
ABU DHABI – ICANN GDD: IDN Cyrillic Generation Panel Meeting
Monday, October 30, 2017 – 15:15 to 16:45 GST
ICANN60 | Abu Dhabi, United Arab Emirates

UNIDENTIFIED MALE: This is the ICANN 60 ICANN GDD IDN Cyrillic Generation Panel Meeting on the 30 October 2017, from 3:15 to 4:45 in Capital Suite 7.

DUSAN STOJICEVIC: Shall we start? Okay. It's okay, it's on. Hello to all people in the room and in the virtual room. I didn't see how many we have on remotely. I'm Dusan Stojicevic. I'm coming from Serbia, and I'm acting as Chair of Cyrillic Generation Panel. I'm here to present the final result of the work of Cyrillic Generation Panel. I must say that before we start, I need to express the deepest congratulations to all my colleagues who were participating in the work, and they are the true force for getting this result that I will present in next – I don't know how much – 70 minutes.

So I will speak in my presentation about our work, and I will separate the presentation in five different parts. One is introduction and background, the second one is methodology of the work of Cyrillic Generation Panel, number three is about the results of our work, then number four is about issues in MSR-2,

Note: The following is the output resulting from transcribing an audio file into a word/text document. Although the transcription is largely accurate, in some cases may be incomplete or inaccurate due to inaudible passages and grammatical corrections. It is posted as an aid to the original audio file, but should not be treated as an authoritative record.

and as a conclusion, number five is about next steps, the end, open discussion about work.

So what we are talking here about, we are talking about Cyrillic script. You can see on the first slide all codes and names and what is necessary to know about the script. You can see also what version of MSR we were using. So about something that is – and still is – the matter of big discussion among people who are willing to talk about history, historians, etc., but what is the common knowledge about Cyrillic script? It's based on early Cyrillic coming from the Bulgarian Empire in the 9th century. It was used for languages across – and still we are using in – eastern Europe, also north and central Asia. So mostly former countries which were part of Soviet Union, or part of soviet bloc, let's say it like this. I can tell you that even now, mostly Slavic people are using the script, and [as I said] non-Slavic languages were influenced by old Soviet Union or by Russia.

As a background, we can say that most nearly 250 million people are using this as official script, although this is not exact figure because recently for example Kazakhstan was turning from Cyrillic script to Latin script. So even these days, we have a lot of turnouts in countries because of political decisions made by new governments, etc. And about half of them are coming from Russia, definitely, so our scope of the work was dedicated to find all the languages and all the scripts in Russia.

And this last bullet is about trivia. Bulgaria was entering recently – 10 years ago – European Union, and by that fact, Cyrillic script became an official script of the European Union in addition to Latin and Greek.

More about background, here I was mentioning Glagolitic, or Glagoljica which we called all across Slavic countries. So it's something about history. As I said, historians are still fighting about all facts about the history of Cyrillic script, so this is something that I was trying to extract without any, let's say, dispute with historians. So those facts are solid and proven by historians.

At the end of the slide, you have the list of individual languages and groups using Cyrillic script. Here you can see on this map which countries are using Cyrillic script, and I highlighted with red color Kazakhstan which I was mentioning before that they have turned recently into Latin script. So you can see how Cyrillic script is spread across eastern Europe and north and central part of Asia.

Now we are going to the methodology according to workplan in the proposal for Cyrillic script. With all due respect, we were doing our job according to the plan, not perfectly regarding timeline because we had a lot of problems about what I've just explained about history, etc. So when we skipped those

problems, we were able to finish our task, but we were doing everything according to the workplan which was in the proposal.

So what we did, we initially compiled the language-based repertoire based on the second level, and we were using .su which is Soviet Union, which is a ccTLD. And in the tables we have found inventory for languages currently spoken in Russia. So we were doing a face-to-face meeting in Istanbul, and this was a very productive meeting with core members of Cyrillic Generation Panel, and 80% of the task we managed to do in Istanbul.

The other 20%, we were using e-mail list to share and finalize the documents. And also as you know and as we discussed previously in Copenhagen on ICANN meeting, we worked with consulting integration panel on a crucial query that we had during our work. Some of the questions that we sent to integration panel was including very important query for Ukrainian people and Belarus people. This is about so-called apostrophe.

