## ICANN Transcription

## IDNs EPDP

## Thursday, 26 August 2022 at 13:30 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/LgYVD

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

DEVAN REED:                    Good morning, good afternoon, and good evening. Welcome to the IDNs EPDP call taking place on Friday, 26 August 2022 at 13:30 UTC.

In the interest of time, there will be no roll call. Attendance will be taken by the Zoom Room. We do have apologies from the Lianna Galstyan, Farell Folly, and Joseph Yee. All members and participants will be promoted to panelists for today's call. Members and participants, when using the chat, please select everyone in order for everyone to see the chat and so it's captured in the recording. Observers will remain as an attendee and will have view-only chat access.

Statements of Interest must be kept up to date. If anyone has any updates to share, please raise your hand or speak up now. If you

need assistance updating your Statements of Interest, please e-mail the GNSO secretariat.

All documentation and information can be found on the IDNs EPDP wiki space. Recordings will be posted shortly after the end of the call. Please remember to state your name before speaking for the transcript. As a reminder, those who take part in the ICANN multistakeholder process are to comply with the Expected Standards of Behavior. Thank you and back over to you, Donna, to begin.

DONNA AUSTIN:    Thanks, Devan, and welcome, everybody, to our IDN EPDP call. We're 24 hours late, I guess, so we might have a slightly lower attendance, but I think things seem to be picking up.

So just a few updates. The deadline for the group three draft text is the 31$^{st}$ of August. So, hopefully folks are having the opportunity to work their way through that. I will also acknowledge that Dennis has sent some text back. There's still some concerns around group two text. So I had a bit of back and forth on that. We're going to have to come back to that at some point.

I guess the only other thing is we're getting three weeks from the ICANN meeting. I think the leadership team—and correct me if I'm wrong—but I think we've decided not to have a call the week before and the week after. Just recognition that folks are traveling and people need a week to recover after they've been to an ICANN meeting. So I've decided the meeting we'll take some time out.

To the meeting itself, we're starting to do some preparation. I know that the agenda hasn't been published yet. But just a reminder that we have between 9 and 12 on the Saturday morning, that's when we'll be meeting. We're doing some thinking about what's the best thing for us to discuss, given we'll be face to face for the very first time—oh no, for the very first time just for Justine and I, for the very first time with group. So we're doing some thinking about what we'll discuss in KL, but I think what we're leaning towards is Steve and Ariel have put together kind of process flow, I guess. Is it a process flow, is it a framework? But it's a document and what we're hoping to do is test some of the recommendations that we've already made. So it might be a little bit different from what the CCs did with their stress testing. But just to get a little bit of a better handle on some of our recommendations and whether they made sense or are there other things that we haven't thought of with them. So that's kind of the plan at the moment for ICANN75. But I think there'll be more on that as we get closer to the time.

So what we're going to come back to today is the string similarity small group output. Ariel is going to take us through this. But I know that there were some concerns that were raised on our call last week. But what we're hoping to get to by the end of this call is how we feel about the small group output and is there another path. So I would strongly encourage those that were part of the small group discussion to—I don't want to use the word "defend" but that's what I'm going to ask you to do—contribute to the discussion by really explaining the intent from the small group discussion output so that those in participate understand exactly where you're coming from, and why the recommendations that

you've come up with, why you think they're important. Because I think it's really important for the rest of this group to understand that to be able to make a decision about what path we're going to go down with string similarity. So with that, I'm going to hand over to Ariel.

ARIEL LIANG: Thanks, Donna. So we developed these slides with the hope to better illustrate the rationale why the hybrid model is being recommended. We definitely welcome small team members to chime in and provide additional inputs and, I guess, arguments for supporting that. So this is based on staff's understanding of the rationale. And we also included some additional information. Hopefully, it will be helpful.

Before I talk about the rationale, this is just a quick refresher what the hybrid model means. In paper, it means it's a mixed level approach between Level 2 and Level 3. So basically, all kinds of variants are taking into consideration in the string similarity review that includes the primary applied-for labels, the allocatable label. That's all of them, and all of the blocked variants. But what is not being compared is basically between blocked variants and blocked variants, but everything else is being compared against each other. So that's the gist of that, and I hope that folks do remember how it works when we have these Chinese and Arabic examples that were shown in the previous meetings.

There's some relevant background for your consideration, not the hybrid model, even though it wasn't explicitly discussed during small team meetings, but it was kind of overarching information at

# EN

the back of the mind of the members when they discuss why such a conservative approach needs to be adopted. So we collected some of this information with the help of Sarmad. He pointed us to some historical document that basically has this overarching theme that root zone needs to apply restrictive rules and to mitigate confusion as much as possible, because it is a very, very important zone. So I'll just show you some of these references and documents.

The first one is RFC 5891. It's actually also part of the IDNA2008 standard. So what it says is that any domain name registry, including that of the root zone, should develop and apply additional restrictions as needed to reduce confusion and other problems.

Then the second reference is RFC 6921. It says zones higher in the DNS tree tend to have more restrictive rules. And the reason is that the root zone serves the entire Internet population. So that's why the restrictive rules need to apply, and then we really need to be very rigorous in terms of what can be delegated to the root zone, because it will have a much greater impact to the entire Internet population.

The third reference is the SAC089. It's SSAC advice. It says confusability cannot be considered in isolation from other issues related to security. Phishing and other social engineering attacks based on domain name confusion are a security problem for end users. So it just emphasizes that confusability could cause a greater impact to the security and stability of the Internet. So that's why. Especially for the root zone confusability needs to be mitigated as much as possible.

