EN

ICANN Transcription

IDNs EPDP

Thursday, 07 October 2021 at 13:00 UTC

Note: Although the transcription is largely accurate, in some cases it is incomplete or inaccurate due to inaudible passages or transcription errors. It is posted as an aid to understanding the proceedings at the meeting, but should not be treated as an authoritative record. Attendance and recordings of the call are posted on agenda wiki page: https://community.icann.org/x/rABmCg

The recordings and transcriptions of the calls are posted on the GNSO Master Calendar page
http://gnso.icann.org/en/group-activities/calendar

TERRI AGNEW: Good morning, good afternoon, and good evening. Welcome to the IDN EPDP call taking place on Thursday, the 7th of October 2021 at 13:00 UTC.

In the interest of time, there'll be no roll call. Attendance will be taken by the Zoom Room. If you're only on the telephone, could you please identify yourself now? Hearing no one, we do have listed apologies from Christian Dawson and Lianna Galstyan.

All members and participants will be promoted to panelist for today's call. Members and participants, when using chat, please select panelists and attendees or everyone depending on your Zoom update in order for all to see your chat. Observers will have you only to the chat access.

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.

Statements of Interest must be kept up to date. If anyone has any updates to share, please raise your hand or speak up now. Seeing or hearing no one, if you do need assistance, please e-mail the GNSO Secretariat. All documentation and information can be found on the IDN EPDP wiki space. Recordings will be posted on the public wiki space shortly after the end of the call.

Please remember to state your name before speaking. As a reminder, those who take part in ICANN multistakeholder process are to comply with the Expected Standards of Behavior. With this, I'll turn it back over to our chair, Donna Austin. Please begin.

DONNA AUSTIN: Thanks very much, Terri. Welcome, everybody. My name is Donna Austin. I am taking over, I guess, officially today as chair of this group from Edmon Chung. As all of you probably know by now, Edmon is stepping up onto the ICANN Board at the end of October and he has thankfully agreed to stay on in a co-chair role until that time. But as from today, I'm officially taking over. So many thanks to Edmon for getting us to this point. If everyone can please bear with me as I get my feet underneath me, hopefully we can have productive sessions and make good progress on where we're headed. So thank you, everybody.

Just one thing I wanted to note is that ICANN72, as folks know, is coming up at the end of September. We do have time on the schedule. I can't remember what that is right now. But what I want to ensure is it won't be at this usual time so I want to ensure that we are going to have critical mass on the call as we start to think about what the agenda would be for that call. So it looks like we're

scheduled for Tuesday, October 26 from 16:30 to 17:30. So that's something we'll come back to the group about just to see whether we would have critical mass. And if we do, we can discuss some of our charter items. If not, maybe we need to think about how else we want to spend that time.

Okay. As a result of some requests, Sarmad is going to take us through a presentation on the Root Zone LGR today and I hope that it will cover many of the questions that were posed by ALAC. I'm sorry. We haven't actually shared those questions on the list and we will do so after this call. And to our ALAC colleagues, if there's anything that hasn't been covered, then maybe we can come back to that afterwards and see how we want to manage that.

One of the issues that we potentially saw is that Sarmad is going to do a presentation very much from a technical perspective, some of it may overlap into some of our policy discussions. Please bear in mind that Sarmad is an expert in this field so we will give him the benefit of the doubt, I suppose, that if he's saying something and some folks think that it's getting into policy territory, please flag it but don't get too upset about it. We appreciate that some of this stuff may go a little bit that way, but for the most part, it will be very much a technical presentation and on the process for the Root Zone LGR. So unless I see any questions before we get started—I don't see any hands up—I will hand over to Sarmad. Sarmad, are you ready for us?

SARMAD HUSSAIN: Someone has a hand up.

DONNA AUSTIN: Anil, go ahead, please.

ANIL KUMAR JAIN: Thank you, Donna. This is regarding the timings for the next meeting. I suggest that 14th of October, whatever we have planned for the next meeting, we should continue that. But after that, we can break for two weeks until ICANN is closed on 28th. And after that we can do it because I also suspect that attendance here will be quite low because a lot of us will be making presentations in ICANN and a good number of them will be participating in various sessions which are of interest to all of them. This is just my suggestion.

DONNA AUSTIN: Thanks very much. Thank you very much, Anil. It's a suggestion we can take to the list. I think we also need to bear in mind that following ICANN72, there will be a number of time zone changes and we will have to reconsider whether this UTC time or whether this actual time for what it is for folks now actually works. So that's something else we're going to have to consider moving forward. I know there had been some discussion about whether we keep these calls to 60 minutes or expand them to 90 minutes. I think that's another conversation we're going to have once we get into the consideration of the charter questions more deeply. So if folks could bear that in mind and we'll put something to the list about how we want to handle Anil's suggestion and also what participation would look like at the time identified on the schedule

EN

for ICANN72. So thanks for that, folks. Sarmad, I will hand over to you.

SARMAD HUSSAIN: Thank you, Donna. Good morning, good afternoon, good evening, everyone. Last time we discussed to provide a brief overview of Root Zone Label Generation Rules. This particular presentation is going through some of the history on how the project or the work started. Also, we'll go into some of the design considerations and processes for developing Root Zone LGR. We will also talk about how Root Zone LGR works. Then also some relevant discussion around what could potentially trigger and change Root Zone LGR and how that could look like, and then we'll conclude on what's coming up as far as Root Zone LGR work is concerned.

The aim is to try to provide as much as information to you as possible for Root Zone LGR to help you go through the policy discussion. If you have any questions, please raise your hand or speak up during the presentation. Let's see if we would prefer that we have this very interactively rather than—so please don't wait until the end of the presentation to ask questions. All right, so let's get started.

So we'll start with a brief history of Root Zone LGR. Basically, quite early on ICANN community had identified that there is need for variant top-level domains in addition to just top-level domains in local languages, so IDN TLDs. Technically, the so-called variant top-level domains would be distinct but they call variants because for some reason, they are considered "same" by that particular script community. The definition of what is a same label really

depends or varies across scripts. So it's not really a cohesive definition and each script community determines what this definition is.