Our methodology was very simple regarding the inclusion and exclusion principles. We had only one inclusion principle, and this is on the screen. So any code point which is a letter in established contemporary use in a language. This was only one inclusion principle.

But we had seven exclusion principles. Any code point disallowed by IDNA 2008 protocol, this was the first. Second, any code point representing a security or stability issue, which cannot be resolved at any other stage of the analysis. Any code point not listed in the MSR or listed in the MSR and deprecated or not recommended for use in Unicode Standard. Fourth is any code point representing technical signs only or that does not meet the inclusion criteria. Fifth, the Generation Panel lacked sufficient information on the usage.

Generation Panel could only use for such languages that had an EGIDS rating level higher than five – six or above – as per the guidelines for developing script, etc. 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 communication. So those are seven exclusion principles.

Now we are talking about results. After all that we've done, you have on the screen seven code points recommended for exclusion, and we were desperately trying to short this list to make it the most short that we could, and 86 code points recommended for inclusion. The results about variants, we didn't find variants in the Cyrillic script, but some code points are visually confusable. So we didn't consider that as a variant,

but we provided a table of confusable code points so organizations and other communities can use as needed.

Cross-script variants, we decided to limit. This was the first decision on variants. We decided to limit these to homoglyphs. So included code points which are homoglyphs in the lower case but not homoglyphs in the upper case. Only lower case because upper case is disallowed in IDNA 2008, and decision was made in consultation with the Integration Panel. So Cyrillic Generation Panel found cross-script variants with Armenian, Greek and Latin script. We didn't find any cross-script variants with Georgian script.

I will go with all cross-script variants that we have found. Armenian GP was indicating three variants with Cyrillic script because they finished the work previously so we could use their work. Opinion of Cyrillic Generation Panel was that we have only two homoglyphic variant, and other one is not identical, so it's included in our confusable tables.

With Greek alphabet, we have found three homoglyphic variants. You can see them on the screen. And because the history of Cyrillic is very confusing for Latin people, and we were expecting a lot of cross-script variants and we were right. You have the whole list of variants with Latin script. We have also – sorry, I didn't see you.

BILL JOURIS: Hi. Bill Jouris, I'm on the Variants Working Group for the Latin Generation Panel, and we're nowhere near as far along as you are, but we have identified a fairly substantial number of variants between Cyrillic and Latin so far, and we're about halfway through. And I wanted to share those with you.

DUSAN STOJICEVIC: Thank you. I'm seeing this for the first time.

BILL JOURIS: Yes, I apologize for that.

DUSAN STOJICEVIC: So I cannot be more precise than we have a lot of – I see some signs, and I recall that we didn't find them as homoglyphic variant, so they are not exactly the same. So we put some of these characters in the table of confusable characters.

BILL JOURIS: Okay.

DUSAN STOJICEVIC: So we can discuss even further. We can see the whole document and you can see the whole confusable tables, so we can

compare each of your findings with our tables and we can discuss separate.

BILL JOURIS: Certainly.

DUSAN STOJICEVIC: Thank you. I will follow the work of Latin Generation Panel also. Next, this was mostly about the results of the work of Cyrillic Generation Panel, and also we have opened the public call for comments. So if you have anything on our work to add, you can add through this public comment period.

We have found some issues in MSR-2, and we were discussing those issues in meetings and off the meetings, and we have decided to keep these questions open. First issue is as I said previously, apostrophe from Ukrainian and Belarus language scripts. They're using it as a sign, not as a punctuation. So we got response, but we still want to find a proper solution for this code point, and we want to find the solution to include it in the next version of MSR.

Others are much lower priority, for example Montenegro has two new characters and as I said for – yes?

UNIDENTIFIED MALE: I'm also part of Unicode Consortium, so I can answer that one. Those would never be added as Unicode characters. They can be presented as sequences, and sequences can be added in the LGR. So they're perfectly in fact representable with the current repertoire in MSR-2 because they will never be added as precomposed characters. Because of [normalization] stability, we can't do that. So number two is not an issue in fact, we can do it today.

DUSAN STOJICEVIC: Mirijana?

MIRIJANA TASIC: Sorry for [interrupting]. My last information are that they gave up those two characters. But you see, this is a political question. So nobody knows how....