The last reference is from the staff paper on IDN implementation. It says the implementation of variants must be done in a way that the operation and maintenance of the DNS not be adversely impacted by the introduction of variants. In addition, it should avoid including variant TLDs in a manner that would create user vulnerabilities or a probability of confusion.

Again, all these references and background information kind of reemphasize the importance of mitigating confusion and applying restricted rules in terms of what can be delegated to the root zone. So I think all these backgrounds serve as a kind of argument why hybrid model is being put forward by the small team due to the consideration of the mitigating confusability and applying additional restrictions.

So this is a quick recap of the EPDP team's discussion of the small team's recommendation. So what we have heard so far is that the EPDP team had generous support for comparing the primary labels and the requested allocatable variant labels. So these are pretty straightforward and not much controversy there. But what we also heard is that some members expressed reservations about including non-requested allocatable variants, as well as the blocked variant labels in the string similarity review.

What we want you to remind the folks that the reason why the String Similarity small team recommends to include all these types of variants in the consideration of the string similarity review is that this will help meet the singular goal of risk mitigation of the failure modes. One is denial of service, and two is misconnection. So denial of service and misconnection, these failure modes will likely be caused by confusability of the top-level variants. So that's why

this hybrid model will do the best work in terms of meeting that objective. That's why that's being put forward. Then another consideration of the small team is that they believe string similarity review is a great opportunity to mitigate those risks and mitigates confusability as much as possible. So that's why the hybrid model is a good vehicle of meeting that goal.

But the last point I want to remind folks is that the small team did not consider the implementation complexity of the hybrid model. That's basically one of the key concerns from the EPDP team members when they express reservation about including the non-requested allocatable variants and blocked ones, because implementation-wise, it seems to be very complex. So that's why we will have a next step for the EPDP team to deliberate and discuss the implementation side of it.

So that's a quick recap of the discussion so far. So I think also during our discussion, I think we didn't talk more in detail about the two types of failure modes and the risks associated with that. So here I just want to provide some examples and illustrations to help folks better understand what to these two types of failure mode means and their potential consequences.

So the first type is denial of service. I think it's first defined in one of the SSAC advice, it's SAC060. What it means is that a user attempts to visit, for example, example.x, that's a website. But then the user read it as being the same as example.y. That could happen if the user saw example.y advertisements somewhere. So when the user attempts to type the address of example.x, the connection does not work because example.x is not registered. So that means denial of service in this context.

So if we look at some real world example, on the left it shows there's a web address, art dot [Qua Del]. [Qua Del] is an artist name in Chinese. So it's been posted on a bus as an advertisement to promote this artist's website. Then there's this lady, she saw the domain being promoted on the bus and then thought, "Oh, this artist [Quan Yel] has a website. I should check it out." Because where she saw [Qua Del], she thought it was actually [Quan Yel]. It's another artist or different label. It just maybe she doesn't have a good sight or she just mistakenly thought the one being promoted on the bus is a different label. She just had this confusion in her mind. And then when she went back home to type [http://art.quanyel], a 404 page popped on her screen because there's no such website thing registered. That's basically denial of service in this example, and then she got kind of frustrated and confused. So that's the consequence.

I guess if you talk about the consequence of denial service, it's not much actual harm that will cause to the user, but the user will experience confusion and frustration when they thought they saw this web address. They type it in their browser, but then nothing shows up, nothing worked. That's a consequence of that.

I saw Maxim has his hand up but I'm wondering whether I should continue and at least go through these examples before we open the floor. What do you think, Donna?

DONNA AUSTIN:          Ariel, why don't we have a conversation around denial of service? Because Maxim has his hand up and Dennis has a question, but I

think it might be helpful if we just have a conversation on denial of service now. We actually do that.

MAXIM ALZOBA:          Do you hear me?

DONNA AUSTIN:          Yes, Maxim.

MAXIM ALZOBA:          I suggest we use something like "no service provided" or "no expected service provided" because denial of service is a term describing the attack on servers when malicious caught requests service, and then next packet it requests to stop provision of the service. It's a well established term and I suggest we do not cause confusion by using it. So something like "no expected service provided" or some other description of the situation. Thanks.

DONNA AUSTIN:          Thanks, Maxim. I must admit, that was my initial reaction too. But Maxim has said it is a term that's used in SAC report. So I guess for the purposes of this group, what we mean by denial of service is the end user is not getting the result that it thinks it's going to get. So it's either for a full message or maybe it's something a little bit further than that as well. I appreciate that we try not to confuse ourselves here, Maxim. So in an attempt to do that, we're calling out what we mean by denial of service in this context. Dennis, go ahead.

DENNIS TAN: Thank you, Donna. Yeah, I agree. From a practical standpoint, let's agree that the use SAC060, denial of service, meaning we all know that has another context in the cybersecurity context. But let's keep it practical, I guess.

I do have a more meaningful, substantive comment, but I think it's coming up on misconnection. So I'll wait for Ariel to go through the example, and I'll just come back to the queue. Thank you.

ARIEL LIANG: Sounds good. Thanks, Dennis.

DONNA AUSTIN: Sorry. I was on mute. Go ahead, Ariel. Sorry.

