
COPENHAGEN – IDN Program Update
Wednesday, March 15, 2017 – 15:15 to 16:45 CET
ICANN58 | Copenhagen, Denmark

UNIDENTIFIED FEMALE: March 15th, 2017, Hall B 4.1, starting at 3:15, going to 4:45, and this is the IDN Program Update.

SARMAD HUSSAIN: We'll be starting in three minutes. Could I request speakers to please come here to the head of the table? Thank you.

So, we're going to get started. Welcome to IDN Program Update session, and we have a fairly large agenda which we're going to go through. We have an overview of the IDN program which I will be presenting and share the progress of the work. Then we have a brief update from the Integration Panel which will be presented by Marc Blanchet, followed by a brief update on IDN implementation Guidelines which will be presented by Mats Dufberg.

After these presentations, we will have four community updates from Cyrillic Generation Panel, Ethiopic Generation Panel, Greek Generation Panel, and Latin Generation Panel, which will be presented by the respective Chairs of these panels. Then we'll have hopefully still some time left for questions and answers.

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.

So, let's get to it. Going into a brief introduction, overview and progress of IDN program, IDN program largely focuses on top-level domains, and within that focus, we have one large project which focuses on IDN TLD program, which is currently the programs focusing on defining label generation rules for the root zone which will be used for determining what are valid labels for the root, and the variance of those particular labels.

In addition, we've also been working on developing an LGR toolset. The toolset is focused on using the LGRs which are developed to allow to not only develop those XML-based LGRs but also to use those LGRs to validate labels as well as compare different LGRs and so on. We will talk about the tools and its availability briefly as well.

And then the third aspect of the project is once the LGRs are developed, when the LGR is developed, integrating different scripts, how to use that LGR to implement variance at the top level.

The second part of the focus of the program is on evaluating IDN ccTLD applications through the Fast Track process, and we will also share some progress on that.

Then we also undertake a couple of projects which are focused on second level to assess the community, particularly the gTLDs. They are basically IDN implementation guidelines for the second

level. We have working group which is updating these guidelines, and we will have a separate presentation today where the revised guidelines will be introduced. And we also are now developing some reference second level LGRs, and we'll talk about that.

Finally, Alain with the update on how we are doing [of] our community outreach.

So going into the TLD program, we've actually been working with different communities across the globe and assisting them to develop proposals for how the script should be used in the root zone to develop valid labels and their variants.

The root zone for the first version of the root zone was released almost a year ago, which had Arabic script integrated into it as a single script. Armenian proposal was also received, but it was not integrated at that time because it had shown some cross-script variance with Latin, Cyrillic and Greek scripts, and therefore the Integration Panel decided to wait for those proposals to come in before integrating the Armenian script.

Since then, we've also received proposals from Georgian, Khmer, Lao and Thai scripts. And these slides are a week old, since then we've also received a proposal from Ethiopic script. So, we have five more scripts which are now set for integration into the second version of the root zone, and the second version

of the root zone is anticipated to come out for public comment in the next few weeks.

This is a summary of where different Generation Panels are, beyond Generation Panels which have already finished. We are also seeing very good progress by Chinese, Japanese and Korean Generation Panels, and as you can see, we have activity also by Greek Generation Panel and Latin Generation Panel, and Neo-Brahmi Generation Panel. But we are still looking for volunteers to formulate groups and get work started for Myanmar, Hebrew, Sinhala and Thanaa scripts.

So, if any of you is interested or knows people in those scripts who would be interested, please let us know and we'd love to start those panels. And if you're interested in contributing to one of those panels which are currently underway, I think that's also an area which we are looking for the participation.

Moving on, the LGR toolset has been developed. There is a specification now which is a standards specification to represent IDN tables. It's been standardized to RFC 7940. We also have a tool available based on RFC 7940 which can be used to create LGRs, and also use LGRs to validate labels and determine their variants.

The tool can do more than that. It also can be used to help manage LGRs by comparing new LGRs with older LGRs and

similar functions. It is available online through the link which is available. It currently requires username and password to access it, but those are published in the user guide which is also available online.

And if you would like to install it on your end, the code for this tool has been released with open source license and it can be downloaded from these sites and can be run on local servers.