So when we started on this work early on, around 2010-2011, there was, of course, not as much clarity in this area, and therefore, what we started with as an initial study which was looking into issues around variants rather than solutions, it was called a Variant Issues Project. In that project, what was done was about six different script communities were invited to come and look at variant issues within their own scripts and perhaps share what they think are variants in their scripts and what are the potential issues associated with it.

The scripts, for example, included Arabic, Chinese, Devanagari, Latin, Greek, and I think Cyrillic as well. Eventually, these script communities came together and provided their own issue reports around variants which were all combined into a single Integrated Issues Report, which collated all the information from each of the script communities. That Integrated Issues Report was published in 2012.

One of the things which the issues report noted as a prerequisite to managing variant TLDs is that ICANN must have a way to validate potential IDN variant TLD labels when submitted and to validate all IDN TLDs requested for variant labels and variant conflicts. It also noted that there really should be a single shared definition of these variants from different applicants because these labels are going into the single zone, which is the root zone. That sort of laid the basis of the eventual work which has resulted into

EN

the development of Root Zone LGR as that single shared source of analysis for the root zone.

By the way, in the footer, I've also noted if the information in the slide is relevant to some of the questions which are raised in the charter, so some of this information, for example, could be relevant for questions A1, A2, and A4 in the EPDP charter. So, that information just in case anybody wants to follow up.

So, moving on. Based on that analysis in the Integrated Issues Report, the community came together and developed what is now called the LGR Procedure or procedure to develop Root Zone Label Generation Rules as a single mechanism for defining variant TLDs for the root zone. That procedure eventually was finalized by the community and adopted by the Board in 2013. The Board basically asked ICANN Org and the community to move forward and implement the LGR Procedure to develop the Root Zone LGR. So that is when the work formally started. So, there was actually a call made to announce the Generation Panels in 2013 after the Board's approval and the whole thing got going.

One thing to also note in this context is that SSAD Security and Stability Advisory Committee also did a whole report around IDN variant TLDs and some of the analysis which was going on and being published at that time, and noted also that the root zone must use one and only one set of Label Generation Rules from a security and stability perspective, of course. So, that sets the context of how the work on Root Zone LGR started.

So, moving on to the design and development of Root Zone LGR. Root Zone LGR is developed based on some core principles and

all the work which is done by the script communities and eventually evaluated, finalized, and integrated into Root Zone LGR is really driven by these core design principles which are part of the LGR Procedure and really drive the design of the Root Zone LGR. So I'm not going to go through all these principles. They will be available in the presentation. They're also available and explained in the LGR Procedure.

But in essence, what they're trying to capture is that the work done in Root Zone LGR has to be conservative. I think one of the key principles is this inclusion principle, which says that the way Root Zone LGR is designed—there are two ways it can be designed. One way is that you start with everything possible, and then take out things which are you're not sure of. That is normally referred to as an exclusion based principle versus there is the inclusion based principle which says that you start with an empty sort of set of slots, and then one at a time keep adding things to that empty repertoire to build Root Zone LGR. And every time you add something, you make sure that what you're adding to it is secure, stable, and is being done properly. So it is inclusion based. It uses an inclusion based method, not an exclusion based method.

It also has a conservatism principle, stability principle, which means that it is not possible to put a lot of things into the Root Zone LGR, and then later on find out that some things don't work so you have to take them out that could make the Root Zone LGR very unstable. So the suggestion, of course, the design principle is that be very conservative and when you're adding things and if you're in doubt on whether something could be added or not, if

there is any doubt, it has to be left out until that doubt is removed and then it has to be added in. That means that you will start with a smaller set of things with Root Zone LGR but it will be very stable and it can grow as more learning comes in and as more information becomes available.

So many of these principles, as I said, which governs the design of how Root Zone LGR is designed and these principles obviously not only apply to code points but also how variants are defined, how rules are defined, and so how the whole Root Zone LGR actually is developed by the community members. Again, this lays the basis of and motivation for the design for Root Zone LGR. So now let's get into—

DONNA AUSTIN: Sarmad?

SARMAD HUSSAIN: Yeah?

DONNA AUSTIN: Sorry. I see Hadia has her hand up. So maybe if you go to Hadia. Thanks.

HADIA ELMINIAWI: Thank you. If we can go to the previous slide, please? So in relation to the conservation principle and that any doubt should be resolved in favor of exclusion of a point rather than inclusion, that

means that if there is doubt then the code point will not be included, right?

SARMAD HUSSAIN: Yes.

HADIA ELMINIAWI: But maybe this was good prior the introduction of new gTLDs. But now that we are going forward with new gTLDs and a code point might be actually, what if—so the basic question here which maybe I didn't understand the technicalities that much, but if we are to assign a top-level domain, an IDN top-level domain, and then it turns out to be confusing, right? What happens? What happens next? How can this be resolved? Do we go back to the Root Zone Panel and what happens? I think this is the basic question that really affects our work. Thank you.

SARMAD HUSSAIN: Sure. There are two different layers we are talking about here. The first layer is to have a list of characters or letters which can formulate a label for top-level domain or for root zone. And then the second layer or second step is to choose from those characters to actually make a label or define a label. So I think when we're talking about these design principles, we're talking about the first layer, which is actually saying that these are the characters which are possible. It's not really talking about how TLDs are formed at this time, that sort of next step and we'll talk about that later in the presentation as well.

EN

So just to give you an example, in some scripts when we were, let's say, working on Latin script or Arabic script, we found out that there are some characters which are used in some languages. So let's talk about Latin script. So there is the use of Latin script, for example, in some languages in Africa, for example. The Generation Panel found out that there were some potential letters which a language was using but they couldn't definitively say that that letter is actually used by that language or not. Therefore, they weren't really sure whether that character should be included or not. So that's a situation where there is a doubt. And if there is a doubt, then until we get clear information that a particular character is actually used in a language and therefore should be included, if there is a doubt that the conservative principle says that you have to leave that character out of the repertoire at this time until eventually you find more evidence and clear evidence and that removes the doubt, and then it can be later added on maybe in a future version of the Root Zone LGR. So that's what we are talking about from a principle perspective. The example you were sharing was the next step when we then start using these core points to actually develop top-level domains. Okay. Edmon, please.

