ICANN81 | AGM – Progress on IDN Related Projects for the New gTLD Program: Next Round
Wednesday, November 13, 2024 – 15:00 to 16:00 TRT

PITINAN KOOARMORNPATANA: Hello, and welcome to the Progress on IDN Related Projects for the New gTLD Program. My name is Pitinan Kooarmornpatana, and I am a participant manager for this session. Please note that this session is being recorded and is governed by the ICANN Expected Standards of Behavior and ICANN Community Anti-Harassment Policy.

During this session, questions and comments will only be read aloud if submitted within the Q&A pod. Interpretation for this session will include English, French, and Spanish. If you would like to speak during this session, please raise your hand in Zoom. When called upon, virtual participants will be given permission to unmute in Zoom. Onsite participants will use the physical microphone to speak. Please state your name for the record, the language you will speak if speaking a language other than English, and speak at a reasonable place. With that, I will hand the floor over to Sarmad Hussain. Over to you, Sarmad.

SARMAD HUSSAIN: Thank you, Pitinan. Next slide, please. Next slide. Okay. Just to give you an overview of what we will be covering here, we will start—I guess just to get everybody at the same page, what we will do is we'll start with some core recommendations which come from SubPro and IDN EPDP to give you an idea on what is being recommended from policy for IDNs, specifically at the top level, but also some at the second level. Based on

that, we will share some progress on some of the projects. We'll talk about the Root Zone Label Generation Rules where we are, String Similarity Review Project, and also the LGR Tool Enhancement we are doing, which I guess will eventually be integrated into the application system. Starting with just some background information for everyone. Next slide, please.

Basically, there are a few policy recommendations based on which the IDN implementation is dependent for the next new gTLD round. There's a set of policy recommendations which are part of topic 24, which is related to string similarity, and topic 25, which is related to IDNs, as part of the final report on the new gTLD's Subsequent Procedures Policy Development Process. This was followed by more detailed recommendations through Phase 1 of the final report on the Internationalized Domain Names Expedited Policy Development Process. We normally refer to these two as SubPro and IDN EPDP Phase 1 and Phase 2.

Phase 1 recommendation, most of them have been accepted by the Board based on the scorecards. There are still a few recommendations which are pending, which are related to fees. Then Phase 2 final report has been finalized by the IDN EPDP team, which adds more details at second level registrations mostly in addition to some other recommendations related to IDN implementation guidelines. It was very recently, actually, in the last session approved by the GNSO Council, and I guess is still in the process of adoption. These are some of the policy documents which form the basis of how IDNs will be implemented in the upcoming next gTLD round. Next slide, please.

Just to capture some of the more core recommendations, there are, of course, many more and more detailed recommendations in IDN EPDP Phase 1. We encourage you to actually go and read those documents and also the SubPro document. But starting with, for example, a recommendation in SubPro 25.2, which asks that compliance with Root Zone Label Generation Rules must be required for generation of top-level domain and their variant labels, including the determination of whether the label is blocked or allocatable. IDN EPDP Phase 1 Final Recommendation 1.1 also says that Root Zone LGR is going to be the sole source to calculate variant labels and disposition values for existing gTLDs as well. Basically, what this means is that the new applied-for strings in the next round and their variants and whether those variants are allocatable or blocked will be solely defined by Root Zone LGR, as well as if an existing gTLD would like to apply for variant labels at the top level, those will also be determined by the Root Zone LGR and also whether they are blocked or they are allocatable. Next slide, please.

There was also a recommendation on single character gTLDs. The SubPro recommendation suggested that these may be allowed for ideographic scripts, given that the confusion risks are addressed. Those which were raised by the report by SSAC, for example. IDN EPDP Phase 1 Recommendation 3.17 further clarified that in the context of ideographic script, currently the scripts which meet that criteria is the Han script which is used to write Chinese, Japanese, and Korean languages. Basically, they said that the recommendation suggested that as part of the implementation, ICANN reaches out to the Chinese, Japanese, and Korean communities, their Generation Panels and

basically look into what would be the mechanism to implement single character Han.