ARIEL LIANG: Okay. Thanks, Dennis, and thanks, Maxim, as well. One thing I just want to mention and also to respond to Dennis's comment, [Quan Yel and Qua Del], they're not variants. They're completely different things. But they look very similar. If you look at the second Chinese character, they just have a minor difference. So that kind of confusion could happen. So that's why we use this example. It's also one of the examples developed by the small team.

So moving on to misconnection. In the SAC060, it defines misconnection as a user attempts to visit example.x but read it as being the same as example.y. For example, this user saw in an

advertisement. But then after clicking on example.y, the user arrives at a site controlled by a registrant different to example.x. Because the user had some expectation of visiting example.x, but then eventually arrived at example.y, and then that's misconnection.

So to illustrate it, we're using the same scenario. It's this lady saw a website, art dot [Qua Del] being advertised on a bus. And there she was thinking, "Oh, the artist [Quan Yel] has a website." Again, she's making this confusion already in her mind. She thought it's another artist's website. She was thinking, "Oh, maybe this artist [Quan Yel] is selling art online." And then she just had this kind of thought in her mind. When she went back home and then received an e-mail to her mailbox, it's actually an e-mail about art dot [Qua Del]. It's the one that's being advertised on the bus. But she still has this kind of misconnection in her mind. She still thought it's a website for [Quan Yel]. So she clicked on the URL, and then tried to check out the website. This website is also for an artist but it's [Qua Del] and it's selling art for this different artist. But then this lady just had this misconnection already in her mind, and then she thought, "Oh, it even look a little bit different. I should still try to support [Quan Yel]," the other artists that she's a fan of, and then put her credit card information and try to buy some art from this art dot [Qua Del] website.

So this whole thing is about misconnection. So she had in her mind for something, but then when she clicked it, there's something else. But then her brain just didn't make the switch, and then she continued some actions on the different website. Then eventually, she's doing something maybe she shouldn't do is

# EN

paying money for art that she wasn't even a fan of. So that's a misconnection example. I know Dennis has more comment about this misconnection scenario. So I will stop here.

DONNA AUSTIN:          Go ahead, Dennis.

DONNA AUSTIN:          Thank you, Donna. Thank you, Ariel, for going through the example. We'll talk about misconnection. But I want to offer a different angle of it. Let me just start with setting up the context. So we're talking about this failure modes discussed in SAC060, Recommendation 7. Edmon put a link to the chat so everybody can look at the exact language. So the first thing about Recommendation 7 is that this is in the context of a variant set. Take your example, variants at the second level or variants at the top level, but a variant set nevertheless. SAC060 discusses these two failure modes in that context. What if a variant does not exist or what if a variant is allocated to a different entity that sets up a different user experience. Going through these examples, we are using those out the context of a variant. So crossing namespaces, if you will. That to me, we are using the recommendation setting and put it in a different environment or context, and that would be non-precedent.

I get, I understand what is being discussed and explained here. But I think—and I want us to be very cautious as to what type of conversation we're having here because it's time to shape as—we want to solve for misconnection across namespaces, which we

know happens every day. But that's a much larger conversation than just IDN variants. That's a different type of issue, and again, it will be unprecedented conversation in how to solve that. How do you even start solving for misconnection across namespaces?

So really, I want us to be very conscious as to how we are using these failure modes outside of variants. I think that was my area of question where these two TLDs using an example where they are not variants of each other, so we're using effectively these examples outside and jumping from one namespace to another. So I'll just leave it there. Thank you.

DONNA AUSTIN:          Thanks, Dennis. Edmon?

EDMON CHUNG:          Thank you. Edmon here speaking personally. I think, Dennis, you're on the right track and also what the small team thought through I think as well, but I think Ariel still has a number of different slides. But what we're talking about is—that's why we go with the hybrid approach, right, which crosses that line by allowing things like string similarity review and confusable strings to be decided not directly by a technical policy approach, but by the processes that will be put in place, like for example, string similarity objection. Then there would be a particular place, and it will be decided not by ICANN, ICANN staff, or these variant tables, but actually people looking at it and say, "Hey, wait. Let's give the arguments. Is this a case where the intervention needs to be had?" But if we don't identify it this way, then it won't even

come to a situation where we would look into it. That's what I understand. Just to not preempt the possibility of addressing the issue, just like, as you said, other namespaces and also non-IDN names also have this situation whereby string similarity and objection and review are also taking certain names beyond a little bit, not to make the decision right there, but then allow a process to then make further decision. That's my understanding. Hopefully that's useful.

DONNA AUSTIN:       Thanks, Edmon. Sarmad?

SARMAD HUSSAIN:     Thank you, Donna. I wanted to add that I agree with Dennis, that in the, I guess, recommendation which is done in SAC060 or written in SAC060 is certainly the discussion is happening in the context of variants. However, I think even though that particular recommendation is in the context of variants, there is still that basically, two kinds of those failure modes, which can occur, which is denial of service and misconnection. There is a possibility that misconnection actually can be motivated by string similarity, not just, for example, variants. And so, in some ways, I think some of those arguments which are presented there are also applicable in a string similarity case where there is sufficient confusion where one end user can look at one string and consider that string as another string. Therefore, using that misperception, types in or clicks and goes to another website which is active, and therefore, misconnects. I guess what SSAC is saying is that in case that other website didn't exist, the user experience would be that the

confused URL will not resolve and that's bad, but that's not as bad as the end user ending up in a different website, because in those cases, that kind of confusion can eventually be also exploited by bad actors. So in some cases, that kind of example is also applicable in other confusability cases where SSAC, I guess, ends up saying that misconnection is worse kind of experience than denial of service. Thank you.