EDMON CHUNG: Yeah. Sarmad, I'm not disagreeing with anything you said. I actually very much agree. Just to highlight one thing that part of A3 in the charter is to consider potential issue that, let's say, a character was considered to be excluded in the LGR but an applicant thinks that particular character is very important for their TLD, then if they do actually apply, is there a process or should

there be a process to trigger a review of the LGR is something that is useful for consideration for this group? So I just want to highlight that.

SARMAD HUSSAIN: Thank you, Edmon. And yes, we'll actually do the later part of this presentation actually come to those specific questions on how can LGR, for example, be revised. But then I think what you raise is also important in other ways that can then be—I think that's a separate discussion on how objections like that or, I guess, requests from applicants like that would be treated, but that's obviously a policy discussion.

Coming back to the presentation again. Again, when we go forward and see how Root Zone LGR is developed, actually, these principles are the frame against which each LGR proposal is evaluated and eventually finalized. Actually, even before that, these are the principles which drive the development of each of the LGR proposals for different scripts. All right, so moving on.

So we are then going into how the LGR is developed. We now know the guiding principles which motivate the development. One of the things which was realized very early on in the process was that when we're developing solutions for different scripts, obviously the expertise and know-how of that particular script lies with the community which uses the script. Therefore, for us to get a solution which is "right," we really have to get the community involved and ask the relevant script community to provide the solution to us. Therefore, the procedure which was developed was that—so the LGR Procedure said that we identify initially all the

different scripts for which we want Generation Panels. And then we actually go to those script communities, we request them to formulate a panel of volunteers and look at these design principles, and based on these design principles, develop a solution for us. Then that solution comes to us, meaning ICANN as a community, and goes through some processes. I'll take you through the processes as well. But technically speaking, the script community as a Generation Panel develops a proposal. So there is a Generation Panel for Chinese. There's a Generation Panel for Japanese, there is a Generation Panel for Cyrillic. So these are script based Generation Panels, not language-based Generation Panels. They look at all the different languages which choose the script, and based on that develop a solution. That solution is then eventually given to an Integration Panel. The role of Integration Panel is the two or three different roles of Integration Panel.

DONNA AUSTIN: Sarmad? Sorry, can I just interrupt before you move on?

SARMAD HUSSAIN: Sure.

DONNA AUSTIN: Just a question about how easy or difficult is it for you to identify a language community that would form a Generation Panel?

SARMAD HUSSAIN: That work was done by the Integration Panel. That was one of the roles of the Integration Panel, to start with what is referred to as the maximum starting repertoire. So, for example, Unicode has currently 154 scripts listed in it. One of the things which Integration Panel has done—this is again back in 2013—was they actually went through the whole of Unicode and did two things. First of all, they shortlisted some of the scripts which are clearly broadly used by the communities across the globe, and therefore, we really need Generation Panels for them. So they identified 28 scripts from the list of 154 scripts and said that "These are the scripts we really need to get going with." Then for each of those scripts, they also did a short listing of the initial set of characters which are good candidates. When I say "good" candidates, it means that they actually looked at the design principles and saw that, "Okay, there are some characters, for example, in let's say Arabic script which are encoded in Unicode but Unicode says that they should be deprecated, meaning their Unicode standard says they should be used even though they're encoded. Since Unicode itself is saying that the code point is deprecated, it should not be, for example, shortlisted for root zone labels. So this initial analysis was actually done by the Integration Panel. The Integration Panel, once they published the MSR or maximum starting repertoire, provided ICANN Org with the initial list of scripts we needed to work on. And then we took that list and started outreach to the communities to help develop the Generation Panels. Does that answer your question, Donna?

DONNA AUSTIN: When you did that outreach, how difficult was it or how easy was it to form those Generation Panels?

SARMAD HUSSAIN: It was a bit of effort, I think. Obviously, there was a lot of interest in some communities. Some communities, obviously, didn't know about this work. So there was a certain bit of outreach which was needed to actually, first of all, inform people of this work and then get a group of people together who can actually have sufficient expertise to form a Generation Panel and do the work.

DONNA AUSTIN: Okay. Thank you, Sarmad.

SARMAD HUSSAIN: There is actually a slide which is coming which shows when each of these Generation Panels was formed and how long it took. That may actually provide some more information.

It did vary from script to script. Some script communities were already there and ready to go, whereas some obviously needed outreach. But the process has been that each script community develops a solution. They keep working with the Integration Panels, so there's a back and forth between Generation Panels and Integration Panel. They develop sometimes solutions, sometimes Generation Panels have questions. It's a very iterative process where Generation Panels and Integration Panel actually work together towards a proposal. Once a proposal is developed and accepted, then it gets integrated into the Root Zone LGR. It is

done one script at a time. So it's not that all scripts have to be done in parallel. Each script community reforms and finalizes its proposal. It can actually be evaluated by Integration Panel and then taken forward, though there are some dependencies which we'll talk about later. All right. Let's keep moving on.

It's obviously important to know what Integration Panel and Generation Panel would look like. Sorry. You have a question? Lisa, please.

DONNA AUSTIN: Lisa, if you're speaking, we can't hear you.

SARMAD HUSSAIN: The mic is muted. You can't unmute.

DONNA AUSTIN: Terri, can you work with Lisa in the background and we'll continue on and come back to Lisa once we've sorted out the problem?

TERRIE AGNEW: I sure will, Donna.

DONNA AUSTIN: Thank you.

