
ICANN Transcription

IDNs EPDP

Thursday, 10 November 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/M4YFDQ>

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, good evening. Welcome to the IDNs EPDP call taking place on Thursday, 10 November, 2022, at 13:30 UTC.

We do have apologies from Farrell Folly and Maxim Alzoba.

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 is 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.

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

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 say 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 our Chair, Donna Austin, to begin.

DONNA AUSTIN:

Hi, everybody. Welcome to today's call. Sorry for my tardiness this morning. I have a new laptop, and Zoom wasn't ... Well, it wasn't being very friendly.

Okay. So we've had a request from the ccNSO for a joint call to discuss ccPDP 4, particularly the issue of string similarity, and we've agreed to do that. So that will be next Tuesday, from 14:00 to 15:30. And we should be able to share an agenda with you between now and then.

Ariel, do you have any additional information to share with folks on this? I think it's essentially at the request of the ccPDP 4. But if you've got more information, that would be helpful, Ariel.

ARIEL LIANG:

Thanks, Donna. Not so much information, but it's mainly a call to compare notes with EPDP and learn about the way we deal with the string similarity review and share their preliminary

recommendations on that. So it's really just an information exchange.

And then I think the folks who are familiar with this ccPDP 4 deliberation can correct me if I have the wrong impression. They're focus on string similarity review seems to be different from ours, and I think they're focusing on which panels we conducted and what the procedures and criteria. And then they haven't gotten into in-depth discussion about variants and [inaudible] that. That's my observation so far. But happy to be corrected if that's the wrong impression.

So basically they want to hear from the EPDP team about our recommendation and then see what the similarities and differences are. But I think this time, the differences are pretty big because the focus areas are different.

DONNA AUSTIN: Do we have information we can share with the group about where the ccPDP is on the topic? Or don't think they have anything available to share at this point?

ARIEL LIANG: I can check with Bart and then see whether they are comfortable sharing any document ahead of time. And then that will get back to this group.

DONNA AUSTIN: Okay. Alright. Thanks, Ariel.

The other thing that we discussed last week—actually, the week before—was that we have some dates coming up where we wondered whether we should go ahead with the meeting or not because it's close to holiday periods and also the IGF. And in particular the two dates were the 1st of December and the 22nd of December. I'm really keen to go ahead with these meeting because I'm very conscious that we've got a target deadline that we've sent for a draft of that first report, which is in March, and I'm concerned that we're not going to make that. And I didn't mean to say that out loud. So I'd like to push ahead with those meetings and encourage folks to, if you're unable to attend, please listen to the recording and provide your inputs if you have any online.

Just on the schedule adjustment, for those of who follow the council list, you may have seen that Farrell as our liaison has submitted our change request to the GNSO Council. And that will be considered at the next council meeting, which I think is in about ten days' time.

There is a question from Pau McGrady about the impact of our work on the SubPro and its implementation. So we're going to discuss that on the leadership call that will happen later today and work out what's the best response for Farrell to go back to the council. So I understand Paul's concern, but I think there are elements that we can't control. So we don't know currently what the timeline is for the implementation of SubPro. We think, based on a conversation we had with Karen Lentz, who is responsible for SubPro, that if we have a draft report in March of next year and a final report no later than this time next year, then that should be timely for SubPro implementation. The piece that's a little bit

unknown is Phase 2—so the second-level questions that we will need to consider. So we'll have a conversation about that on the leadership team later today and work out what's the best response to Paul.

So with that, any questions on any of that? I understand that the notification about the joint call with the ccPDP is a ... Well, the 29th of November is still a ways away, so hopefully folks can make that. And then we will have the call on the 1st of December. So we will do our scheduled call two days later. So any questions or comments from folks on either of those topics?

Okay. I don't see any hands.

So with that, I'm going to hand it back over to Ariel, and we'll continue our discussion on the options for legal rights and community objections. So are you good to go, Ariel?

ARIEL LIANG:

Yes. Thanks, Donna. And, actually, not too much from my end because this is really just trying to wrap up the discussion. And I reached out to Jeff. I hope Jeff saw the e-mail from me because he offered to write up some points regarding the, I think, exception process and how it interacts with the objection process. And I'm hoping Jeff is able to provide some input on that. And I see he unmuted himself, so maybe I'll stop and ...