Based on that, we actually had two public comments. One was there was a statement by Chinese, Japanese, and Korean Generation Panels. ICANN followed up as part of implementing this particular recommendation with those Generation Panels. They published their statements which we took for public comment. That was the first instance where we received public comments from the broader community. Then more recently, we also published the IDN module of AGB for public comment. That also referred to the single character Han based on the recommendation. We've also received additional public comments through that second public comment which was more recently closed. There has been quite a significant amount of feedback on this. What we are doing now is we are collating that input from the Chinese community and other communities and generally larger community which has provided feedback on this topic. We plan to take it back to, first, the subtrack of IRT for IDN EPDP and then to the full IRT later in November and early December and discuss with them on what would be the next steps on this topic. Next slide, please.

Also, in SubPro, there's a recommendation which says that the IDN gTLDs which are identified as variants of each other can only be applied-for by the same entity, meaning the same registry operator. In addition to that, the variant should be managed by the same backend registry service provider. Next slide, please.

Basically, there are also some recommendations which go at the second level registrations. The first one, which is 25.6 from SubPro,

basically, talks about that if a second level label is registered under a TLD and its variant TLD, then the label must be registered either by the same registrant or should be withheld to be registered by the same registrant. Next slide, please.

This is then extended further in Recommendation 25.7 of SubPro, which actually says that not only the same label under the variant TLDs should be registered by the same registrant, but also variant labels of the same label at the second level should also be registered by the same registrants. Next slide, please.

Finally, this Recommendation 25.8 says that even though these domain names would actually be registered by the same registrant, it does not mean that they have to behave or have to act the same way, meaning that these domain names can be used in different ways by the registrant. One reason being that one of these variants could actually be in one language and another variant actually could be in another language, which may actually even cross geographies. Therefore, these can actually be used differently by the same registrant because it is serving different audience, possibly.

More recently, we've actually also published some draft modules which are related to IDN topics as I shared. They're actually closed public comment. But in any case, if you want to see those drafts, there is a module on String Similarity Review which was published. There was a draft on Internationalized Domain Names and a draft on Root Zone Label Generation Rules. So these are three different modules within the Applicant Guidebook and drafts, for these were already discussed with the IRT and have recently published for public comment. We'll be

obviously taking input, which was provided through the public comment and updating these drafts. These will then go through another second public comment process later next year. At that time, I guess you will have another chance of reviewing them. With that, let me hand it over to Pitinan to talk about Root Zone LGR.

PITINAN KOOARMORNPATANA: Thank you, Sarmad. This is Pitinan. On the next topic, the updates on the Root Zone LGR project. As earlier mentioned, that by the policy, the Root Zone LGR will be used for determine the applied-for string and also its variant string. The current version that we have is version 5 published in May 2022. Currently, we include 26 scripts as listed here. Many of you in this room are also involved in the project. Thank you very much for that. This work actually involved more than 10,000 hours of volunteer works from 43 countries, 270 plus members. Now, even though we have 26 scripts, one script can cover multiple languages. We cover, as per the document, 386 plus languages. We hope that this cover most of the languages used in the modern time. But on the updating side, I see hand from Hadia. Maybe, Hadia, if I can hold to the end of this section and then I'll call upon you. Recently, we have the Thaana Script Generation Panel recently formed in September 2024. With this Thaana script, they are still working on it but this is how it recently looks like. They have about 25 consonants, 11 vowels, another big repertoire. What to be included or excluded still being discussed. This is something the work going on and potentially coming up in version 6 as per the timeline by the Thaana Script Generation Panel. Since the formation in September, since then there's been working on finalizing the draft.

Because they already have some work going on at the second level, some solution's already been there. Based on that, they modify to be the Root Zone LGR. Tentatively by January next year, the draft should be published for the public comment. By Q1, it should be finalized, and by Q2, we plan to integrate it into the Root Zone LGR version 6. This is something as planned. If it's integrated, it will be the base for the next new gTLD round.

With that, I take the hands from Hadia. Hadia, would you like to unmute yourself? Please, go ahead. Hadia, can you unmute yourself? Okay. Maybe if you could type your question in the chat also or come back later, that would work, I guess. Let's move on. I'll hand it over back to Sarmad for the next session. Thank you.