SARMAD HUSSAIN: Okay. Thank you, Donna. So moving on. The integration as per the LGR Procedure, Integration Panel should consist entirely of experts selected by ICANN on basis of established expertise in certain areas. The expertise is listed in the second bullet. Well, of course, they need to be impartial and therefore shouldn't have any conflicts. Also, the panel should consist of at least one expert in Unicode issues, at least one expert in IDN and DNS issues, or one for each, and at least one expert in linguistics and writing systems. Or this one could overlap with the one person who also is an expert in Unicode issues because I know that people with expertise in Unicode have also expertise in linguistics.

LGR Procedure actually notes that it's worth emphasizing that the supply of actual general experts in any of the relevant areas of expertise is extremely limited, which obviously was the risk of in the procedure. But early on, there was a poll made and invited applications for Integration Panel members. Based on that, there were five members which were selected. Three are required but we've included more experts just to make sure that this is a robust process. We have two Unicode experts, two IDN and DNS experts, and one linguistics and writing systems expert. These experts and their bios are posted online and I've included the link in the presentation in case you want to take a look at who these people are. Moving on.

Generation Panels obviously also need to be composed of a set of people who bring in some broad expertise. There was actually a call for Generation Panel made in 2013 which explains the purpose and requirements on Generation Panels as well as the expertise required for each of the Generation Panels. As far as

expertise is concerned, all Generation Panels should have significant expertise in the writing system concerned but have neither overall expertise in all of Unicode nor expertise in any other writing system. It's a very script-specific Generation Panel. From a diversity point of view, the Generation Panels need to have some diversity of participation in order to be useful. They must have sufficient number of participants and should be diverse in economic interest. The work of these panels is considered with the technical issues and involves linguistic expertise and is not a representative in nature.

I think one of the things which it says is that the panel requires general expertise in the script and diversity needs to be that you are capturing diverse perspectives. But it is, I guess, realized that you can't make a representative panel. Representative means that, for example, you may actually have a script which is used by 100 different languages. It may or may not be possible to have one representative of each language in the panel. Therefore, even though the panel should have general expertise in the script, it may or may not have actual expertise in every single language which uses the script in writing. As far as diversity and capturing linguistic diversity is concerned, we actually raise more details later in the slide as well. But let me stop here. I see Satish's hand up. Satish?

SATISH BABU: Thanks, Sarmad. On this point about that it is not meant to be representative of what the language is, yesterday we had a presentation from one of the Latin GP experts on the panel. He was saying that they were about—the Latin script that all the

[inaudible] are used by about 400 different languages. They have shortlisted based on various criteria about 200 plus out of these. There are seven experts and 200 plus languages. For me, this looks like a fair amount of lack of diversity. How do you respond to this?

SARMAD HUSSAIN: There is actually a slide which we will come to which addresses precisely this question. Would you like to wait until that time and then we can discuss it then?

SATISH BABU: Yeah.

SARMAD HUSSAIN: Thank you. As far as the composition of each Generation Panel, I guess, asked for, it should contain a chair community representative, at least two linguistic representative or experts, at least one to two registry/registrar representatives, at least one DNS IDN Unicode expert. The minimum size is normally five to six people. But based on scripts and based on the actual work which has been undertaken, we've seen that different Generation Panels membership has varied from as low as seven members to as high as more than 60 members in some panels. So there's actually been a great diversity of, I guess, experts in each panel. Basically what we do is when the panel publishes their work, we request that the documents include that then also the list of people who would actually contributed for everybody to know who has been involved. Moving on.

GP work organized. The GPs have expertise in the script as shared but not required to cover all the languages quoted by the script. The way we've run the process, even though normally the procedure basically said that ICANN Org … For each GP membership, people could apply for membership and ICANN Org would review it and finalize the GP members, but we've actually kept it open with the no upper limit, as I shared earlier as well. Anybody who was interested could join our Generation Panel. Even the procedure did require—it was said that there could be a review but we've actually kept the Generation Panels completely open.

For the languages, supported additional expertise sought through outreach to the relevant community, online research and consultations with the invited experts as needed. I guess, Satish, this is going back a bit to your question as well, that it is not possible for any Generation Panel to have expertise in all the possible languages which are used by the script. In some cases, it's possible where there are not a lot of languages. In some cases, that may not be possible. In such cases, the Generation Panels have undertaken various strategies to get the information they need to get. That has included outreach to the relevant community. We've actually supported them. So we've had them reach out to some community members for a particular language community in case there were specific questions about that language community.

An example is that Latin Generation Panel actually needed some clarification on how dotless i is used, for example, in Turkish language. So we'd actually reached out to experts from Turkey,

they came on call and had conversations with the Latin Generation Panel for a couple of meetings, explained how they, for example, use dotless i and i in Turkey into this language. That, obviously, dialogue helped the Latin Generation Panel figure out how to formulate rules for the i and dotless i for Latin Generation Panel.

So that's an example of outreach of the community. There is obviously online research, which is done by the panels. They're in consultations with invited experts. Where there was information required, we went and got that online or reached out to the people to consult.

Most GP had regular online calls for members to be able to conveniently attend. So most of the work was done online. Basically, ICANN Org had presented all the Generation Panels facilities to host mailing lists, Wiki pages, as well as online calls. Where the panel is agreed to review ICANN's facilities, those were provided. In some cases, the panel's decided to manage this on their own. In those cases, obviously, ICANN supported them to the extent they requested. Where the community is less geographically distributed, some panels we're even having face-to-face meetings and outreach events with local communities.

For example, in Laos, the language is geographically not very distributed, the script is used only in Laos PDR, largely. So they were having face-to-face meetings for their review and at the end of the face-to-face meeting, they actually held a public session where they presented their research and invited the larger community before presenting the work to ICANN and going to the ICANN Public Comment process. So different communities

obviously organized this work in different ways based on obviously their own geographical distributions.

The GPs reached out to additional experts and community through their internal outreach as well and also to presentations and local forum. So for example, I know that the Arabic script community not only reached out to people to ICANN community forums but also through, for example, their multiple local forums within Middle East where the Arabic word was presented.