Moving on, as far as the Fast Track program is concerned, we continue to receive applications for IDN ccTLDs and process those applications. This is a snapshot of where we are, covering all the successfully evaluated IDN ccTLDs from different countries and territories.

So far, we've evaluated 57 different labels from 35 different countries and territories. In addition, from these 57 labels, 48 IDN ccTLDs have already been delegated and they represent 38 countries and territories, and they cover 21 scripts and 35 languages.

IDN ccTLD Fast Track process also goes through an annual review. Last time it was open for annual review was in January 2015, and there were some comments received on the second similarity review process. Based on those comments, the Board resolved that ccNSO look at those comments and revise the

second similarity process in collaboration with the stakeholders, including GAC and SSAC.

ccNSO has recently finished that work and published the report, and it is currently being considered by the Board. And based on the final decision, that public comment will close and the EPSRP process which is the string similarity review will be updated based on the final decision.

I'm going to skip the IDN Implementation Guidelines for now because we will come back to it in a bit. So, moving ahead on to second level reference LGRs, basically IDN tables have to be submitted by the new gTLDs during the pre-delegation testing process, and if they are adding another language in their repertoire, they have to submit an IDN table for which ICANN has to review for security and stability considerations.

So, what ICANN has done is developed and published some reference LGRs for the second level which community can use as, as I said, references so that when they submit an IDN table for testing or review, they have at least some reference point which can be compared with.

It does not mean that these reference tables have to be followed. They're just guidelines, and it is up to the registry itself to decide eventually what exactly they would want to put in the IDN table they're applying for.

So, that's sort of an overview of the different projects we undertake at the IDN program at this time. In addition, we also reach out to the community to tell them what we're doing, and also to engage them in the work which we are doing, because a lot of work which we are doing is actually based on community participation and community leadership.

So, we obviously publish all of our materials through the IDN website, so it is reasonably straightforward to get to with ICANN.org/IDN, and all of our projects and their statuses are posted at that website. We also hold update sessions like this one at ICANN meetings and also go to SOs and ACs during ICANN meetings and present them with update of IDN program as well.

And then we participate at relevant events directly and through our GSE team across the world as well during the year. And obviously, we also maintain our wiki pages and mailing list to keep the community engaged.

So, these are a couple of contact points for us. If you have any questions, queries, comments, please feel free to e-mail us at idnprogram@icann.org.

I'll stop here and pass on to Marc Blanchet who's a member of the Integration Panel, and he's going to give us a quick update on the work by Integration Panel since the last update at ICANN 57.

MARC BLANCHET:

Thank you, Sarmad. So, just a short summary of what we're doing. 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.

This morning, Asmus actually presented a lot of the details of what the integration means. The decisions by the Integration Panel are required to be unanimous. There are five members, Asmus Freytang here, Nicholas Ostler, Will Tan, Michel Suignard and myself.

We receive the proposals from the Generation Panel, but we don't wait until we receive the panel the proposals. We actually work with the Generation Panels in advance of this by reviewing their drafts at the various stages to make sure that the necessary information is present when they're going public comments, and therefore help the overall process.

So, this is a list as of a few weeks ago, and I should add additional, which is we've received recently an early document from the Japanese Generation Panel, which is not listed there. Obviously, the scripts Generation Panels have a different level of where they are. For example, the Ethiopic, we've been reviewing

the final drafts and it's roughly almost ready for public comment.

Georgia went to public comment, so now we're integrating it into the root zone LGR. Korean and Chinese, we got multiple interactions and drafts that we've been discussing. Cyrillic is similar, which is almost ready for public comments. Thai and Lao have been to public comments and are now being integrated, and Neo-Brahmi we received an initial proposal for Devanagari.

So, as you could see, we will have a lot on our plate for integration in the next months. Other tasks that the Integration Panel has to do is the maximal starting repertoire. This is kind of the sandbox. If you look at the Unicode code table, then you have IDNA 2008, a set of possible codepoints, and MSR is a smaller set given the restrictions of the root zone.

The current MSR, MSR-2 has been there for some time. We don't have any update yet. It's not finalized because additional scripts will need to be added, but we don't see the pressure at the moment. There were some requests for adding some codepoints to the MSR. For example, we got some discussion with the Cyrillic GP for specific codepoints, but it didn't [hit] the appropriate requirements for adding it to the MSR, and there are two we've been in discussion for adding to codepoints.