DONNA AUSTIN:          Thanks, Sarmad. Maxim?

MAXIM ALZOBA:          I think since the work of the group is quite relevant to security community and to SAC and to [inaudible] SAC. I suggest staff, maybe Sarmad, talk to maybe to CT Office or SAC to check if the use of denial of service is a good idea. The reasoning is that well-known attack distributed denial of service or DDoS, as you may heard, usually people will just omit the first word and say just denial of service. Since we reference to SAC documents a lot and they're going to read what we write, I strongly suggest to change this term to something else. Or we need a good explanation why this particular term, despite being used outside differently, is pushed in this forum. Thanks.

DONNA AUSTIN:          Thanks, Maxim. I recognize the point that you're making. It's not just this term denial of service that we have that can be problematic for what we're doing here. But we have recognized that we will put a glossary together. When we have these

# EN

discussions, the terms that we use, if there's any confusion, will be quite specific about what we mean, and I think we've done that here. We've been very clear about what we mean by denial of service. I guess, in my mind, what that means is that the end user isn't getting the result that they thought they were going to get, or they put in a URL but it doesn't go anywhere. So, in my mind, that's pretty clear what we mean by denial of service. So if we can just accept that, yes, it may be a little bit confusing, but how do we put it in another way, I don't know. But I think we've been pretty clear about what context we're talking about here. So if we can just push on and accept that that means that somebody's not getting to a website rather than the more serious security term. Okay. So, let's move on, Ariel.

ARIEL LIANG:    Sounds good. Thanks, everybody, for the discussion. The next slide is basically to reiterate the point as Sarmad mentioned regarding the consequence of misconnection. The first one, as he said, it's more problematic than denial of service because it may cause more harm to the user beyond just frustration and confusion.

For example, if the user arrives at a wrong site, even as a legitimate site, it can still result in credential compromised and accidental exposure of information. So in the previous illustration, we tried to show that the user tried to buy some art from another artist that she doesn't even know of. But then in her mind, she made the misconnection with an artist she's a fan for, and then she decided to put her credit card information in that site and buy

art. So that could result in credential compromised and exposure of her credit card information. So it's one consequence.

Another worse consequence is that if confusion is maliciously leveraged, it can be a DNS abuse vector. So as we know for DNS abuse, what we talk about right now is mostly at the second level, but it's confusion already happened at the top level, some bad actor may leverage that, and then cause greater risk for DNS abuse, compounding the top-level confusion and second level confusion. So that will be a worse consequence if some really smart bad actor take advantage of the confusing similarity at the top level. So that's basically reiterating what Sarmad mentioned earlier.

Another thing I want to emphasize is that there are some concerns or reservations raised by folks in the previous meeting as to why we should still care about blocked variants since they cannot exist in the root zone. So I want to emphasize here is that even those variants, they cannot exist in the root zone, it doesn't mean they do not exist in everyday life because they definitely exist in some context in the language or the script people use. That's why we even have them as a variant. But they were blocked by RZ-LGR due to different considerations. Then when the user have this in their mind a blocked variant, they can still perceive and intend to access a blocked variant label domain name without knowing it does not exist in the root zone, because in your mind, it exists in real life but they just doesn't know RZ-LGR and the rules associated with that.

Then just to give you some additional information why some variants are blocked in the root zone, some of the reasons

considered by the script community is that they want you to apply more restrictive rules to the root zone and reduce mutation issues where overproduction issues, and then they also some may consider, if they blocked the variant, maybe it will give more flexibility for gTLD applicant to apply what they actually want to apply without forcing a variant on them. So they have those different considerations. That's why RZ-LGR blocked variants based on the script community's recommendations, but it doesn't mean those variants they do not exist at all in real life. So that's another point I want to mention here.

This is another illustration or example I want to show that a blocked variant may play a part in triggering this misconnection risk. So for example, this is a lady and then she saw a different web address being promoted on a bus. So, shop dot Arabic character—I'm sorry, I don't know how to pronounce it. So welcome, Sarmad, or others who can pronounce this one. But it's advertising for an online shoe store. Then she got interested and was thinking maybe she should try it out and check out the site. So when she went back home, she typed shop dot another Arabic character. This is the red one in her browser. It looks almost exactly like the blue one being promoted on the bus, but there's no web address registered under this domain. It just doesn't exist. So she got a 404 page. But then at the same time, she just thought maybe she misremembered it, and maybe it's just a different way of writing the same character. So she thought the green one, the green Arabic character is basically the same as the red one but they look different. They're variants to each other. So she just tried to write shop dot this green Arabic character and the label, and then a web address popped, but it's actually showing a

handbag store. It's different from what's being advertised on the bus so she got confused there.

In this scenario, even it takes some mental gymnastic, but it may happen. Maybe a good Arabic speaker will make the connection of the dots and such scenario could happen. So basically, a blocked variant, this red one, served as a bridge between the blue, the other domain being promoted on the bus to the green one that this lady mistaken this address was. So that's basically one possible scenario that a blocked variant may play a part in making misconnection happen. That's another illustration.

I think the next couple of slides it's basically to talk about next step, what's the EPDP team needs to consider. But I want to pause for a moment and see whether there's any comments or confusions or additional inputs you want to make based on the examples shown so far.

DONNA AUSTIN:          Thanks, Ariel. I think we're all doing a fair amount of mental gymnastics at the moment. Again, I appreciate that you're trying to give us a visual to help our understanding of what the small team was trying to get to. So I really appreciate that.