Eventually, obviously, once all this work was completed, for each of the panels, the panels would submit the proposals to ICANN. While even the work was being done, all the script panels which have been actually active, ICANN Org has been regularly inviting them to the IDN update sessions we've been holding at ICANN meetings. At each of those meetings, we invite all the active panels and they present the work under progress and the work they're doing, the challenges they are facing and how they're addressing it. So it allows an ongoing interaction with the ICANN community as well for all the different Generation Panels. Let me stop here and see—Satish, does this address some of the concern you have raised? But if you want to take the floor again—

SATISH BABU: Thanks, Sarmad. Don't get me wrong. I completely appreciate the voluntary work that the GP and the IP has been doing. My concern arises from the fact that we are moving into an automated regime. And one of the points that yesterday's presenter was pointing out was that among the seven experts, there was a degree of lack of coherence on some of these points. Now, what

EN

is it confusable to an end user is not necessarily confusing to an expert. Firstly, because they are experts, it's their job to know. But secondly, also because they are comparing side by side two things. When the end user is going to eventually use the TLD, there is no luxury of side by side comparison. So if you have something that looks like something else, that's going to create a problem for the end user, for the applicant, and registrant. So my position is that we should be careful in what goes into the Root Zone LGR so that when we move to LGR as a single authoritative zone, sole for variants as well as for validation, we should not get into undue trouble. And in this case, the seven experts for 200 plus languages, it has taken them seven years. And they were saying that they did not think that they were going to go back and do this job again. So that's the other side of it. How easy is it to trigger revision of the GPs, re-examination based on challenges or whatever else?

SARMAD HUSSAIN: Thank you, Satish. There are the two separate questions and I'll take them one at a time. The first, I guess, question is about— you're talking about confusability or similarity. I do want to point out that the LGR Procedure from a similarity perspective has very clear direction in that it says that things which are absolutely obvious are those from a visual similarity perspective could potentially be captured as variants. But from the confusability part, largely would eventually go not through the variant process but to the string similarity review process and that larger net, which is the string similarity review process is, obviously, there to make sure

that the concerns you are raising, Satish, are actually addressed beyond the Root Zone LGR.

As far as the second question is concerned on how easy or difficult it is going to be, yes, it has taken Latin Generation Panel and some other Generation Panels years of hard work to get to these proposals. It has been tremendous effort by the community. But when we go and change or review these rules on LGRs, please note that the review is going to be very limited in scope. Currently, if they were assessing inclusion of, let's say, 300 different characters, the review maybe for two of those characters or one additional character, which is going to be much limited in scope. As you can see, a much more contained problem than the initial problem which they were solving for, let's say, 300 different characters. Let me stop here and see if—Edmon has his hand up. Edmon, please.

EDMON CHUNG: Building on, Sarmad, what you just said, I think in a little bit in response to Satish. I think, Satish, what you raised is very important and is within scope of this working group. However, that is not part of the LGR. The LGR deals with IDN variants which are linguistic or technical base that Sarmad explained. It does address some of the confusingly similar issue, but not all of them. But if you look at the Charter Question E, later on down the path, we do also come back to string similarity or confusingly similar issues. This is certainly something that this group needs to deal with, but just not within the LGR process. I just want to highlight that.

SARMAD HUSSAIN: Thank you, Edmon. Satish, do you want to follow up?

SATISH BABU: No. I'm good. Thanks, Sarmad and Edmon.

SARMAD HUSSAIN: Thank you. Okay. Then moving on. One of the questions which came to Integration Panel very early on and also the Generation Panels very early on was that when they were developing the Root Zone LGR proposals for a particular script, there were some scripts which were actually being used to write many, many, many different languages. We just heard that for Latin, there are more than 400 languages, which are actually using Latin script as the writing system. There are more than 100 languages, for example, which use Arabic script for the writing system. So there are many different scripts out there which are being used by many different languages. The question then came that how diverse or how much support should actually be considered by the Generation Panels.

So the LGR Procedure actually had already as one of the principles said that "We're looking for stable orthography." If the orthography is unstable, then obviously that raises a question mark and then the design principles say that if there is a writing system which is for a particular language, which is probably not very stable and not very widely used—widely used being obviously also one of the ways one could say that it is stable— then it obviously creates a problem. So I guess the question then was at how do we know whether which languages to include,

EN

which languages to not include. For that Integration Panel actually went to a third party source, this is work done by ethnologue.com. It's one of the foremost sources of linguistic research and documentation available and they present what is called the language status. For each of the languages across the world, they document thousands and thousands of languages. The language status is basically developed using what they call is the EGIDS value. They take the EGIDS value from zero to 9 or 10. Basically going from zero which is international language like perhaps English or French, all the way to languages which are now extinct, and therefore no longer spoken. Of course, if there is an extinct language, we shouldn't really probably be working on that for Root Zone LGR but we should certainly be working for international language and including international languages in the Root Zone LGR analysis.

The question was that where do we draw the line? The Integration Panel had recommended to Generation Panels that anything which is six and above is probably not very in widespread use. Anything which is four and below is certainly something which has to be captured. Then level five is in the gray area, in the middle, where the Generation Panel should discuss and decide whether they want to include languages which are at level five. But for all the languages which are used at level five, the Generation Panels need to make a case of why they should be included because it is an inclusion based standard. So number five is not automatically included. If they want to include something, they have to make a case. Number six and above are not included at this time because community is not really using that language actively. But to ensure that there is reasonable diversity in the Root Zone LGR, all the

EN

relevant languages are considered. The Integration Panel actually asked the Generation Panels to include at least all the languages which are up to level four and possibly some languages which are level five using that particular script.

Let me stop here and see if there are any questions. Including, at least, level four and possibly some level five and showed that there will be sufficient diversity in the solution for the Root Zone LGR. Then excluding some of the higher numbers ensured that any instability actually is prevented.