So, the MSR is obviously the sandbox. It may not be perfect, but I think we've seen that it's a very good sandbox. But we are open if we are missing codepoints to be discussed. Obviously, the bar should be pretty high, even the early work.

There's a variant rules document that we have updated, and it's also being processed to the IETF since it's a good informational document on variant rules in general, and hopefully become an RFC at some point in time. But that's not a standard track, it's an informational one.

LGR-2 integration, I think I'll skip this one because it's been discussed this morning extensively. So, we are doing the integration for what we call the LGR-2, the second version of LGR with the following scripts, and hopefully soon be for public comment. That's it for me.

SARMAD HUSSAIN:

Thank you, Marc. So, we will now request Mats Dufberg to give us a brief update on IDN implementation guidelines. So, Mats, please.

MATS DUFBERG:

Thanks, Sarmad. Yes, the implementation guidelines are for the next level in the DNS tree, so this is not for the root zone, this is for the second level. That is for the TLDs. Their purpose is to give

good rules and description for the registration of domain names under the TLDs.

For the gTLDs, these guidelines will be bound by the registry agreement. For ccTLDs, these are recommendations to be considered. And the purpose is to minimize the risk of confusion and cybersquatting, etc. So, we don't want the IDN names to be confused by the users, so that is the background.

The status right now is that we have created a final draft for public comment, and we really want all of you in this room and online to take a look at the draft if you're interested in this matter. We have presented interim draft earlier, but now we have something that we want public comment on.

The IDN Guidelines Workgroup represents various stakeholders in ICANN, and you can see the list of all the members of the working group. We have regular meetings, bi-weekly or weekly. The guidelines contain recommendations in six categories: transition – that is for example, going from IDNA 2003 to 2008, format of IDN tables, consistency of IDN [tables] and practices, IDN variants, similarity and confusion, and then terminology.

On a previous ICANN meeting, it has been raised that we should cover registration data and EPP, but we have decided to exclude those topics. But it's still open, the public comment can suggest that we cover those areas, and with more concrete suggestions.

Here, you have the link to the draft guidelines for comments, and you have a link to the wiki page. So please read, and you can e-mail any comments or feedback to the address listed on the page. Thank you.

SARMAD HUSSAIN:

Thank you, Mats. So now we're going to go into the community updates, but before we do that, let me pause here and see if there are any questions around the room or for comments for the material that has been presented so far before we proceed. Please.

EDMON CHUNG:

I'll bring up my favorite project, project seven. Those who don't know what that is, we've covered it very briefly. It is after all these LGRs are created, how we will implement the variants into the root, the processes up for gTLDs and ccTLDs. Is there any more information, Sarmad, you can provide us on that?

SARMAD HUSSAIN:

Generally the status is the same, but it's not that we've not made more progress. Basically, where we are is that some internal homework has now matured reasonably, and we have an internal draft report which is being presented to senior

management and the Board IDN Working Group for consideration and for review.

As soon as that is cleared and feedback from management and whole IDN Working Group has been incorporated, then we're going to go forward to the public comment. So, the report has been done and is now currently under consideration by the senior management and the Board IDN Working Group. That's where we are.

EDMON CHUNG:

The reason why I bring this up, recently there [have been] increasing enquiries from those – especially IDN TLDs that have implemented the Chinese IDN TLDs that are in the root right now and operational, because some of the users are starting to see the confusion, starting to report confusion that the second level they may have variants and top levels and they start failing.

Those are the issues that a long time ago, the VIP, the Variant Issues Project Team study teams have already identified, and they are actually harming our internet users right now. So, it should be a matter of urgency. I understand that the LGR needs to be processed, but the process of actually putting it into the root is probably equally important now in terms of time-wise. And I know you probably won't want to put a date to it, but I guess the question is still, what is likely the date that we can

start seeing this come out and the community can deliberate on it?

SARMAD HUSSAIN:

It would eventually depend on the feedback we receive from the senior management and the Board IDN Working Group members, but if that goes through, we should be able to get a draft to the community in the next few months, probably after the summer ICANN meeting. But soon after that, so a few months.