SARMAD HUSSAIN: Thank you, Pitinan. Let's talk about String Similarity Review. Next slide, please. Basically, SubPro reaffirmed that strings must not be confusingly similar to an existing top-level domain or a Reserved Name and that the criteria for similarity needs to be visual. The wording basically says that the visual similarity has to be not just possible but probable for this confusability. There have been actually more additional details which have been added to these affirmations in SubPro where IDN EPDP Phase 1 suggests that, first of all, identifies the scope of comparisons. We'll talk about that in a bit in the next slide. Then also states that the decision has to be done eventually manually by a String Similarity Review Panel. There was a potentially a tool used earlier but it would not, of course, be used in the next round. It also states that even though there are now many more comparisons, the

String Similarity Review Panel may actually follow some guidelines to omit or even include some comparisons. Those deviances must follow some guidelines and then will also be documented. Again, this is just a paraphrasing of the recommendations. We encourage you to actually read the actual policy recommendation language in the actual documents.

As far as the string similarity comparisons are concerned, this is a summary of what would be compared with what else. You have the applied-for gTLD string at the top there on the right. That applied-for gTLD string, that is actually represents in some ways a set. It includes the primary gTLD string which is applied-for. It also includes all the allocatable variant strings of that primary string. Then it includes all the blocked variant strings which are generated from that primary string. The primary string is your starting point. That's a very important point to note. You'd use the primary string and the Root Zone LGR to determine what is the allocatable variant strings and what are the blocked variant strings. That is how it would work. Please note that the middle column allocatable variant string does not distinguish the applied-for allocatable variant strings and the not applied-for. All allocatable strings generated are, I guess, in scope here.

These then will be compared against whole set of strings for string similarity. If you look at the first column on the left here, these will be compared with the existing gTLD strings. These will also be compared with the strings which have gTLD strings which have been applied for in the previous rounds and still are in process. They're not delegated yet but they were already applied for in one of the previous rounds. They

will be also compared with the existing country code top-level domains or ccTLDs and also any requested IDN ccTLDs which are in process. They will also be compared with all the other applied-for gTLD strings in the same round. They will be also compared with Reserved Names, blocked names and any two-character ASCII which are, of course, reserved for by ISO 3166. Not only these strings but their primary. These will be considered the primary strings. Then using Root Zone LGR, these primary strings will generate the allocatable variant strings and then also blocked variant strings. For each string which is applied for and its allocatable and blocked variants, they will be compared against all these other types of strings, primary and their allocatable and blocked variants for String Similarity Review. The only comparisons which will not be made would be the blocked variants of the applied-for string against the blocked variants of all these types of strings on the left-hand side. That's a no, so that is out of scope, that comparison. All the other comparisons are going to be made.

Here you also see that there are some gray areas where you have a yes. These gray areas are comparisons between blocked variants and two allocatable or primary variants of others. These gray areas is where the String Similarity Review Panel will have some level of decision-making that they can omit some of these comparisons if they have a good reason and they are following the guidelines. But the white areas where they are the central, where it says yes in the white blocks, those are comparisons which the String Similarity Review Panel obviously must do.

This is the scope of the String Similarity Review. It is part of the manual process. You can probably see that this will create a lot of comparisons because you're not just comparing applied-for strings against existing strings, but you're also comparing their variant labels against variant labels of the existing strings as well. This will likely create thousands of comparisons. Before we move on on how we are trying to address this in a way that this can actually be managed, let me stop here and see if there are any questions on the scope of comparisons. Okay. We don't have any questions. Then let's move on.

We've actually developed an initial set of guidelines for String Similarity Review. We put it through public comment process and received comments. We will obviously integrate those comments and bring these guidelines back for public comment review for a second review by the community. We'll obviously develop it as we go through this project. But as you realize that because of all these variant strings which are now part of this whole process, the number of comparisons has increased very significantly. To assist the String Similarity Review Panel to go through these comparisons, what we are suggesting is a two-step process. There is a first step which will largely do some pre-screening. For example, there may actually be a label in Arabic script which may not actually be confusable with another label in Chinese script, Han script. Maybe that doesn't need to be manually reviewed, but then there may be actually two different labels within Latin script or a label between Latin script and one label between Cyrillic script or some other scripts which are related. There may be potential confusability and that is something which may need to be reviewed by expert panel for String Similarity Review.