Based on the design principles, the Integration Panel suggested GPs to consider at least all languages in the EGIDS value of four or lower to ensure linguistic diversity is captured. Languages with EGIDS scale five may be considered if GP find sufficient evidence of general purpose use of that language. There were many cases where they were included and many were included.

Then each GP has been requested to document the relevant languages in their proposals and is reviewed by IP to make sure linguistic diversity is maintained. So for each script, the Integration Panel asked the GP to document all the different languages using that script and their EGIDS value and then share what is the shortlisted languages they're using based on EGIDS values and other factors. This is one of the things which IP evaluates to make sure that linguistic diversity is captured. Let me stop here and see if there any questions.

EN

DONNA AUSTIN: Sarmad, just a time check that we're at the top of the hour. We've got 30 minutes to go. Thanks.

SARMAD HUSSAIN: Sure. Thank you. All right. We'll go slightly quicker. Process to create and integrate script proposal and Root Zone LGR, this is script community forms a GP, the GP explores the languages that should be supported, and their needs in terms of code points, variants, and rules. So they do some initial analysis. The GP develops proposal for Root Zone LGR based on that analysis and submits a formal solution through XML, it's a machine readable format, and an explanatory document or additional documents like appendices so there are some supporting documents in addition to the XML. But XML is the normative form and that's the one which is fully integrated. Supporting document is to just understand what the XML is encoding.

GP normally would have iterations with IP and discussions with IP over to develop a solution which meets the design principles, and once they have finalized it, they will publish it for public comment. After the public comment is received, GP inputs the comments and finalizes their proposal and submits to IP. IP reviews the proposal based on the design principles and accepts the proposal and integrates it. If it meets the design principles, it could actually return the proposal back to GP and say that maybe there's some issues which still need to be resolved. If it is integrated, GP actually updates the Root Zone LGR to the next version and publishes the integrated version for public comment, and then finalizes the Root Zone LGR's next version based on public comments. So there are two public comments in each cycle. First

EN

is for each script, and then when the scripts integrated, then the Root Zone LGR also goes through a public comment.

Summary of work which has been done so far. So this is the list of all the 27 panels and scripts which have worked. It gives the starting and ending duration times and also the number of actual days they have worked. Averages around 1100 days per panel, and the minimum, as you can see, is 160 for Georgian, and the longest one is for Japanese which is 2389 days, but we still have my Myanmar GP going hopefully that will be done sooner. But so far, that's the effort put in. So you can see that it's for most of the Generation Panels, it has been a multi-year effort. It has taken considerable amount of work and discussion in these panels to formalize their results.

So this is a history of how Root Zone LGR was developed. In 2013 the call was announced, 2016 the first panel finalized their work and it was integrated in the first version of Root Zone LGR. In 2017, a few other script panels have also finalized their work—Ethiopic, Georgian, Khmer, Lao, and Thai—which were integrated into the second version. And then since then, we've had a couple of other versions. The current version is Root Zone LGR-4, which includes all the different languages which have been integrated through LGR-1 through LGR-4. In 2022, we are anticipating that we will be able to integrate many of the remaining scripts which are listed here. Many of them have already completed. Japanese and Latin are currently open for public comment, and Myanmar is just finalizing their work.

This is again a summary of where we are, these many scripts which are integrated. There are some scripts which are finalized

but waiting for integration, and that's because the Integration Panel thought that Romanian, Cyrillic, and Greek are too close to Latin to be integrated independently and they were waiting for Latin solution to make sure that they are all coherent with each other and based on that now since Latin is also finalized and they will integrate it.

Japanese and Latin are in public comment right now. Myanmar is almost final. For Thana and Tibetan scripts, we've not been able to put the Generation Panels together yet. So we have been working with these communities. There are some challenges but we continue to work with them. The last couple of years have been even more difficult because of COVID, there's been travel issues. But in any case, we will continue to work to see if we can get them on board as well.

Just to, I guess, wrap up on the process of development of Root Zone LGR and how it's actually going to be used. So now Root Zone LGR is at a stage where it's almost final. Hopefully, next year we'll have the Root Zone LGR-5 in the next six months or so. Eventually, this Root Zone LGR gets used for defining variants. The initial work was done by staff. When they published the recommendations for managing the IDN invariant TLDs, there was an initial report which actually included a special—there was six reports in this publication. One of them focused on the rationale for Root Zone LGR and talked about why root zone is a good source or a single source to use for labels in the root zone.

Also, in 2019, the Board resolved to take the IDN variant TLD papers. They resolved to share them with GNSO and ccNSO, and to request GNSO and ccNSO to take these into account while

developing their policy. And they did a follow up. The Board also asked technical community to do a follow-up study on technical utilization of Root Zone LGR, which was also published and then also forwarded to Board resolution to GNSO and ccNSO for consideration. In 2021, GNSO published the report on new gTLD Subsequent Procedures, which incorporates the use of Root Zone LGR for the next round of new gTLDs. So that's sort of the current status and some of the recent documents and the Board resolutions around Root Zone LGR.

Okay. So let me take a quick pause here and see if there are any questions. Otherwise, we move into how the Root Zone LGR works. All right. So, this is a schematic or illustration which explains how Root Zone LGR is going to work. As we've discussed, we have all these 28 different scripts or LGR proposals which will be eventually integrated into the Root Zone LGR. What will happen is that we have a tool and we also will eventually have a tool which will load up these files and with these files be able to answer some questions against the labels which are sent to the LGR tool.

So there can be two kinds of questions, one related to the existing TLDs. In the case of existing TLDs, we are not going to ask whether the TLD is valid or not, it is obviously already delegated so it is considered valid. But what we want to find out is whether it has any variants or not. So the tool is going to calculate or will be able to tell whether there are variants and which, if any, of those variants are allocatable variants and which of those are blocked variants. So, actually, it gives two pieces of information whether something is a variant of existing TLD or not, and then if it is a

EN

variant, whether it is an allocatable variant or it is a blocked variant. If you apply with a new label, you get this information, but before you get this information, you also actually do another review, which is whether the applied for label is actually valid or invalid. So, the first question to answer is whether the applied for label is valid. If it is valid, then the next question is what are the variant labels for that particular applied for TLD? And then you get similar kind of answers. You get a set of variants, some are allocatable, and some are blocked.