EDMON CHUNG:

Sorry for hogging the mic and belaboring this issue, because I do think it's very important. Unfortunately, some of my Chinese colleagues couldn't join us here to add to my voice here, but you mentioned that it depends on some of the processes internal to ICANN.

I think I'd like to point you to one of the documents coming from the RYSG as well, noting that I hope the proposal coming out is a strawman and not an ironman so that people cannot make any changes to it. Because if it goes through so much process internally, there may be a lot of ownership in that document.

Hopefully, that's not going to be the case and useful input from the community can actually still shape the eventual process.

SARMAD HUSSAIN: Certainly a strawman. It's nowhere close to what you're suggesting otherwise. Any other questions?

Okay, then let's move on to the second part of this session, which are community updates. I will request the panelists to limit their presentations to about seven minutes. So, the first presentation is by Dusan who is chairing the Cyrillic Generation Panel. Let me hand it over to Dusan Stojicevic. Sorry for mispronouncing that name.

DUSAN STOJICEVIC: That's why you need a Generation Panel on Cyrillic. No, I'm joking, and this is pronunciation: I'm Dusan Stojicevic. I'm chairing Cyrillic Generation Panel, and I will be as short as possible. As I said, I am chairing Cyrillic Generation Panel, so obviously, we are dealing with Cyrillic. This is basic stuff about Cyrillic, what is ISO called, English name, transliteration in Latin and under which MSR we're doing our task.

Background on scripts and principal languages, a little bit about history on Cyrillic script. Let's say first Cyrillic was invented in First Bulgarian Empire in 9th century, and we are using Cyrillic from that time across Eastern Europe, and North and Central Asia. You can find basis in alphabets in all languages past and

present, especially those of Slavic origin, and non-Slavic languages influenced by Russian.

The figure is really impressive. 250 million people are using officially as official script Cyrillic. So, another trivia. With accession of Bulgaria to the European Union, Cyrillic became the third official script of the European Union in addition to Latin and Greek.

How it looks on the map, you can see the area where the script is used, and in light green you can see couple of countries where the script is used equally with other scripts. Dark green is for countries where this is the only official script in constitution. So, you can see the parts of the world on the slide.

Let's go to business: what we did. We did everything according to the plan which was in proposal for Cyrillic script. We initially compiled one list of, one repertoire of codepoints based on second level IDN tables and used by different ccTLDs. And most importantly, we were using .cu – Soviet Union – ccTLD, which contains inventory for languages currently spoken in Russia, because in Russia you have more than 100 languages which are using Cyrillic script.

So, we started the real work on a face-to-face meeting in Istanbul last November. Afterwards, we were continuing to use mailing list to share, shape and finalize the documents. We had

special enquiry – like Marc said – for some of the codepoints, especially for Ukrainian and Belarus apostrophe, which is not included yet in MSR, but we are suggesting.

So, the results. Can we move the picture to the left? Thanks.

The result is 84 codepoints are recommended for inclusion, eight codepoints recommended for exclusion. Those who are for exclusion are in the table on the screen.

This is the main variants. Let's say no variants in Cyrillic script. Some codepoints are visually confusable and we are not considering that as variants, and at the end we provided the table on confusable codepoints so anybody can use it as needed.

Cross-script variants, decision of the group was that we limit this to homoglyphs. We have also decided to also [inaudible]. Sorry. Also with talks with Integration Panel in that we are using only lowercase. We found cross-script variants with three scripts: Armenian, Greek and Latin, and we didn't find any variant with Georgian script.

I will go fast through the variants. This is with Armenian, this is with Greek script, this is with Latin script. And what we are doing next, as it was already said, we are preparing the final document

for public comment. We are shaping this document, and those are the steps from I was pointing to.

Three phases. One is short phase, second is after public comment phase, and long term phase, because in our region we have announcement for new letters inside some of Cyrillic in some countries.

Members of Cyrillic panel all listed here. Thank you. If you have any questions –

SARMAD HUSSAIN:

Thank you. I think there's a comment, but it's not clearly said whether it needs to be read out, so while we clarify this, let's move on to the – in the chat room. So, let's move on to the next presentation. We have a presentation by Dessalegn, who chairs the Ethiopic Generation Panel. Over to you, Dessalegn.