JEFF NEUMAN:

Yeah, I apologize. I've been delinquent on what I owe you, but I will write it up. It'll be the same for the legal rights and for ... Well,

there'll be slight differences, but essentially it's a process for the applicant to show that is not confusingly similar to the either the existing TLD in the case of—well, that's the string similarity review, but in this case (either string confusion or legal rights objection) that it's not confusingly similar to, in the case of legal rights objection, the existing TLD or another—whatchamacallit?; sorry, now I'm confusing myself here—application if it's put into a contention set, depending on who the objector is, right? There's two types of objectors. So, yeah, I just need to put it in words, but that's pretty much the concept that would apply.

Now, we haven't discussed or we sort of discussed that in the community objection, and I think it'll be some similarities there but obviously has to be tailored to a community showing that it's use of string that is similar to a blocked variant is not ... Well, the community objection is a little different in terms of the standards. But, yeah. So instead of confusingly similar, it would be that the community ... Sorry. I'm slow today. I'll have to go back and just tailor it to the actual criteria of the community-based objection, depending on the challenge.

Well, the concept, again, is that—and it's just one of the difficulties with the hybrid model—because the either existing TLD or, if it's an application that's put into a same contention set, the other application because it can't launch that blocked variant that, being confusingly similar or looking the same, is not going to cause the type of consumer confusion or the, in way you all have put it before, disconnection. Or non-connection? Disconnection. That's, I think, what you've mentioned. Thanks—misconnection. Thanks.

DONNA AUSTIN: Okay, thanks, Jeff. Apologies, folks. I'm having some technical difficulties where everybody is currently yelling at me when they speak, so I'm trying to find the volume controls where I can turn folks down.

DEVAN REED: Donna?

DONNA AUSTIN: Yeah?

DEVAN REED: I'm so sorry, but there's an echo when you speak, so it's like we're hearing two Donnas.

DONNA AUSTIN: Okay, folks. Ariel, can you kind of fill the gap? I'm going to try to sort this out. Sorry, folks.

ARIEL LIANG: Yeah, no problem, Donna. It's actually not too bad with regard to the echo. Hopefully, it's sorted out.

And thanks, Jeff, for elaborating on what you plan to write up. And I think that's very much related to the string similarity. And I think we'll wait to see your write-up.

And I think the objective today is that hopefully the group can agree on which option to choose for the two objection processes

at hand that we're discussing right now. So for legal rights objection, we have two options that were discussed by the small team. So Option 1 is basically that only the applied-for string and applied-for variants can be subject to a legal rights objection. But the non-applied-for variant, as well as the blocked variants, should not be subject to legal rights objections. So that's Option 1.

And then Option 2 is basically that everything is on the table—so not only the applied-for string and the applied-for variant but also the other non-requested allocatable variant and the blocked variants. So someone could file an objection against these variants based on legal rights.

So there are two options. And I think I just want to go back to the group and see which option you think you will support and then the rationale behind that.

And I see Jeff has his hand up.

JEFF NEUMAN:

Thanks. I think the answer is going to be it depends on the exception process and what we're actually seeing. Again, the whole concept is that an applicant should not be prevented—in this case, legal rights ... So a trademark owner should not be prevented from launching a domain that matches its trademark in subsequent rounds unless it was a variant of what has actually been allocated. The difficulty is not so much with the blocked variants but with the non-requested allocatable variants. I think that's the harder one because, if we say that variants must be with the same entity, whether it's allocated or not, and we're giving

them the first right to get those variants, then I think we have to allow the objection to be filed against all the non-requested and the blocked—all of it. It has to be Option 2.

But I think that the remedy for the objection should not necessarily be ... Sorry, I'm just trying to go through the objection process in my head. If someone objects—let's say a legal rights objection—and it's only similar to, let's say, an allocatable variant but not necessarily similar to the primary string that's applied for, then the applicant shouldn't necessarily be prevented from launching its primary, but it may be prevented from launching an allocatable variant, if that makes sense. So the differences may be in the remedy, not necessarily what the objection could be filed against.

In other words, normally, in a legal rights objection, if there's an application, and a third party objects to the application based on legal rights and succeeds, the remedy is that the applicant cannot move forward. But here, if a trademark owner is objecting based on an allocatable but non-requested allocatable variant, that doesn't necessarily mean, if it succeeds, that the applicant can't go forward with its primary string, if that makes sense.