So, basically, the Root Zone LGR contains a list of characters or code points, a list of variants and a list of rules which can apply on code points or code point sequences. So for a TLD label, all code points. So if something has to be valid, all code points forming the label must be in the list of code points which are available. If Arabic script or Latin script or Cyrillic script, it has a list of code points and you're applying for a TLD label which is formed with the code point which are not in that list, it will come out as an invalid label.

None of the code points or code point sequence should violate applicable rules of them. So sometimes for complex scripts, there are additional rules on code points. So, for example, there are tone marks in Lao or Thai languages. And the tone marks can only, for example, come on top of a consonant or it can follow a consonant or it can follow potentially a vowel which is following a consonant and so on. So, they're very strict kind of constraints in the orthography of that particular script. So if you, for example, put five tone marks on a consonant, that's not going to work because that's not how the script works. So it's totally unpredictable for

EN

local script users. Therefore, there are rules in these particular, I guess, LGR solutions, which in a way contain those restrictions. If you have a label which is violating any of those restrictions, the Root Zone LGR is going to say that your label is invalid.

So the label can be invalid in two ways. One, that it contains some code points which are beyond the ones which are shortlisted by the Root Zone LGR. The second one could be that it has all the right code points but they are coming in the way that they're violating some of the constraints on our label can be formed for that script. So, if the label constraints are met and the code point constraints are met, the label will come out to be valid. Otherwise, the label will come out to be invalid. So, once the label has been determined as valid, then one can calculate the variants of that label as well.

So, first of all, for any label, the variant rules are applied to determine if it is part of the variant set. So there's now variant rules which are separate from the label rules. So variant rules say that these are the two code points, for example, which are variants of each other, or these are the three code points which are variants of each other. So, if you create one label, it will use the variants to create all the other variant labels as well. Once the variant labels are created, one of the first things the tool does is actually goes and validates all the variant labels and make sure that each of the variant labels is also valid. Because if one of the variant labels is invalid, it cannot proceed. So, the initial label should be valid, all the variants which are generated should be valid as well. We cannot have invalid labels.

EN

Once the validity of variants is calculated, then it also calculates whether a particular variant is allocatable or blocked. Allocatable means that that particular variant is a potential candidate for possible allocation or delegation at a later stage or at the same time or depending on that policy procedure information. Blocked variant means that even though it is valid, it is not a candidate for possible allocation and delegation. So whether a particular variant is a good candidate for allocation or delegation or not a good candidate and therefore should be blocked is a decision which has been made by the script community. So the script community is in the Root Zone LGR solution telling us that for a particular label, these kinds of variants should be allocatable and these kinds of variants should actually be blocked. So that is actually how it is encoded into the Root Zone LGR and then that is how it's determined. Based on a label, it is algorithmically done through the tool. So it's not an arbitrary solution from a tool perspective with this, just running the XML file which has been designed by the script community to get those answers.

This is sort of a schematic for the top part of that is validating our label. And then the lower part of this is to determine whether a particular variant label is, first of all, valid and then if it is valid, whether it's allocatable or blocked. In each stage, it is actually using the different parts of LGR for that purpose. I'm not going to spend a lot of time. This is just trying to capture the previous slide.

DONNA AUSTIN: Sarmad, before we move on, Jeff has his hand up.

SARMAD HUSSAIN: Okay. Jeff.

DONNA AUSTIN: Go ahead, Jeff.

JEFF NEUMAN: Thanks, Donna. Thanks, Sarmad. I think that it's one thing to see these kinds of things on slides. And I think we said sort of the last time if there's a way to show this stuff in the actual tool itself, that would be really helpful. Well, I understand the rules you presented on the last slide. It is so difficult in the abstract to understand. For example, I didn't understand—I'm trying to think of an example of a label that's valid but has invalid variants, and therefore, you can't allocate the valid initial label. First of all, that doesn't sound right to me but it's so abstract at this point. We really need to see these things in a live demo, not a PowerPoint slide but kind of actually going through it with examples. So if someone applies for this TLD, even though the code points for that TLD are valid because these are the variants that would be produced and this variant in particular is not valid … We need to actually see that I think in person.

SARMAD HUSSAIN: Sure. If you like, we can organize a demo as well for the working group. We are more than happy to create some examples and do a live demo as well.

EN

JEFF NEUMAN:	Yeah. That would be really helpful. That's what I thought we had requested last week. I mean, this is great and this is good background and we need it. But I think the demo is really going to be important for us, too. It's hard to design challenge rules, for example, if we can't see an actual case that could produce a challenge, right? So if we could have a demo, I think that's a necessity for this group. Thanks.
SARMAD HUSSAIN:	Sure. We'll work on that and then come back to the working group, and maybe we can take another small chunk of time in one of the subsequent meetings to provide a demo as well. Okay. So we move on.
	So here's a worked out example. So what I did was I took Shabaka which is one of the gTLDs in Arabic script. It's one of the delegated gTLDs right now. I just ran it through the XML for Arabic Root Zone LGR. It basically shared the allocatable and blocked variants, and it gives all the different U labels, A labels, and the code point sequences. We can take you through sort of a more detailed guided tour of the tool and how it works, actually, as just discussed. We'll try to put something together and share with you in one of the future meetings.
JEFF NEUMAN:	Can we get back to that? Thank you. This is great. Just to get things kind of straight, just because something's technically labeled as allocatable, they were not allocatable in the last round, right? I mean, there was a policy decision or there was a decision

made that said that no variants can be allocated. But what you're saying here is if the policy, which is what's recommended, allows the allocation of allocatable variants, then these could be allowed. But in the last round, they were not. Is that right?

SARMAD HUSSAIN: Right. Actually, last round, we didn't even have these definitions. So we didn't even know what the set was.

