnese Achinese Aceh	ace	Austronesian, Malayo-Polynesian, Malayo-Chamic, Chamic, Acehnese	3,500,032	5	Indonesia, Sulawesi, Indonesia, Sumatra	
Acholi , Acoli Acooli Akoli Atscholi Dok Acoli Gang Lëbacoli Log Acoli Lwo Lwoo Shuli	ach	Nilo-Saharan, Eastern Sudanic, Nilotic, Western, Luo, Southern, Luo-Acholi, Alur-Acholi, Lango-Acholi	1,197,000,	5	South Sudan Uganda	
Achuar-Shiwiar , Achual Achuale Achuar Achuara Jivaro Maina	acu	Jivaroan, Jívaro	7,000	5	Ecuador Peru	
Adzera , Atzera , Azera , Atsera or Acira ,	adz	Austronesian Malayo-Polynesian Central-Eastern Malayo-Polynesian Eastern Malayo-Polynesian Oceanic Western Oceanic North New Guinea Huon Gulf Markham Upper	30,000	4	Morobe Province of Papua New Guinea	

Table contains, at the moment, information on 455 Languages using Latin script. It is developed using data from <http://www.omniglot.com/writing/langalph.htm> and www.ethnologue.com.

Appendix B: Latin MSR-2 Code Points with Attestation

Status CD	Status MM	Unicode	Glyph	Example languages	EGIDS	ISO 639-3	Population	Ref	Unicode	Comments
	OK	00DF	ß	German	1 German	deu	78,093,980	CLDR	LATIN SMALL LETTER SHARP S	May only appear in the middle or at the end of a label. IDNA version issues.
	OK	00E0	à	Portuguese, French	1 Portuguese	por	203,352,100	CLDR	LATIN SMALL LETTER A WITH GRAVE	
	OK	00E1	á	Spanish	1 Spanish	spa	398,931,840	CLDR	LATIN SMALL LETTER A WITH ACUTE	
	OK	00E2	â	Portuguese, French, Turkish	1 Portuguese	por	203,352,100	CLDR	LATIN SMALL LETTER A WITH CIRCUMFLEX	More information about circumflex use in Turkish in the Distinctive Features section.
	OK	00E3	ã	Portuguese	1 Portuguese	por	203,352,100	CLDR	LATIN SMALL LETTER A WITH TILDE	

Issues for Discussion

- ⦿ Amount of work – do we need extra help?
- ⦿ Lack of expertise in some language areas – how to solve?
- ⦿ Pinyin – should we consider it?
- ⦿ Upper and lower case similarity – should we consider both cases?
- ⦿ Proposal for next face-to-face meeting - Belgrade, Serbia
- ⦿ ...

Engage with ICANN



Thank you and questions

Website: <https://community.icann.org/display/croscomlgrprocedure/Latin+GP>
email: Mirjana.Tasic@rnids.rs



twitter.com/icann



[gplus.to/icann](https://plus.google.com/icann)



facebook.com/icannorg



weibo.com/ICANNorg



linkedin.com/company/icann



flickr.com/photos/icann



youtube.com/user/icannnews



slideshare.net/icannpresentations