DONNA AUSTIN: Jeff, can I ask one question?

JEFF NEUMAN: Yes.

DONNA AUSTIN: So I understand what you're saying and why you're saying it, but I wonder if there's a question here about, how does that impact on one of our first recommendations, which is basically the sanctity of the use of the RZ-LGR? So I just wonder, what's the consequences on that recommendation? If you have any thoughts on that.

JEFF NEUMAN: I'm not sure it does affect that. I think we're, in this [inaudible], we're still using it. And as Edmon said in the chat, what we're talking about is a similar string, not the string itself.

DONNA AUSTIN: [inaudible] same string. Okay, thanks, Jeff.
Hadia?

HADIA ELMINIAWI: Thank you. So I have two questions here. So does a non-requested allocatable variant block a future IDN TLD? And does a blocked variant block a future IDN TLD? If the answer is yes, then I think we need to go with Option 2. But if the answer is no, then Option 1 is good. However, the applicant will suffer from the possibility of not being able to allocate the allocatable variant in the future. At least this is how I'm trying to think about it. Thank you.

DONNA AUSTIN:

Thanks, Hadia. I'm trying to get my brain back into this discussion, and it is tricky. And the consequences of if we go a certain way—how does that impact future applications?

I think I agree with you, Justine, that we need to set out—and Jeff, about—is it possible to set out some use cases so that we can look at this in a practical way, rather than an abstract way?

Jeff?

JEFF NEUMAN:

Thanks. Absolutely. We need to come up with just a bunch of examples just to walk through all the scenarios.

But also, to Edmon, I think we discussed this the last time: the wording of these—whether something can be filed or can't be filed—I think is not exactly the right wording. It's whether an objection could succeed or not. I think this is just shorthand. Obviously, anyone could file an objection to anything, but whether it can succeed is a whole other story. And I think that's what we mean when we say, "Legal rights objection can be filed," because of course you can't stop someone from filing an objection. They will or they won't. And it gets dismissed or it gets accepted. So I think, when reading these, we shouldn't be reading it as to whether someone can actually physically file a complaint or an objection. It's whether they have any chance of success if they do file it.

DONNA AUSTIN:

Okay. So I think we need to do a little bit more work on this to come up with some scenarios or test cases so that we can work this through a little bit and see where we end up. And I agree with Jeff. If you can say you can file an objection but there's no chance you'll succeed, then really what's the point?

Okay. So does it make sense we just move on from this discussion, accepting that we need to do a little bit more work to understand what we're talking about here?

Ariel?

ARIEL LIANG:

Thanks, Donna, and everybody. Just based on staff's observations, so far I think—correct me if I'm wrong—we have to work around Option 2 for a potential recommendation, but that depends on the write-up from Jeff in terms of the exception process and then what the potential outcome would be if the objector prevails. What strings would be taken out of the process, whether it's affecting the entire application or just the affected strings? So I think that's our observation so far.

So maybe we can work around Option 2 as a potential recommendation and then incorporate the exception process and then also think through the potential outcome of the objection. So hopefully I summarized it right.

DONNA AUSTIN: I think that's right, Ariel. Okay. Alrighty, it looks like people are supporting that approach, Ariel. So let's draw a line under this for now, and we'll come back to it.

ARIEL LIANG: Sounds good. So, moving on, the community objection is basically actually the same: two options. And I'm wondering whether we treat them differently from the legal rights objection or we have the same treatment, which means we are converging on Option 2 but we need to also consider how the exception process could possibly work for this option and then what the potential outcome could be. I wonder whether that's what this group thinks or whether we're thinking differently.

DONNA AUSTIN: Jeff, go ahead.

JEFF NEUMAN: I think, in my mind at least, if we are moving forward with this hybrid model, even with the exceptions, you have to do Option 2 because we should never have a position where someone is not able to object, even if it means that they would be blocked in the future from launching something that matched their community or their legal rights or whatever it is. I think the hybrid model necessitates having Option 2.

DONNA AUSTIN: Thanks, Jeff. And I see you have support from Michael with that statement as well.

Dennis, was your hand up?

DENNIS TAN: I'm sorry, Donna. Was that me?