DESSALEGN MEQUANINT:

I will try to highlight the progress that we have made since the Hyderabad meeting. Yes, okay.

So, these are the agenda items that I'm going to focus on. First, I will give you background information Ethiopic script, and then I will say a few words about GP members and the challenges that we encountered while designing the LGR proposal, then a few

words about the summary of the progress and timeline for completing the remaining activities.

Ethiopic script is one of the oldest script. Its use dates back to the 1600s. Currently, it's being used predominantly by the national language of Ethiopia. It's used as the national language for two countries: Ethiopia and Eritrea.

In this label generation ruleset proposal, we have only considered eight languages even if there are many languages being spoken in Ethiopia because of the obvious reasons that most of the languages do not satisfy the requirements set out in the requirements guidelines. Particularly, their status level is not four or less, which is required. [inaudible] is less than four.

There are significant number of user communities living in many countries, so we can say Ethiopic script is usable in many parts of the world.

Just to summarize what Ethiopic script is all about, it's a syllabary script which mixes, of course, consonant and vowel.

Among the eight languages being considered for the LGR proposal, one of the languages which is the Amharic language exhibits peculiar property. Ethiopic script version which is the Amharic language makes use of – has one peculiar characteristic, which is phonemic decay. Most of the codepoints

the Amharic language uses are already descendant from the ancestor language which is [inaudible]

Due to that historical reason, there is what we call codepoint redundancy in the Amharic script, the Amharic writing system. I will come to it momentarily.

These are the panel members. As you can see, the panel members involve a diverse set of people with a diverse set of skills. We have members who are linguists by profession. We have also people from the government organizations, Ministry of Communication, Information Technology, from the telecom sector, post-graduate students.

In the course of developing the LGR proposal, we have encountered a few challenges. One of the challenges was generating enough evidence, as particularly for [under-resourced] languages, particularly Amharic and English. Except these two, remaining six are not well developed in the sense that they are only being used in their local community, particularly for primary educations. So, we have some challenges of generating evidences for these script versions using those six languages.

We also encountered one small challenge in terms of dealing with variants among the eight language. The Amharic language has one property, as I said earlier. There is some codepoint

redundancy because of historical reasons. The script [inaudible] inherited from an ancestral language which we call [Ge'ez.] That specific phenomena caused us to [do this,] the variant things that forced the GP members to treat this variant things in special way.

The obvious reason why variants require some special treatment is because variants bring or they create some observable influence on aspects of the security of the root zone. Due to that, among the 318 codepoints in the repertoire, 90 of them exhibit this property, the variant property.

So for security reasons, we made them blocked variants, meaning non-allocatable variants. Security always is a prime concern in designing LGR proposal. Due to this reason, we took the safest option of making these 98 redundant codepoints non-locatable variants.

One of the strangest criteria the IP want the LGR proposal to meet is making the document readable for the larger audience. Due to this restriction, we were also forced to make some sort of analysis, like what is the impact of making the 98 codepoints non-allocatable. For that, with the support of the IP, we did some sort of analysis.

We generated evidence showing that the impact of variants on the impact of security or the impact of making these 98

codepoints non-allocatable variant is little. In fact, the [inaudible] and Amharic language, the issue of using [inaudible] is much more serious than some of the variants being non-allocatable codepoints. Some of the challenge that I have earlier said has contributed to drag our project timeline to some extent.

This is a summary of what we have done so far and what remains to be done in the immediate future. The Generation Panel was formed in December 2015, and it took us nearly 30 members to reach to this level. We have already submitted a final version for the Integration Panel. We are hoping that it will be published for public comment anytime soon. Thank you.

SARMAD HUSSAIN:

Thank you. So, let's keep moving forward. We are getting short on time. Next presentation is by Panagiotis. I'm not going to attempt to say the last name, I'm sorry. I'd like you to please introduce yourself. He is the Chair of the Greek Generation Panel. And give us a brief overview of the Greek Generation Panel work.

PANAGIOTIS PAPASPILIOPOULOS: Thank you, Sarmad. Hello, everybody. My name is Panagiotis Papaspiliopoulous, but in short, you can call me Panagiotis or just Panas. I'm the Chair of the Greek Generation

Panel. Together here with us, there are also three other members of the Greek Generation Panel. They're sitting in the first row behind me.