ARIEL LIANG:          Okay. Sounds good. So the next step, just having these examples in your mind and the rationale at the back of your mind of what the EPDP team needs to determine, is if the hybrid model is the appropriate path forward. Then there are some key considerations

that we should take into account because those were not deliberated by the small team.

So one is risk analysis. So basically, what you think the likelihood and impact of those failure modes if they happen. Especially for the misconnection risk, do you think the risk is great enough so that the hybrid model is the appropriate way to deal with it? And if not, what will be the appropriate method? So that's the first consideration is about the risk.

Then the second consideration is operational impact of the hybrid model. So that's really related to the implementation, how difficult or easy this can be done in the real string similarity review, so operational impact.

Then the third is the cost and the benefits of the hybrid model. So it's very kind of similar in a way, like the cost in particular, it's related to implementation impact and the benefit is the consequence of that. So these are some considerations for the EPDP team to determine whether the hybrid model is the appropriate path forward.

I know, Dennis, you have your hand up, but I only have one slide left. To provide some, I guess, resources or data to help the group consider these factors. May I? I just want to finish that before we open up for discussion.

So when we tried to think of the operational impact and cost and benefit, definitely hybrid model is something new and we never dealt with variant before. But what we know is about how string similarity review was starting 2012 rounds, and there was some

preliminary analysis, I guess, done by ICANN work. I believe there was a report published in 2016. It's the program implementation review. So they did a pretty thorough review of the actual 2012 New gTLD Program, and there's one section specifically about string similarity review. So I just included some key points mentioned in the report.

So the first one is for the String Similarity Evaluation Panel. ICANN contracted or asked InterConnect Communications and University College London act as the evaluation panel. So they were the two parties that conducted the string similarity review at manual kind of basis mostly, I believe. Then what the panel decided to figure out or find out is that they identified 234 contention sets composed of 754 applications. So you can see it's quite a lot of contention sets and a lot of applications they identified that may have string confusion problems. Then just as some additional information, 230 of that contention sets are exact match. So that includes both ASCII strings and IDN strings. But then there's also non-exact match contention sets. Two of them are ASCII. So one is .HOTELS and .HOTEIS. And the other is .UNICOM and then .UNICORN. I just did a quick search and I know that .HOTELS and .UNICOM, they're active new gTLDs. And then the other two, I believe, they were withdrawn as [inaudible] the contention resolution, I believe. Then there are two other non-exact match of IDN contention sets, and they actually do have varying relationship, if you put them in the RZ-LGR, you should be able to see they are regarded as variants.

Then the first is dot [inaudible]. It's a restaurant. So one is simplified Chinese, the other is traditional Chinese, and then the

other is dot [inaudible], then that's basically .com. One is simplified Chinese, the other is traditional. Then I saw only dot [inaudible] is active top-level domain. So that's basically the outcome of the string similarity review.

Then there are some, I guess, impacts or consequences of the review. One particular point the report mentioned is the results were published much later than originally scheduled because what ICANN forecasted was that all the results will be published in November 2012. But they were not published until the end of February 2013. So that's like three months delay or so, maybe a little bit more than that. Then the delay is mainly due to the volume of unique strings because there are 1380 unique applied-for strings in the application rounds, and that resulted in over one million combinations that require review. So that's a key factor for the delay.

Another consequence of the delay is that the results were not published until two weeks before the deadline for applicant or someone to file the string confusion objection. So that really leaves very limited time for the objector to prepare for the objection. So it does have a consequence for that.

Then there's another dissatisfaction from the community is that the String Similarity Review Panel believe there's no confusion between singular and plural versions of strings, but some applicants, they object to that and then they leveraged the string confusion objection to challenge that result. Then I think Jeff mentioned that SubPro took this into consideration and they actually have the recommendation related to singular and plural version substrings.

# EN

So that's some of the information from the 2012 round string similarity review. I think one key takeaway is that the actual results were published much, much later than originally scheduled due to the volume of strings and the complexity related to the review. So in our consideration of operational cost and operational impact, that's something we probably should keep that in the back of our mind. So that's the data I like to present and that's the end of the staff presentation.

DONNA AUSTIN:          Okay. Thanks very much, Ariel. Ariel, can you just take us back to the slide before, please? Edmon, just your question about did SubPro address this issue, what issue were you referring to?

EDMON CHUNG:          Thank you. I was just typing it as well. Basically, Ariel was saying the deadline for submitting objections ended up being just two weeks after this publishing of the string similarity review. I'm guessing this approach should have addressed that and say the clock doesn't start for the objection for string similarity until after the string similarity review is published. I don't know whether the SubPro actually did address that, but it's definitely not our job here to address this issue, I guess.

DONNA AUSTIN:          Right. The purpose of providing that information, Edmon, was really we've got a couple of things here about operational impact and cost and benefit as a hybrid model. So we're just trying to give a flavor of what happened in 2012 because we had nothing else.

As far as the processing and the implementation of the SubPro recommendation, you might have a little bit more insight into that the most because of the ODP work that's going on that will inform the Board about those recommendations. So I accept that, yes, that's not our role to address that. But the purpose of providing that information was just to give folks a flavor of what happened in 2012. If we're trying to assess the operational impact to the hybrid model, that's a really difficult thing for us to do because we don't have much data except perhaps what happened in 2012. So we can see that just the volume on strings that were applied for made that process perhaps more complicated than folks thought it would be and took longer than what folks thought it would be, so one would hope that moving into the next round that those issues would be resolved.