DUSAN STOJICEVIC: You interrupted me. I was trying to put a parallel between Kazakhstan and Montenegro because Montenegro started a process of – I don't know – to put those two characters in Unicode, but somehow political climate in that country was changed and they erased those two characters from mostly governmental, official websites.

MIRIJANA TASIC: [The official, yes.]

DUSAN STOJICEVIC: So we don't know the real nature about those two characters, and as I said, it's constantly changing, the question. And it's not high-level issues in MSR, all I say. Or Cyrillic script also. It should not be used for root zone because there is no, let's say, practical use. But some of the churches still use old script Cyrillic. Then using uppercase and lowercase Unicode code points in Cyrillic is a problem. Well, we need to address that. This is something that is not MSR question, this is just a question whether we should or should not address the usage of uppercase characters. So those were the issues that we have found.

I was talking about timeline previously. So you see that three years ago, I think it was L.A. meeting, there was a formation of Cyrillic Generation Panel, and you can see that we actually started to work from December 2015. So one year and a half we were losing in some not asked work of Generation Panel, so everything else is according to the plan.

And as I said – and I want to highlight again – public comment period was started in October. And I put question marks, is it the end of the work of Cyrillic Generation Panel [as does we] finish

our work? We will see with public comment period what will bring this, what comments do we get from the community. And also as I said, political situations in countries are constantly changing some issues, and some of them are script, so probably those changes would require our further work. But we will do occasionally this work, that's how I see the future.

So as I said, next steps, very short term we were doing that, so we had last night the dinner when we celebrate actual end of the work. And as I said, long-term phase I explained, so that's mostly about the plans of next steps.

And at the end, I want to highlight who was working, who am I representing, and also to highlight that in bold, you have the core team which was on the meeting in Istanbul, and I'd like to thank to all of these people here, and especially those who are highlighted in bold. So that's all from me.

UNIDENTIFIED MALE:

I have a question about the exclusion principle. Could you take that slide? The next one. Yes, in number six there, you say EGIDS rating higher than five, and then you say six and above. One is the highest, isn't it? And two is the second highest and three is... it goes that way. So which?

DUSAN STOJICEVIC: [There's a number higher.]

UNIDENTIFIED MALE: Six or above.

DUSAN STOJICEVIC: It's higher than five.

UNIDENTIFIED MALE: Yes, but if you look at the scale, it starts with one, so one is the highest, the most dominating language.

DUSAN STOJICEVIC: Okay. Here is the question how you look at this list. Yes, one, they are live and the list is starting with one. But we....

UNIDENTIFIED MALE: Which ones did you include?

DUSAN STOJICEVIC: No, it's a matter of term. Term is we were including higher – okay, [Sarmad.]

[SARMAD HUSSAIN:] Thank you, [inaudible]. Thank you, so I think [to interpret these are] exclusion principles, not inclusion principles. So any languages which had an EGIDS scale six and higher were excluded. Anything which was smaller were included.

DUSAN STOJICEVIC: I was thinking that you're referring to the numbers. Yes, higher than five. So it's six, seven, that's how we meant to exclude.

BILL JOURIS: Perhaps the English term you're looking for is not higher rating, but higher ranking, so that you're excluding those that are a lower rank than five.

DUSAN STOJICEVIC: This was written regarding the numbers, not regarding the ranks because the higher rank is one.

BILL JOURIS: I'm just saying that if you phrase it ranking, it will be clearer.

DUSAN STOJICEVIC: Yes. Thank you. Any other question?

[SARMAD HUSSAIN:] I think we have Latin Generation Members here, so we would like to request Latin GP members to look at the proposal by the Cyrillic GP which is open for public comments, and please share your comments through that public comment process as well. I just saw Bill provided a list to Dusan, but I think it would be useful if the same list can be I think formally submitted to the Cyrillic GP through that process as well. Thank you.

BILL JOURIS: As I mentioned, we're still working on it. it will be another three or four weeks before we have finished going through the Latin repertoire looking for conflicts. But yes, we will. [Sure].

UNIDENTIFIED MALE: [inaudible] Even if you're not done, you could send – I would highly suggest that you send the current list that you have and do a cover letter saying, "We're still working on it," so at least there's the input and knowledge that there's a variant set that needs to be managed.

BILL JOURIS: We can do that.