So, [these are some] small piece of information for the script. The script is Greek. It is the modern Greek. It's the script that is used in Greece and in Cyprus as official script for the Greek language. And it's also used by large amount of Greeks who don't live in Greece and Cyprus and live in other place all over the planet.

So, the composition of the Greek Generation Panel. The panel consists of individuals who are actively involved in policy development process related to the telecommunication, research and development related to the Greek language, standardization, computing and maintainers of the domain name system in Greece and Cyprus. The panelists come also from governmental, regulatory authorities, academia, private sector, ccTLDs registries from both countries.

Here is a, let's say, unofficial timeline, because the inaugural meeting of the Greek Generation Panel took place in December 2015. Actually, before officially submitting the proposal for the formation of the Greek Generation Panel to ICANN.

Since the Greek and the Cyprus governments give a lot of attention to the significance of the work being done by the

panel, there is also a ministerial decision of the Minister of Infrastructure, Transport and Networks of the Hellenic Republic that officially forms the Greek Generation Panel as also a working group within the Greek administration.

But the official formation of the Greek Generation Panel took place on October 2016, and we have to admit that our steps are relatively slow. We apologize for that, and we need your understanding because this is not the only thing that we are doing. But we are determined to conclude our work.

This first presentation in ICANN meeting of the work of the Greek Generation Panel is actually a milestone for us. That's why you can see it on the slide. And we hope that we will be able to finalize our work by June, or a bit later. We'll see.

This is the structure of the proposal. The Greek language, as you may know, is one of the ancient languages of the world. Through the years, through the centuries, have been changed. So, there were some questions about it.

Also, the orthography has been changed, so we have to consider those before conclude to the repertoire. We are drafting the proposal piece by piece, and the [process and] methodology will be written at the end.

Here, you can see some examples of issues that the panel has taken into consideration. There were two forms of the language. Katharevousa the older form comes directly from the ancient Greek, and Dimotiki is the modern form of the country.

As you can see from the two texts, this is the Lord's Prayer written in both forms. One has many letters have signs on them, and the other in modern form, they don't. So, the panel decided that only monotonic characters are to be allowed, because these are the characters that are used both countries officially. Polytonic characters could be used maybe in lower levels, second or third levels, but not for the top-level domain.

Here's another significant issue that the panel took into consideration: the final sigma issue. Sigma is a very common character, very common letter. Most of the male names end at sigma, so they use the small final sigma, not the third in the slide, and not the first one which is the middle sigma.

The problem was that in IDNA 2003, the mapping between them, if you use the capitals, was not precise, where in IDNA 2008, those two characters, sigma, middle sigma and final sigma are handled as completely different characters, so they are both accepted.

In the meantime, I cannot show you now, but there are still other issues under consideration and are under process. These are the

ones that you see on the screen. Within script variants, vowels with or without diacritics and cross-script variants: Greek and Latin, Greek and Cyrillic you already saw from the presentation of the Chair of the Cyrillic panel that that is such a case. And of course, whole label evaluation rules.

These things are being worked from the panel, and we hope soon we'll have some progress. That's all. Thank you for your time, and I'm available. This is the mail if you want anything to ask. We are at your disposal, and I'm here as well as the members of the panel for any questions you might have. Thank you.

SARMAD HUSSAIN: Thank you. We do have time for a comment or a question if anybody has one around the room.

UNIDENTIFIED MALE: A curious question about Greek. There is no other language besides Greek using the Greek alphabet?

PANAGIOTIS PAPASPILIOPOULOS: The Greek language uses the Greek alphabet.

UNIDENTIFIED MALE: No other language?

PANAGIOTIS PAPASPILIOPOULOS: No other language.

UNIDENTIFIED MALE: Okay.

PANAGIOTIS PAPASPILIOPOULOS: Yes. Greek and Cyprus, as I said, has Greek as official language. Cyprus has also two other languages, but that's irrelevant. Yes, these are the language. Thank you.

SARMAD HUSSAIN: Any more comments, questions? Asmus?

ASMUS FREYTANG: Well, the only other language that uses the Greek alphabet is, of course, mathematics. But we're not covering that here.