But if I think about it, my thinking on this as well, too, if we introduce this confusing similarity based on what the hybrid model is proposing, how much more difficult does that make the string similarity review process? I know that's not the only consideration but I think that's where, when we reviewed the hybrid model last week and the week before that, I think what folks were jumping to is the operational impact and how difficult will the hybrid model be to implement. That's something we need to discuss because the small group hadn't done that.

Dennis, I think you might have had your hand up a long time ago to come back to something. So we'll start there.

# EN

| | |
|---|---|
| DENNIS TAN: | Thank you, Donna. No, I think I echo what you just said. This is very important for us to consider. We have heard the concerns that the string similarity review of variant sets present, but at the same time, we need to think about the cost, the operational impacts. We want IDNs and variants, for that matter, to be successful. But we keep in mind that it needs to be affordable for applicants to go through the process and manage them, operate them, and offer to the operators. I mean, that's the end goal. So we have to think of what's our precious goal at the end of all this process. And we don't want to raise the bar, higher cost for operators to go through, and then what? So I think putting this into the balances to make sure that the utility that we get, it's reasonable. I think that's what I just want to say. Thank you. |
| DONNA AUSTIN: | Thanks, Dennis. I think that the goal here is twofold, right? We need to find a way to introduce IDNs and their variants at the top level so that there's no security and stability issues. I think confusion falls into that bucket. So we do need to be mindful of that. We do want to introduce IDNs and their variants in a way that won't be confusing for the Internet user. So I think that's a goal. I think we also need to keep in mind that this is a priority for the ICANN Board and the ICANN community when we talk about IDNs, and we're trying to work out how to introduce that variant piece. |
| | All right. I do agree that another goal is how do we do this? How do we come up with a recommendation that's implementable? When we think about that, it is possible for us to think about, well, maybe with the string similarity review was used for 2012, it's |

# EN

going to be used again in the next whatever the subsequent round is. But is there something that perhaps we need to do differently for IDNs to get through that review process? So let's not be constrained, I suppose, is what I'm saying about how we think about implementation.

So what we really want to try to get to here is what are folks still thinking about the hybrid model? I know that there were reservations last week or the week before, actually, when we introduced the hybrid model. So I really want to understand whether those reservations still exist. And maybe we if we can understand why that reservation is there, I'd really like a conversation with those that are in the small group, to perhaps explain that they may have had the same reservations but how did they overcome them? So can we have a conversation around whether we still have reservations about the hybrid model, or whether we're in a position to accept the hybrid model for a string similarity, and then move through "Okay, we accept it"? This is probably the best way to go. Can we implement it, and how do we do that? How are folks on the hybrid model? Do we still have reservations? Or based on what we've seen here today, do we think there is a valid reason for going down this path? Dennis?

DENNIS TAN:              Hi, Donna. To your question, whether we do have reservations or not on the hybrid model, aren't we going to do these next steps and then have a holistic view of utility and cost, and how do we see the hybrid model in the full context of it?

DONNA AUSTIN:    We can if that's the conversation folks want to have. So if your reservation is because of the operational impact or the potential cost, then that's a valid reason for having reservations about the hybrid model. I'm happy to throw this over to folks and to see where we get to.

The real challenge here for us is that I don't know that we have the expertise or the ability to do some of this work. I think Michael Karakash might be on the call, but it might be some time before we can get ICANN to have a look at this.

The other thing that occurs to me is we've got the hybrid model now as a possibility, we could try to do some comparison of how that stacks up with I think what was initially our option two, which was applied-for and allocatable variants but didn't include the blocked. We could try to do that kind of analysis. I don't know whether the small group did any analysis in that regard. So look, it doesn't matter which way you're thinking about this, but just your thoughts on whether the hybrid model is a good model and one that we should pursue because of X, Y, Z, whether you have reservations because of the operational impact, it doesn't matter. Let's just see if we can have a conversation and see where we get to, see if we can identify where the real issues are. Satish?

SATISH BABU:    Thanks, Donna. If I remember right from the previous meeting, the objection or the reservations were largely due to the counterintuitive nature of why we should consider blocked variants at all. So I think this is not very straightforward, it takes an effort to

understand what could be the consequences of not going for the hybrid model.

We're not discussing this in the ALAC team, but it's very much clearer for me after this presentation, after the last example, that this is a kind of edge case which we tend to kind of ignore because it's a blocked variant. But blocked variants have their own kind of consequences, and that has become very clear from this particular example on the screen now. Because we want to be conservative about it, on a precautionary approach, I will certainly personally go for the hybrid model. I agree that we have to consider the cost aspect and the complexity aspects, at least on an indicated basis, because we can't maybe drill down to the final kind of position, but at least to know whether this is going to be a prohibitively expensive option or not. To me, it doesn't appear so but there could be better opinions on that aspect. Thanks very much.

DONNA AUSTIN:          Thanks, Satish. Michael?

MICHAEL BAULAND:          Also on the topic of operational cost, which Dennis talked about, the safest approach, the most restrictive one would of course be the complete Level 3 comparison where every blocked variant would be all to compare to all other blocked variants. But if you look at the usual variant distribution, so to say, the vast amount of variants is in the blocked disposition. So, if we were to compare blocked with blocked, we really get an exponential amount of

# EN

required comparison. And that's why we, as the small team, came up with this hybrid model where we still catch most of the cases. But because we do take into consideration all the blocked variants, but since we just compare them with Level 1 and Level 2 variants of the other string, the number of required comparisons is by a magnitude lower than the full Level 3 comparison. For that reason, we thought that hybrid model is the best tradeoff, so to say, to catch most of the cases without going into too large numbers. Thanks.