DONNA AUSTIN: Yeah, I thought I saw your hand go up, but maybe I'm just seeing things at the moment.

DENNIS TAN: Maybe you were reading my mind because I was typing in the chat. So perhaps it's best to just speak my mind. Just wanted to note, as it looks like we are inclining or leanings towards Option 2 here, from the Registry Stakeholder Group, that we noted, on our position regarding the hybrid model and the consideration and using all blocked variants in any evaluation process, that blocked variants—some—are valid labels in the sense of protocol valid as a DNS label, but many of the blocked variants are what I'm calling or what we have referred to in the past as not-well-formed labels—for example, mixed script labels and whatnot.

So we should be careful about putting everything in the same bag here because those labels will never be delegated, regardless of the calculation of the root-zone LGR. So let us be careful as to how we refer to all blocked variants here because we don't think that all should be considered in terms of the mixed bag that you

can find in the blocked grouping. And I will leave it at that. Thank you.

DONNA AUSTIN: Thank you, Dennis.

Jeff?

JEFF NEUMAN: I think that's a really good point. Maybe they deserve different terms, right?

But let me ask—and this is more a question for ICANN staff or actually Michael or people who know more technically—how difficult is it to figure out which variants are blocked and are mixed scripts or not well-formed names? Is that a relatively trivial function, or is it actually going to be so time-consuming to separate out those malformed ones?

And Michael is saying it's easy. Okay, good. That's the answer I was hoping for. So, yeah, if there were a way to not include whatever the term where using for those malformed or mixed scripts are or whatever—we didn't exclude those from the hybrid model to begin—that would be ideal.

DONNA AUSTIN: Thanks, Jeff. So Sarmad is saying in chat that mixed-script variants can be determined computationally. So he agrees with Michael. Alrighty. So we've got a bit more work to do on this one.

And just on the hybrid model, that is one of the things that our ... Our initial thinking on that has gone to ICANN Org. And I understand from Michael that we should be getting some feedback from ICANN sometimes this month. So I think that's going to be really helpful to our discussion, informative to our discussion, so we're looking forward to that input from ICANN Org.

Okay. So with that, I think, Ariel, we'll move on to what's next on the agenda.

ARIEL LIANG: Thanks, Donna. I guess you want me to take over.

DONNA AUSTIN: Yeah. Just before you start on this, on this topic and this particular question, I'm going to take Chair's prerogative and say there's an obvious answer with this. There's really no reason to permit the registration of decorated two-character Latin labels, primarily because there's potentially a political conversation that you get into with ccTLDs and what they are and that the confusingly similar conversation ... the likelihood that, if somebody did submit a two-character Latin label as an application, that it's very unlikely that that it's going to succeed. So I just want to put that out there: in my mind, I think there's a pretty simple answer to this question, but the challenge for us is the rationale.

So with that, Ariel, I'll hand it back to you. Thanks.

ARIEL LIANG: Thanks, Donna. And I think, with some of the background I'm going to refresh folks on, it's probably not super difficult to develop the rationale to answer this particular question.

The other thing I want to just quickly note is that Jeff sent out the proposal in the mailing list about developing a similar recommendation as the one related to single-character TLDs. So we can use the time today to discuss that: whether the group wants to develop something similar to that kind of recommendation. So that's a little bit broader than the scope of the charter question at hand, but I think it's worth discussing.

And I see Jeff has his hand up.

JEFF NEUMAN: Thanks. And I would love to go through that e-mail because it points to the fact that we've been just referring to the Latin script or we've mostly been talking about the decorated Latin, but there's a lot of scripts that have characters that look a lot like Latin scripts. And I put all those examples or at least a lot of examples in that e-mail. And I'll open one up and display that. But that was sent on ... When was that? October 27th.

ARIEL LIANG: Thanks, Jeff. I have put together a slide to show the example, so rest assured that it will be reviewed by the group together today.

JEFF NEUMAN: Great.

ARIEL LIANG:

[inaudible]. If it's okay, I will go ahead with the slides here. So this is a refresher for folks about ccTLDs' treatment of the two-character ASCII labels. And I think a lot of you already know, but I just want to provide some additional information, and that was after checking with Bart to learn more about this topic.

So basically maybe all of you know that the two-character ASCII labels are derived from the ISO 3166-1 wish. And then that list is external standards dictating country codes. And then that can be two- or three-letter alphabetic codes.