SARMAD HUSSAIN: Thank you. So, let's move on to our next presentation. We have an update from the Latin Generation Panel because Mirjana Tasic who's the Chair of the Latin GP. Over to you, Mirjana.

MIRJANA TASIC:

Thank you, Sarmad. Latin Generation Panel had some background. There were a few times that people tried to make Generation Panel to put it in function, but the last trial was during the last year, in the middle of the last year.

We revised the membership of the panel and we started to work on the proposal. The scope of our panel is the Latin script which is used all over the world. Those four things are the four or five slides on which specific terms are written, let's say it like this.

Okay, we have been revived during October of 2016, and in three months, we made a draft proposal and sent it to our IAP for comment last week. Now, we are waiting for Integration Panel comments on our proposal, and we shall continue to work on it after this ICANN meeting.

What is the subject of our work? It is our codepoints, Latin codepoints which are part of the MSR-2. Only lowercase letters will be taken in – oh, I'm sorry. I'm not speaking in the mic. We are taking care only about lowercase letters, and not all UNICODE ranges are included in MSR-2, so this is the list of the parts of UNICODE which are included in SMR-2 and which will be the subject of our work.

Latin script, as you know – and we are hearing this for the third time today – same origin as Cyrillic, Greek and Armenian, so there might be cross-script variance with these, Cyrillic, Greek

and – Greek is not yet finished, but Armenian and Cyrillic has finished their work and they have proposed their cross-script variants.

Latin is widely used by missionaries in creating alphabets for languages all over the world for the languages which didn't have their own script. So, Latin is very intensively used everywhere.

At this moment, we have some list of languages using Latin script, and at this moment, we have 455 languages on the list. This is just part of the languages for which we could find information on the Internet. We shall probably take into account only those languages which has EGIDS codes 1 to 5, and at this moment, we have a list of 300 languages with these EGIDS codes, qualifications.

I have prepared this slide before I realized that we have an error in the last statement. We have found 279 MSR codepoints for attestation in the Latin script, but there is more of them, so we shall recheck it later.

This is just to get an impression how Latin script is intensively used all over the world, and it is used – as I have found somewhere – for about 70% of world population or something like this. Maybe I mixed something.

The countries marked with dark green color are those ones where Latin is used as only script. Light green areas show countries where Latin coexists with some other script. And in this gray area, Latin is used for some other reasons. It is not used as an official script.

Our Latin panel is composed, at this moment we have 11 members although we are trying to find more of them. We have a lack of expertise for some languages like Chinese Pinyin which we shall discuss if we shall include it or not. For example, from native language in Asia Pacific region, for native languages in North and South America, and also we have a lack of expertise in the area of languages spoken in Africa.

On the other side, diversity of the composition of the Generation panel is pretty good. We have representatives for all categories which are expected to be included in a Generation Panel. And we have developed some workplan which consists of five steps.

Most of you who have finished your Generation Panels, you have passed all these steps. Our plan is to finish it maybe in July 2018, so at least we shall need one year and a half because there are a lot of languages and a lot of codepoints to be investigated. But that's the rough estimation. We shall see at the end when we start to define how shall we work in the future, how much time we shall need to finish the whole task of the Generation Panel.

This is all I have to say about it. Thank you very much.

SARMAD HUSSAIN: Thank you. We're done with the presentations. Let's open the floor for questions. We'll take any questions specifically for Latin Generation Panel first before we take more general questions. Any questions for the Latin Generation Panel?

Okay, so are there any questions anybody would like to ask or any comments you'd like to make for any part of this session today? Andrew.

ANDREW SULLIVAN: Imagine. I was just slightly worried about a comment in I think it was Ethiopic about not being able to do corpus analysis for some of the languages. This strikes me as slightly worrisome, so maybe somebody could say something more about that.

SARMAD HUSSAIN: Just to elaborate, I guess your worry is that if there is no corpus available, then what is the motivation of including that inside the analysis. Is that what you're suggesting?

ANDREW SULLIVAN: That's half of it. The other half of it is if you don't have a corpus, then how do you know you're getting it right? You don't have

anything to test against either, so you've got really two problems. First that it's not clear what the motivation is, and secondly, it's very difficult to know what success would look like.