DONNA AUSTIN:          Thanks, Michael. Dennis?

DENNIS TAN:          Thank you. I have two comments. I understand what Michael is explaining and Justine just confirming. That's the compromise that the small group reached out to. I'm just trying to process that in my mind. I'm just assuming how this work within this small team, is that you look for examples and reverse engineering to get to the primary label, the source labels, and that was an easier exercise to find those confusable labels. But in the practice, you will have potentially two primary labels, applied-for labels that don't look alike at all. But somehow one blocked variant is going to be look-alike. How is that going to be identified? The machine readable format will not tell you that. Somebody has to go through the whole list of blocked labels to see if that matches in applied-for labels. That comes into the operational cost. So that's one thought that is still working through my mind, how's that going to work? How do you identify a blocked variant that is confusable similar to

applied-for variant? There are no means, no machines that can tell you that unless, of course, somehow you define that. Otherwise, the RZ-LGR will tell you.

The second point is one that I think I have said in the past but I'd like it to be reflected somewhere. Some of the blocked variants are not well formed labels. For example, in the Latin LGR, many, if not all, of the blocked variants are going to be cross-script variant labels, which are prohibited from delegation. So, in those cases, and I certainly cannot speak with how other scripts work or other scripts LGR, that is, in the Latin example, Cyrillic and Greek, which are related, then those blocked variants ought to be dismissed, not even considered for visual similarity. Because again, cross-script labels are not to be delegated to the root zone so that do not even entertain making a visual confusing similarity review for such labels. Thank you.

DONNA AUSTIN:          Thanks, Dennis. Michael?

MICHAEL BAULAND:       Thanks. That's an interesting approach, Dennis. I have to admit that as the small team, we did not take into account that some of the blocked labels are actually not just blocked but also labels mixing script which would never appear anywhere. So maybe it's an idea to have hybrid version to model in which we say we still compare the blocked variants with Level 1 and 2 variants. But out of those blocked variants, we only considered those which are valid labels. Thanks.

**EN**

DONNA AUSTIN:             Thanks, Michael. Sarmad?

SARMAD HUSSAIN:          Thank you, Donna. Just to add to that last bit about mixed script variants, I think those would eventually be not valid labels except where script mixing is allowed in certain languages, for example, in Japanese. I guess in those cases, we can devise the tool to just take out those labels, which are, for example, mixing Latin and Cyrillic. So that part, of course, can be done by the two. Thank you.

DONNA AUSTIN:             Thanks, Sarmad. I have a question that's related to what Dennis said that you could apply for an IDN TLD as a top level so it has a certain meaning and a certain purpose for the applicant. Then it turns out that that is a blocked variant of another TLD that somebody has applied for. So what's the consequence of that? Michael?

MICHAEL BAULAND:         If those two labels that are paid for are actually in the same variant set, independent whether one of them would be a blocked variant of the other, then this case is automatically caught by the case that they are variants of each other and they are automatically in the contention set.

# EN

DONNA AUSTIN: So I think there are still some reservations about the hybrid model and whether this is how the test will be done. So, I guess there's a difference between what is confusingly similar and what is an exact match. So if exact match, ASCII say, would go into string contention. It would be the same for IDNs, but where it gets tricky is their associated variants. Nigel, go ahead.

NIGEL HICKSON: Thank you very much, and good afternoon. This has been really, really worthwhile and much, much clearer than I had understood initially. Thanks so much for the diagrams, they help. In terms of the hybrid model and implementation, and perhaps there's variations as a hybrid model, as Michael said, but I just think we're in a difficult position without all the facts and figures in front of us. The sort of gut feeling that we have, I suppose, as some of the government's, is that there's always a risk, I suppose, but we don't want to deny opportunity—I was very struck by what Edmon said earlier—and certainly, there might be difficulties. But, in general, I think the hybrid approach has some benefits. Interestingly, it took us back to the string similarity review. I seem to remember this, the tensions and the problems around it when it was published back in 2013 or whatever. Of course, there was a lot of controversy about whether HOTELS and HOTEIS were confusingly similar or not. Thanks.

DONNA AUSTIN: Thanks, Nigel. I wonder whether it's possible to do—and I'm kind of moving to the operational implementation of if we go with a hybrid model. But this isn't just with the hybrid model. I think we're

going to have similar issues if it's just what would have been our option two. So blocked variants are another consideration but we still have allocatable variants.

When somebody applies for a TLD, they have to identify the string and the purpose of what they can use that string for. And also with IDNs, the applicant had to identify what the meaning of the string was. Is it possible that with the implementation, we could recommend that the string similarity or confusion for IDNs be done in a quiet way? What I mean by that is that the initial analysis is just the primary TLD that's being applied for or the source label to see whether there's any confusion there or whether there's exact match labels. So you do it at the top level, and then you kind of work your way down. With the variants that are applied for, is that problematic? Then is there a problem with the blocked? But I think maybe if we can try to keep this at the top level, what are the consequences for the source label that's been applied for, and then see if there's some kind of layered approach that we could have to implementation. I just wonder whether that's some way that we can take some of the challenge out of going down. As I said, I don't think this just applies if we do the hybrid model. I think it applies regardless of whether we go for the hybrid or we go for option two. Maxim?