What we do is in the first step, we do some pre-screening. What we're doing is we are collecting data from script experts on string similarity. We'll go into that in a little more detail. Also, developing a tool which will extend the variant mechanism. It's going to use the Label Generation Rule format, RFC 7940. We'll use the same mechanism, but we'll extend to include similarity data as well. The tool and the data will be able to calculate what are potentially or possibly confusable labels, which are not as confusable labels. That pre-screening data will then be given to the String Similarity Review Panel to actually process it further. The final review is actually still going to be manual and by the String Similarity Review Panel. So nothing changes from process wise, but this is just to make sure that the scope can be managed.

Basically, then we have two speakers here. We have Michael Bauland from Knipp who is going to talk about the work which is going on data collection from script experts. And we also have Julien Bernard who's going to talk about a little more on the String Similarity Review Tool work which is going on. With that, let me pass it to Michael to share more details on the data collection.

MICHAEL BAULAND: Thank you, Sarmad. As mentioned by Pitinan before, the Root Zone LGR in the current version covers 26 scripts and all repertoires of all these scripts need to get analyzed. This covers a lot of languages and code points and there's obviously no single person capable of knowing all the details of all these scripts. That's why the decision was taken to delegate the task for each of those scripts to experts within the respective scripts. For this, each of those experts has been assigned a

scope of work to analyze their repertoires. This basically consists of four items. First, all of the repertoire characters need to be compared to ASCII characters, both the small and capital versions, A to Z. Then there's, of course, the in-script comparisons. This means to compare all elements of the repertoire to each other. Some of the scripts have built a script family with related scripts for those. Also some cross script comparisons need to be done. Finally, there are some miscellaneous comparisons that do not apply to all of the scripts, but for example, the capitalized elements in the scripts need to be compared to each other and to the lowercase ones, at least, for the scripts that have capitalizations. Same for conjuncts and compounded characters. All scripts also need to look at simple shapes like circles, lines, curves, which may occur in any script. Finally, there's also the need to look at underlined versions of the elements, because usually in URL, domain names are displayed with an underline and this may hide some part of the elements. Next slide, please.

For this categorization, we have created five categories. All pairs need to be checked in which of those categories they belong. There's Category 1, which are the identical or near identical cases. This also contains the already defined variants. In the following slides, I will go into more details for each of the categories and provide some examples. Then we have Category 2, which are the highly confusable cases. Those are strongly similar or near homographs. Category 3 are the similar, Category 4 are the distantly similar, and Category 5 are the distinct characters. Next slide, please.

For Category 1, by default, all existing variants that are already part of the Root Zone LGR belong into this category. This is independent of their actual visual appearance. This is really all the variants. For the further categorization, we only look at visual similarity and not any other kind of similarity, like meaning or something like that. Then we have, within this category, the cases of identical characters. Most of these should already have been defined as variants by the different Generation Panels, but some might be missing due to the extended scope of the review. Some other cases will likely get introduced, for example, due to uppercase form. If you look at the Latin E and the Greek Epsilon, they may be slightly similar, but definitely not really confusable in the lower-case form. But if you look at their respective uppercase form, which have the code points 0045 and 0395, these are actual identical, and therefore, these definitely belong to Category 1.

Another case would be the near identical cases. Those are easy to confuse if looked at side-by-side. One example would be the underlining case if you have the E and the E with the dot below. Unfortunately, on this slide, the underline is missing. In the second part, both E and E with dot below should have been shown with an underline. If that's the case, also depending on the font, you would not be able to see a difference between the E and the E with dot below, because the dot below would fully be covered by the underline. For that reason, this pair should also belong to Category 1. Next slide, please.

Then we have Category 2, which are the highly confusable. This means that they need to be confused by a large majority of attentive users when not looking at the labels side-by-side. One example would be the

E with an accent. One from top left to lower right, and the other one from lower left to top right. When you look at them side-by-side, you definitely see that they are different. If you just have one label in mind, for example, example .copy, and then see the other one, a large majority of users would be confused by this and would not recognize that the accent is in the other direction. Maybe except French speaking. They would recognize that. Other examples from other scripts are the ones listed in the second and third example. They also have been categorized into two by the respective experts. Next slide, please.