DESSALEGN MEQUANINT: We performed corpus analysis only for [inaudible] language. So we used the corpus analysis as auxiliary evidence, not a primary evidence. So, we could have only done for – with one evidence, everything could be done, but we used that additional evidence just to strengthen our argument. But that's it.

I said for the eight language we used some primary evidences. For the two language which we think are relatively well-resourced, we also used additional evidence. So, we have used for example the online encyclopedia of the Omniglot, which publishes script versions for language.

So for the remaining six, we obtained that. Even for all of the eight languages considered, we have also eight script versions. So, the corpus analysis then is simply [secondary,] auxiliary evidence. We could have also avoided that.

SARMAD HUSSAIN: I guess another question to ask would be that, what level of linguistic expertise was there on the panel to evaluate the other languages in addition to the ones, the two for which you were

able to find corpus? So beyond online sources, was there any native expertise within the Generation Panel itself to do that work? Or if there was not native expertise, was there local material? What other evidence was used to develop that proposal?

DESSALEGN MEQUANINT: Okay. As I tried earlier to describe the composition of the GP members, we have two linguistics professors working for the university, so beyond that expertise, we can say that [their] expertise in many ways for most, not only for the eight languages, but for many languages. We used them as a resourceful person.

Besides that, even if their name was not included in the list, we also have been in continuous communication with scholars who are working on the digitalizability of the Ethiopic script. So, I think we have tried to use as much evidence as we could. To add one small comment, we used the corpus analysis as auxiliary evidence because we are cognizant of the inherent limitations of corpus analysis.

SARMAD HUSSAIN: Andrew, do you have a follow-up on that? Or should we move on? Okay. There is a question in the chat box that states that –

question is from Meikal Mumin. He asks, "Historically, there has been some use of Greek to write some variety of Turkish, known by the term Karamanlika. Did Greek GP consult linguists to verify if there is any contemporary use left among diasporas which may be eligible for consideration? Similarly, minority uses could hypothetically continue with Pomak, Jewish, Armenian or Macedonian communities.

PANAGIOTIS PAPASPLILIOPOULOS: Thank you for the question. As I said before, there is a range of scientists in the panel. We also have two linguistic experts in the panel, and also UNICODE expert. And actually, when it was the issue of thinking of other cases that Greeks maybe used, we excluded the Greek symbols like the case of mathematics and physics like Asmus said before, and also cases that small minorities of different cases might use some Greek characters. So, we'll not deal with this. Thank you.

SARMAD HUSSAIN: Asmus?

ASMUS FREYTANG: I think at this point, sometimes it's useful to remember that the key focus of this work is to cover everyday common use of a writing system. Some usage scenarios are, while real people

may be writing real text that way, not necessarily rise to the level of common everyday use for things like business transactions and similar things.

Not just for the Greek script, but overall, there are many scripts that are used to communicate a certain amount of cultural heritage in a way that is not associated with actual daily transactional life at this point, and maybe not in the future. One doesn't know how languages develop, and one has to be very careful in this business not to go and chase after stuff that isn't fitting the scope we have set for this project.

On the other hand, you want to make sure that you don't dismiss communities where some writing system is in fact used for transactional everyday life, and just because it happens to not be the government-authorized one or a high status one, it still exists.

So, from the Integration Panel, we are very careful in looking at the presented evidence to see whether it appears satisfactory that the right scope was met.

PANAGIOTIS PAPASPILIOPOULOS: Just to add something, we have the polytonic characters. The use of polytonic characters in Greece is extremely more frequent than the cases that the question referred to. There are

people of older age or from their belief they use some polytonic characters.

[inaudible] also uses polytonic characters. But as Asmus said, we thought that we have to be strict of the characters that are used in everyday life by the governments of those two countries, by the people, by the companies, by the press, by everybody for the economic transactions. That's why we chose to exclude the polytonic characters and deal only with the monotonic characters. So, this set of polytonics are a lot more than the cases in the question. Thank you.

SARMAD HUSSAIN:

Thank you. Any more questions around the room? None in the chat room. Okay then, we can then close the session. Thank you all very much for attending, and especially those who are attending online remotely. We will now be closing the session. Thank you. Bye.

[END OF TRANSCRIPTION]