MAXIM ALZOBA:        Also, I suggest that we do not forget that if we, as a result, forbid something, it will take ages, if at all, to allow it again. Example was international IGO's names which are on restricted list, like Olympic Games, Red Cross, and in order to get it, basically many years passed and nobody got anything out of this list. The same

# EN

organizations who wanted there to be blocked, they haven't blocked but no access. Thanks.

DONNA AUSTIN:          Thanks, Maxim. Michael?

MICHAEL BAULAND:       Thanks. To quickly respond to Maxim, yes, it may take a while if we block something to later make it available again. But I think on the other hand, if we blocked too little and those get assigned, then there's almost no way to take that back again. So blocking too much might cause a problem but that can be remedied. But blocking too little, it's almost impossible to take it away again. Thanks.

DONNA AUSTIN:          Thanks, Michael. Maxim, is that a new hand? Okay. All right. So where do we think we are? Based on the conversation, what are people thinking about? Is this still really hard and we still need some time to digest this? Should we think about how this could be implemented? To be honest, I don't even know where to start to begin on how we would do that. But perhaps we could take a shot at it. Dennis?

DENNIS TAN:            Thank you, Donna. I'm thinking along those lines, Donna. I'm not sure I'm going to give you an answer, but I have a question. I'm just thinking as far as next steps. For the sake of this example

# EN

conversation we say, "Okay, a hybrid model is okay and that's what we recommend." And of course, the policy recommendation would be really in a way that establishes what the objective that it wants to achieve, but it doesn't go into the how, right? That's an opportunity for us to give some guidance in implementation notes and what have you, but really, it's going to be up to the IRT in order to devise the artifacts, processes, algorithm tools in order to provide or try to meet the policy recommendation.

So that's what I worry. I don't know whether we should worry about it in terms of the how. I know we want to establish the what and why it's important. But having a few PDPs under the belt, and looking at the experience, I think we want everybody collectively, we want a policy recommendation that is implementable and it can be done in a reasonable fashion. But I'm struggling here to separate whether that's one thing that we need to do worry about or not. I guess we have to take it from there. I don't know. That's my thought process right now, Donna. Again, I'm not giving you an answer. It's more of sharing my own thoughts here and try to find a way forward for us to come up with the policy recommendation and beneficial notes and other artifacts that we should produce.

DONNA AUSTIN:            Thanks, Dennis. You are correct. Really, our remit here is policy recommendations. But we've also put a lot of thought into how the recommendations will be implemented. I think that probably for this one, that's what we're all struggling with. So we could legitimately just develop a recommendation and the policy perspective as it relates to string similarity. Because I think those on other processes that we discussed last week, I don't think there

# EN

was this type of angst related to those other processes. I think we were in a reasonable place with those, which is good.

So we could for this, if folks are leaning towards the hybrid model, we could develop a policy recommendation around that, noting that we do have concerns about whether this can be implemented or not. And perhaps what we could do is seek to get some feedback from ICANN Org. Michael Karakash said that they would be looking at our recommendations and providing feedback. So this is one that we could flag, and then come back to it once we've got that information. But I think we could put a recommendation in draft report that we send out for a hybrid model, if that's where we think we're leaning here, and then see what feedback we get. We can identify, as part of our rationale, some of the challenges that we think could be created in implementation. So that's always an option to us as well.

Any thoughts from folks on that? That's if, of course, we think we think the leaning of this group is towards the hybrid model. I know Jeff isn't on this call and probably absent. Being on this call, we'll probably still have reservations about this approach. Dennis, it may be too early to ask you this question, but if that's the approach that we just focus on the policy recommendation, where do you think the Registry Stakeholder Group would be?

DENNIS TAN:         Thank you, Donna, for the question. I'll need to take it back. I don't want to get ahead of myself. We do have an upcoming meeting with the Registry Stakeholder Group. We'll present the problem at

hand, how the EPDP is suggesting/recommending how to proceed forward. We'll discuss it and come back with their input.

DONNA AUSTIN: Okay. Thanks, Dennis. Of course, the other—I don't know that we've discussed this at any great length. But we're only talking about language that goes into the draft initial report. I think what I'm seeing in chat is that there is a leaning towards the hybrid model. So we could identify that this doesn't have full consensus, but it does have good support at this point in time. And that still gives an opportunity for folks to comment on the recommendations. As part of the public comment process, if we get more detailed feedback, then we will consider that and see how that changes the recommendation.

Okay. So, in summary, I think where we are, we're leaning towards the hybrid model recommendation. We're going to set aside the operational impact and cost-benefit, and even to some extent, the risk analysis, because that's kind of outside our expertise. But we hope that ICANN Org, as part of their feedback, would have a look at the recommendation and give their thoughts on it. Then we can put this out of the draft recommendation and see what response we get through the public comment, and if there's strong pushback or some good information, new information as to why that shouldn't be the final recommendation, and we take that on board.

All right. Thanks, everybody, for your patience. Thanks again to the small group for the work. I think we have a path forward, notwithstanding that, Dennis, I want to leave the window open for

you to come back with the Registry Stakeholder Group's position, but with that additional context, maybe the conversation is a bit easier. Okay. Thanks, everybody. We will see you all again at the regular time next week. You can end the recording now, Devan.

DEVAN REED: Thank you all for joining. Once again, this meeting is adjourned. I'll end the recording and disconnect all remaining lines.

**[END OF TRANSCRIPTION]**