And then also another important thing we want to note is that the two-letter ASCII country codes are not only used in the DNS but also in other aspects, such as passports, currency, transportation, and other things. Governments use those two-letter combinations to indicate their country, territory, other areas.

And then another piece of information I learned is there is only a very limited number of two-letter ASCII codes that are currently not in use. So for example, I think the two-letter ASCII codes that start with the letter "S" as in "Sarah"—I think all of them—are already in use, except for two possible combinations that are not already claimed by governments of countries and territories. So there's only very limited ones that are vacant, I guess.

And then another thing I want to note is that ICANN has been maintaining a restriction on the use of two-letter ASCII combinations for TLDs because it wants to ensure there's no conflicts with future country-name assignments. So, so far, for the

two-letter ASCII combinations, they're basically all ASCII labels, and there's no IDN ccTLD labels in the Latin script. So there's no decorated two-character codes currently. But that doesn't mean that it won't happen in the future. So that's an important thing to note.

And then, furthermore, what Bart told me is that there's a presumption in the ccNSO that the two-letter Latin codes are automatically excluded from gTLD applications because of the political sensitivity that Donna mentioned. And also, as Jeff mentioned, there are other scripts that may cause potential confusion with two-letter ASCII combinations, such as Cyrillic, Greek, and some other scripts.

And I'll stop here. I see Sarmad has his hand up.

SARMAD HUSSAIN:

Thank you, Ariel. I just wanted to add one more thing in the list here. ISO 3166 is actually maintained by the ISO 3166 Maintenance Agency. It's an independent body which actually has rights to assign any unassigned two-letter codes to any country, territory, or for other purposes. So that's not, in that way, under ICANN's control. And therefore, all the two-letter codes are reserved for that reason because, even if they're unassigned at this time, they could actually be assigned in the future by an independent agency. Thank you.

ARIEL LIANG:

Thanks very much, Sarmad, for that information. And thanks, Jeff, for mentioning that two-letter ASCII codes are reserved, and that was affirmed by SubPro. So thanks for the context.

And then another thing we did some research on is regarding the 2012 round and what happened in terms of two-letter TLD applications. So we found out the following. The first one is that two-letter ASCII strings were not permitted in the 2012 round, and that's exactly the case. It's to avoid conflict with current country codes but also future country codes based on the ISO 3166 standard.

And then another thing we found out is that there are indeed two-character IDN strings, and then there were applications for that. But there's some differences there. So we found out there are two-character IDN strings in other scripts, like Chinese and Japanese, for example. I [can't] confirm that Arabic also had two-character string, but at least we saw in some Latin scripts there are applications for those. But then they are only limited to those specific scripts. But for the Latin script one, we definitely didn't see any two-character TLD applications. And there are applications in Cyrillic, for example, but they're not two-character. They are at least three. And then we found there are some IDN Latin script applications, but it's in German. And it's actually on the slide here. And it's definitely more than two characters. It's a long string, basically. So that's what we found out.

And I think the reason why we didn't see any two-character TLD applications in the Latin script is that there is a very important factor in the string similarity review: all applied-for two-character IDN strings were compared against any possible two-character

ASCII combinations. And then the reason is to protect possible future ccTLD delegations. So, in other words, if someone applies for a two-character IDN TLD in the Latin script, it's most likely not going to pass the string similarity review in the 2012 round. So probably that's the reason why we didn't even see any application for that, even though, in the Applicant Guidebook, it wasn't explicitly stated that two-character Latin script application is forbidden. That wasn't the case. So that's what we observed in the 2012 round.

So back to the question itself, when we tried to analyze the charter question, we stratified or separated the question into three sub-questions, just to help us understand it better. So the first sub-question is, can a two-letter gTLD label in the Latin script be applied for in future rounds? So the answer to that is, if it's a two-letter ASCII label, then it's definitely not allowed. And that is also confirmed by SubPro. But currently there's no restriction on an application for two-letter IDN TLD labels in the Latin script. So someone could potentially apply for it, but to get that delegated, that's subject to the successful evaluation of that application, which includes the string similarity review. So that's what we think the possible answer to the first sub-question is.