JEFF NEUMAN: Okay, true. But if we were applying the last round rules and we did know it, we still wouldn't have been able to allocate those two strings.

SARMAD HUSSAIN: Right. Well, first of all, we didn't even know what is allocatable. Was it blocked even if we knew what was variants in some cases? And then once we know what is allocatable and blocked, then we need actually policy to decide what to do with allocatable ones and what do with blocked ones. That's also, of course, was not defined. For example, how to take an allocatable variant to, for example, something which is activated or allocated and activated and delegated, that policy obviously was not available before in the previous round. Sorry. Does that answer your questions, Jeff? I'm sorry unless I misunderstood your question,

JEFF NEUMAN: Yes, that does in the sense that just because something is now labeled allocatable, that's what we're deciding the rules about in our PDP, which is if it is allocatable, who can it go to, under what conditions, etc.?

SARMAD HUSSAIN: Right. All these strings, by the way, which are visible in some part of Arabic script community, they would consider each of these strings to be the "same". So that is why it is suggested that these variants go to the same entity, otherwise, it can create confusion.

Another thing I wanted to note—and see a couple of hands up as well—just a quick comment, that you will see that there were fewer allocatable variants than blocked variants. There are many more blocked variants. And that's part of the design of the Root Zone LGR as well, which actually asked to maximize the blocked variance and minimize the allocatable variants. So we have Hadia and Edmon, please.

HADIA ELMINIAWI: Thank you, Sarmad. I would just like to note that allocatable, that does not mean that it would be necessarily awarded because it will still need to go through a string similarity review process. The way I see it, all of those allocatable would be deemed as similar if they go through a review process. So am I right? Thank you.

SARMAD HUSSAIN: Right. Actually, the similarity review process would need to be reviewed in the context of how to resolve contentions when it

EN

actually is a variant label being applied by the same entity. I guess that's something which is a separate discussion. But variant labels, in many cases, may actually be visually same or similar. Not always, but they could be really same or similar. So obviously, that's something which needs to be eventually considered in the similarity review discussions.

HADIA ELMINIAWI: Yes. So just to make sure that being allocatable does not mean that it's a pass for it to be awarded, right? It still needs to go through other processes in order to say that string could actually be awarded. And this is actually the policy that we are going to be working on, right?

SARMAD HUSSAIN: Yes, certainly. One of the things which needs to be defined beyond what is already defined in the SubPro is that what are the steps which will take, for example, allocatable variant label all the way to delegation? Thank you. Edmon?

EDMON CHUNG: Yeah. Just quickly on building on, Sarmad, what you said in response to Jeff, three things. One, first of all, it's about the 2012 round, the self identified variants and looking at the LGR and understanding whether it's allocatable and blocked, then this group will need to think about how to then allow the 2012 round applications to move forward to actually allocate and delegate the variants. Number two is, of course, going forward actually using this and dealing with allocatable and blocked. And number three—

and this is a good example here that kind of it's displaying right now—there are a number of allocatable. One, two, three, four, five, six, seven, eight. Eight allocatable. One of the questions—I forgot which number exactly—is to think about whether it's appropriate to allow all eight to be actually allocated and delegated, if you will, and use, whether it should be just only three or two or four that is actually delegated. So these are three questions that this group will need to answer in response to what the LGR generates out.

SARMAD HUSSAIN: Thank you, Edmon. We have Quoc.

QUOC PHAM: Just to add on to that. If Shabaka wasn't a registered IDN and this is the first time the application was submitted, then the term of allocatable is there a variant of the string that you've submitted as an IDN TLD to have published into the root. I guess then there's the point where ICANN or someone needs ICANN to make the decision—apologies if I don't have the history at that—the concept of variant TLDs where the applicant then has to decide which string they want to use to represent the TLD. If this is the case where Shabaka already exists and someone submits row two as a TLD that they want to have applied for then that should be blocked for them, being that regardless if it's allocatable or not, it's a variant of an existing TLD that's live and based on the variant mapping of Arabic LGR. I'm double loading the term here blocked because you can't have something that is canonically unique from a computational sense. Thanks.

SARMAD HUSSAIN: Right. Thank you, Quoc, for bringing that up. We differentiate this—I think it may actually be the next slide. No, it's not. We differentiate this by saying that there's something which is blocked, and then there's something which is withheld for the same entity. So for a different entity, the row number two is actually going to be unavailable, right? But for the same entity, which has number one, it would actually be available. So in a way, number one is being withheld for the same entity but it is not available to other entities, if that is eventually what the policy is to do.

DONNA AUSTIN: Sarmad, I'm going to have to—we're at time.

SARMAD HUSSAIN: Yes.

DONNA AUSTIN: How much more do you think you have to go through? I know how many slides you have, but how much time do you think that would take?

SARMAD HUSSAIN: We can try to wrap it up here. The slides are available if everyone wants to go through the slides. Maybe we can respond to any questions over e-mail and in the next call, if you'd like. But we can actually close it now because we are actually out of time.

DONNA AUSTIN: Okay. So, folks, I think what we'll do, I'll have Sarmad share the full slides on the list. We'll give Sarmad an opportunity to complete this deck at our next meeting. I'm sorry that it's a week away from now. And then I think we'll see if we can include that live demonstration as part of that as well. I know we're kind of loading into the background of this stuff and we're not getting into the meat of the policy discussion. But I think based on the conversation we've had today, this is really important. So if we can finish Sarmad's presentation next week and also include a live demo that people have asked for, then that puts us in pretty good shape to get stuck into the policy questions whenever we meet next. So if there's no objections to that approach, that's the way I'd like to move forward.

Okay. Thanks, everybody, for your participation and staying with this today, and certainly to Sarmad for going through all this work. It can be a lot to take in. I know I'm starting to wrap my head around it maybe. So thanks, everybody, and we'll continue next week.

TERRI AGNEW: Thank you, everyone. Once again the meeting has been adjourned. I will stop the recording and disconnect all remaining lines. Stay well.

[END OF TRANSCRIPTION]