UNIDENTIFIED MALE: I was just going to comment a bit about the Armenian confusable because as you pointed to, you added two on the list and there was a third one that was in fact suggested by [inaudible] suggested by us, by the way because [it's really] part of the Armenian LGR which has been already accepted. So to some degree as an IP we have no choice but to take into account the third one when we do the integration.

So that also shows how important it is for all the groups to work together because it's very difficult for an integration panel once LGR has been basically finalized to not add one that has been asked by one of the panels of the LGR. So we're getting situation now that you may disagree with the Armenian variant for those two characters, but in fact we have no choice to add it as per integration at this point.

It's not that bad. I look at it, it's in fact pretty reasonable assumption that they're a variant of each other. You can have various opinion if you want, I understand that, but at this point I am [also] bound by the fact that the Armenian LGR is final.

DUSAN STOJICEVIC: The way I see this problem, it's your problem now, so it's Integration Panel who will decide what to do. They have two different opinions, so it's your job, not ours.

UNIDENTIFIED MALE: I understand. [inaudible]

DUSAN STOJICEVIC: We were trying to express our opinion. Our opinion is there. You can just say, “Okay, you're wrong, you're right.” It's your job.

UNIDENTIFIED MALE: It's not our opinion, it's basically what every LGR does, so we are bound by whatever previous LGR did. So that also shows how important it is for the Latin and Cyrillic panel to work together because we would like to avoid having different opinion between two LGRs that we need to be integrated on with these kinds of things.

DUSAN STOJICEVIC: Thank you. Any other remark, question?

UNIDENTIFIED MALE: It's a minor one. On your MSR-2 slide, there's really nothing in that MSR-2 slide you have that we as IP were going to do. There is no action on us, I think, that we have to do anything about it. If you go back, you had the MSR-2 slide.

UNIDENTIFIED MALE: [inaudible]

UNIDENTIFIED MALE: Yes. 02BC is not even – it's beyond us. It's been pretty clear from AIB that they [don't want] it, so while you can consider having it on second level, on the root is absolutely a no-no to have such a character in any LGR in the root level. Second point, it's not a Unicode issue, so it's a [no op] for us. There's nothing to do on that one. On the rest, yes, I think – basically, I don't see any need for us to change anything in MSR to address any of the points I see on that slide.

DUSAN STOJICEVIC: As I said – and I'm here to represent also Ukrainian and Belarus people, so from their point of view which I was expressing pretty much clear, they are using that in their language as a proper character. So their wish is to somehow include in MSR-2. It doesn't mean that you need to do that, this is also your decision. It's my task to represent them and to tell you their wishes. From my personal point of view, by knowing Latin and using Latin and Cyrillic script equally in my language, I can only say that I agree with you that you don't have anything to do with apostrophe because apostrophe in root zone will create a mess. This is my user perspective.

UNIDENTIFIED MALE: Yes, it's way worse an issue in Latin script, by the way. So 02BC is in fact essential to some writing system, much more than like if you go into Polynesian languages, it is in fact part of the writing system. But that [inaudible] this is the root. We're not talking general writing system here, we're talking what you put on the root level as identifier. On the limitation on 02BC is well understood is like I say, way worse in fact in some other writing systems [than] Latin, and there's nothing we can do about it.

DUSAN STOJICEVIC: Any more comments? So I can conclude the session. It was pretty much shorter than I thought. So thank you.

UNIDENTIFIED FEMALE: [inaudible]

DUSAN STOJICEVIC: This is another message for the members of Cyrillic Generation Panel that you didn't find any more problems so far.

UNIDENTIFIED FEMALE: [inaudible]

DUSAN STOJICEVIC: Okay. Thank you. And see you.

UNIDENTIFIED MALE: Thank you. We'll close the session.

DUSAN STOJICEVIC: To type TLD or domain name as such, so we were just highlighting that this is different. Different tables, different variants, different confusables, and we should address this topic. We didn't because it was not asked. It was not our task to address this uppercase.

BILL JOURIS: If you come to the Latin Generation Panel meeting later this week, we have some comments on that very topic as well, which I think is why [Dennis] brought it up.

DUSAN STOJICEVIC: Thank you.

UNIDENTIFIED FEMALE: There is no other – oh, sorry, Dennis is typing. Okay, there are no other questions online. Thank you.

[END OF TRANSCRIPTION]