Then Category 3, this contains the similar cases, which are confused by a simple majority, but not a large majority of users. Here are also some examples that have been classified as Category 3 by the respective script experts. You can see them in the three examples here. One other case of Category 3 similarity is the fallback replacement, which is mainly used for the Latin script. Because quite often, if you have a word which has a character with a diacritic, as a domain name, you might tend to write it in a way without the diacritic, that you only need to use ASCII characters and don't need to use the Punycode algorithm to convert them to a dash-dash version. All of these cases should also be Category 3. Next slide, please.

Then we have Category 4. Those would be confused by a minority of attentive users or maybe not even confused at all. The experts are not actively looking for these cases because those would be quite a big set of characters to include here. But whenever a pair has been looked at and then the decision was that it's not Category 1 to 3, it was requested to still put it into the data and write it down as Category 4 that it's only

consistently similar. The reason for this is so that when the data is looked at by other people, for example, in the public comment, that they can see that this comparison has not been forgotten but it has been looked at and it was a conscious decision to not put it into the Categories 1 to 3. Some examples are again provided here in the three lines by different script experts. Next slide, please.

Finally, we have the vast majority of comparisons that would belong to Category 5, namely distinct. Actually, you can almost use any random pair of elements and almost all of them will be distinct as you see here in this example. There's definitely no need to look at them and also no need to write them down as this case is really not important for the String Similarity Review. Next slide, please.

What's the current status of this data gathering project? For eight scripts, the experts are still working on the initial data. This includes mainly the largest strips like Han, which have a really huge repertoire and that needs a bit more time to look through all the cases. Then we have 16 scripts, whether experts already have submitted that data and the data is now currently in an internal review process, the review is in progress. And for two scripts, we are already ready. The review has been completed and the data is ready. What will be done with the data is best explained by Julien. I hand over to him for the next part of the presentation.

JULIEN BERNARD: Thank you, Michael. I'm Julien Bernard, IT consultant from COFOMO. I do a quick overview of the string similarity tool. This tool will take as

input the list of existing TLDs, gTLD, ccTLD, the Reserved Names and blocked names—what we saw in the table that was shown by Sarmad. Also, the list of applied-for strings and the similarity data collection. That means the report we got from Michael and also the LGRs, Root Zone and the Reference LGRs. The tool will take this similarity data collection and calculate if any of the applied-for strings, existing TLDs, Reserved Names or blocked names are the same or similar to each other also to their variant labels. Then it will output a report with the potential contention set for the review panel. Sarmad, over to you.

SARMAD HUSSAIN: Thank you. I think one detail which we should probably add here is that in the five levels which were being presented by Michael, we are going to use the first three levels to shortlist the data for potential contention sets to be finalized by the panel. Levels 4 and Level 5 will not be considered so that we align with the recommendation which asks for probable visual confusion, not just a possible visual confusion. Next slide, please.

As you can probably see, the work on data collection and tool development is ongoing from last quarter and also in this quarter. We hope to have the data work completed early next year. We'll obviously share update on the tool as well as the data at ICANN82 as well and would have the tool finalized along with data incorporated and tested. We're targeting June 2025 to reach that milestone. I think that's the final slide for this section. Let's see if there are any questions on string similarity.

NABIL BENAMAR: Thank you, Sarmad, thank you, Michael, for the excellent presentations. This is Nabil Benamar for the record. I have seen that we are basing our study on the visual standard definition of similarity. I was asking myself if there is any standard in this regard, scientific standard, meaning that there will be no interpretation from different interested people in the topic on this similarity and this confusion. My thing is that this is not an exact size, so it depends on the user experience, meaning that each user can put any label in one of the categories that we have mentioned here, especially the first three ones. When you go, for example, to the Category 4, you are basing the definition on the attentive users. How can we guarantee that a user is always attentive? This might not be guaranteed every time and in every experience. I think these are some challenges. I'd like to know your feedback on this about the work that has been so far done in different scripts. Thank you.

SARMAD HUSSAIN: Thank you, Nabil. This is, I think, one of the things which was covered in the draft guidelines we published for String Similarity Review. One of the questions we had debated as part of that process in those guidelines or before those guidelines were shared with the community was that are we looking at people, for example, who are familiar with the script or people who are not familiar with the script? There are also questions about what you have raised, whether people are actually paying attention to what is written versus you're somebody just glancing through and skimming it and not really paying attention to those details. What we proposed in the guidelines which we published

earlier was that we are looking at similarity based on readers of that script, as well as we are also basing our scope for people who are not reading it, just skimming it or reading it very casually, but actually paying reasonable attention to reading a domain name. I think that follows from—one of the reasons we were saying that is because the policy recommendation from the PDPs is quite clear that we shouldn't just look at possibility of similarity, but there has to be a probability of similarity. Those string similarity criteria are there, it needs to be somewhat conservative. If you're just looking at skim level, there is certainly a possibility of confusability, but we would want to keep it at a level where people are actually paying attention and still getting confused that would maybe raise to the probability of confusion. Thank you.

NABIL BENAMAR: Thank you. This might be more present in some other languages or specific scripts because there is a difference between these categories of scripts that we have, that we are studying right now. If you take a look at the Arabic script and you are familiar with the Arabic script, we have the dots, we are using the dots. The dots can be confusable with small hamza or small symbol that is used in some languages within the Arabic script community. That creates much more challenge for these scripts. I don't know of the others, but for Latin scripts, for example, this example is not there. Only few characters use the dots, for example.

SARMAD HUSSAIN: Thank you, Nabil. I think that's why we're involving the local experts who use the script to give that feedback to us. We're conscious of time as well, so if you allow, I think we'll go to the next topic, and once we'll come back and take more questions after we've gone through the next topic. Maybe we can move to Pitinan first and then take more questions.

PITINAN KOOARMORNPATANA: Before that, maybe Michael has some—

MICHAEL BAULAND: Just a quick comment to what Nabil said. One important point to remember is also that the whole data and process that is developed here is just to help the actual String Similarity Review team to do their work. These decisions that are taken here are not exact decisions, but the String Similarity Review Team will do the final decision. This is just to help make their work easier. We are not defining clear-cut decisions here. Thank you.

PITINAN KOOARMORNPATANA: Thank you. For the next topic is the Label Generation Rules Tool Enhancement, we call it LGR tool. As a background, currently, we have these two available online and it has provided three modes. The basic mode is you can go in and verify whether the labels you interest. Is it valid as per reference LGR or a Root Zone LGR? This is the basic mode that you can use. The second functionality of it is to review the IDN tables. The registry operators can use these two to compare the IDN table against the Reference LGR or the Root Zone LGR. Then the last

mode is the advance mode, which is the user is the expert who developed the LGR itself. You can use the tool to create a new LGR, do some comparison, modification and so on.

This is the current of the tool. This tool is available in this link, lgrtool.icann.org. If you have ICANN account, the same ones that you use for meeting registration, you would be able to access the tool. For the technical community, this tool also is available in the Open Source. The source code is available on this link as well on the GitHub. The manual is available in the link.

In relation to the next round, we plan to enhance the tool to also include this functionality. First, you will be able to verify a primary label, which also includes the ASCII and IDNs. If it's the ASCII or is it coming as a label, we will check whether this is compiled with the protocol requirement. Is it the string as per the IDNA2008 if it's the IDN? And also is the string valid as per the Root Zone LGR? It will also check the length requirement. As mentioned earlier, we may have a slightly different requirement on the length of the labels for Han script. We separately check that for now as well. Then we will also check whether this primary label collide with the existing TLDs or is variant set. We also check against the Reserved Names and also the variant of the Reserved Names. The functions are separate because if it's the Reserved Name but you meet some criteria, this string may be able to move on. Also lastly, we'll check whether this collide with the blocked names or the variants of the blocked name, which will certainly be blocked.

That's the functionality for primary labels. We will also have an actual function to check the variant labels. It will do all the checks the same as

the primary labels above. It will also check whether this is an allocatable variant of a primary label. You will have to also input what is the primary you want to check against. This is the plan on the tool that we plan to enhance before the next round. This will be available for the community to use before that. That's about it for me. Over back to you, Sarmad. Thank you.

SARMAD HUSSAIN: Thank you. One more thing about the string similarity data. Once we have the final data and it's reviewed internally for any omissions or other syntactic details, we do intend to publish it for community review. You will see that data come through and we certainly look forward to your input on the analysis which has been done by the script experts, for whether you agree or not agree with the categorization of the script experts on not just the similarity but the level of similarity. You will see that come through soon. We look forward to your input into that process to finalize that as well.

Also, this similarity data is obviously at a character or a character sequence level. Sometimes there can be ambiguity at a label level which may or may not be addressed at a character level sequence or character level analysis. Of course, those are things which the String Similarity Review Panel will be looking at in addition to obviously the pre-screening tool. Pre-screening tool has a very limited scope and just generating potential cases for String Similarity Review Panel to see in more detail. But the final decision making will really be done through a manual process through the String Similarity Review Panel. We have

time for one or two more questions. I have Anil in the queue. Anil, please.

ANIL JAIN: Thank you, Sarmad. Question for Pitinan. The LGR testing, what you have demonstrated is wonderful. I think we are able to cover all aspects, A to Z. This will be quite useful for the new gTLD round also. But my question here is that whether this tool is available for ccTLD and gTLD together or it is only designed for new gTLD? This is question number one.

Question number two is that in case registry is not able to figure out the LGR testing as we are talking, in that situation, is there any plan for helpline from ICANN Org to assist the registry to check their desired LGR? Thank you.

SARMAD HUSSAIN: Thank you, Anil. I think IDN ccPDP4 Policy currently is in process of review and approval. Once that policy comes to implementation, we can certainly look into making such a tool if that's useful for the ccTLD community to make that available as well. Currently, the focus of this tool is primarily for the next round of gTLDs. That was the first question.

PITINAN KOOARMORNPATANA: Pitinan here to add on that. The existing TLDs, that will include both G and CC as well.

SARMAD HUSSAIN: There was a second part of your question, Anil. Could you please—

ANIL JAIN: The second part was whether in case with these directions, the registry operator or manager is not able to figure out how to check their LGR before applying a new gTLD, is there any plan to have a helpline available from ICANN Org to assist such kind of difficulties?

SARMAD HUSSAIN: Thank you. I think that there will be support available for the next round of new gTLDs through ICANN. I think the channel which will be made available, queries about the tool could also go there. I would assume the answer is yes. This would be managed directly for the next round program.

ANIL JAIN: Thank you, Sarmad. Thank you, Pitinan.

SARMAD HUSSAIN: Okay. Any other questions or comments? Yes, please?

ASHLEY ROBERTS: Hello. My name is Ashley Roberts. Just thinking about this from a new gTLD applicants' perspective. There's obviously a lot of complex stuff here. I'm just thinking about how an applicant would navigate this in terms of trying to determine whether they may have a string similarity issue. The String Similarity Tool you've been talking about that you're

working on developing at the moment, what exactly does that look like? For example, what would the output of that tool be?

SARMAD HUSSAIN: If you can go back to that slide, Pitinan, on which Julien was presenting. Maybe Julien, you want to address that. I guess the question is what would be the output of this tool will be?

JULIEN BERNARD: This will be the contention set. You will have the applied-for TLD along with any similar existing name, blocked name or TLD along as well as the [score] that would give a level of confusion. That's what we expect to give to the Review Panel here.

ASHLEY ROBERTS: Thank you. Can I just clarify, is that tool just for use by the Review Panel or is it available for applicants as well?

JULIEN BERNARD: At this time, we are developing this tool for the Review Panel. But I think if the community thinks it may be useful, we can certainly look into seeing whether that can be made available more broadly as well. But the current focus, of course, is to help with this pre-screening for Review Panel to do their review.

ASHLEY ROBERTS: Thank you. I think just due to the complexity of this, any tools or additional guidance that you can provide to applicants to help them navigate this would be greatly appreciated. I mean, I've had a bit of a look at some of this stuff already. I've looked at the guidelines document which I've personally found it quite hard to get through. It's quite a challenging document. Just have a bit of concern about how applicants, particularly people that are very new to this space are going to be able to navigate this in an efficient way.

SARMAD HUSSAIN: Sure. We can certainly take that feedback back and see what we can do to make tools like this available. Thank you. We are overtime so I think we should probably close the session. Thank you for attending. If there are any more questions, we are here to answer after this session. Otherwise, please feel free to send your queries to idnprogram@icann.org and we are more than happy to take your questions through e-mail as well. We'll be available here in case you have any follow-up questions. Thank you. We can close the session. Thank you.

[END OF TRANSCRIPTION]