And then the second sub-question is, can a two-character IDN gTLD label in the Latin script be applied for? So the answer to that is, yes, it can be applied for, but it's going to be extremely unlikely that application can pass the string similarity review because of what we just mentioned: the criteria or the mechanism for the review won't change, so all of the two-character TLD label applications will be compared against any possible ASCII two-

letter combinations. So it's very unlikely it's going to pass it because a Latin script application for a two-character TLD will either be very much the same as a variant of a two-character ASCII combination or very similar. So it's just not going to pass the review due to the hybrid model the group is proposing. So that's the answer to the second sub-question we can think of.

And then the third is, can a two-character IDN TLD label in the Latin script that is not a variant of any two-letter ASCII label be applied for? So we think this question is actually flawed because all of the two-character IDN gTLD labels in the Latin script have to be a variant of a two-letter ASCII label because the two-letter ASCII label can be any combination, so you [won't] be able to see if that's going to be a variant of a decorated two-character gTLD label in the Latin script. So the question is actually wrong.

So that's our analysis to these three questions. And I'm hoping I'm not confusing everybody.

But I see Sarmad and Michael have their hands up.

SARMAD HUSSAIN:

Thank you. Just to lay out the potential possibilities, I think we can potentially divide the Latin set of characters in four subcategories in the context of these questions. There is one category of code points or characters in Latin script which are ASCII—the regular English alphabet, for example (A-Z). So that's one subset.

There's another subset, which is plain characters. When I say "plain characters," it means it doesn't actually have a diacritic. But they're outside, for example, the core A-Z characters. And we've

actually been talking about one example of such character as the Eszett, or the German sharp “S,” if you recall. But there are more. So, for example, you actually have a Greek epsilon kind of letter in English and so on. So those are plain characters, but they’re outside A-Z, and they’re not in, for example, the ISO 3166 list.

So the third category of characters is decorated characters, so to speak, or these are plain characters plus a diacritic. But they are identified in root-zone LGR as a variant of a base character.

And then the fourth category is decorated characters, like, again, E with an accent or something like that. But they’re different from the third category in the sense that they’re not identified as variants of a base character. So decorative characters can themselves be of two types: one set or subset which is a variant of the plain ASCII, and the other one, which is not a variant of plain ASCII.

So when we’re making strings in the context of these questions, they can be made up of these four categories of letters. I’m not suggesting anything but just categorizing them just to, I guess, go through this discussion more clearly. Thank you.

DONNA AUSTIN:

Thanks, Sarmad.

Michael?

MICHAEL BAULAND: Thanks. I'm of similar opinion to Jeff here. We should not just look at the Latin script and do any special rules for the Latin script. If we disallow two-letter TLDs for ASCII, in general this should also be the case for Cyrillic and maybe some other scripts. So my opinion is to not disallow this in general but to say you can apply for them, but whenever they are confusingly similar to ASCII two-letter labels, then this will be rejected. And this rule holds for all scripts, whether it be Latin or Cyrillic or even some remote scripts that have similar shapes, like an "O" or an "I" or an "L". So there should be a general rule just based on the confusing similarity to two-letter ASCII but not a specific Latin rule. Thanks.

DONNA AUSTIN: Thanks, Michael. So I guess the question I have is, what would be the confusingly similar test? I think I could have a go and work it out pretty quickly, but it's something we might be hopeful to have in implementation guidance about what would be the test for that.

Okay. So, Ariel, do you have more?

ARIEL LIANG: Yes, I actually do. I just want to confirm my understanding based on what Sarmad mentioned and I think was maybe alluded to by Dennis in the chat. So I think the answer to Sub-Question 3 may be not as simple as I put it because what they said is that a two-character decorated Latin script may not be a variant of two-letter ASCII label. That could be the case. So I think that's maybe what they're trying to tell me, if I understand it correctly.

But I just want to just emphasize that the thing is, in terms in string similarity review, any two-character gTLD application will be compared for visual similarity against any two-letter ASCII combination. So basically all these two-character applications in whatever scripts they are have to go through that comparison against ASCII combinations. And if they have any visual similarity, not matter what scripts they are in, they're going to be caught in the string similarity review, and they won't pass the evaluation. So that's something I want to emphasize, and I think that's what Steve also put in the chat.

So there's some questions we want to ask the group. So the first question is, should the group develop a recommendation to explicitly prohibit applications for two-letter gTLD strings in the Latin script for the future round? So that's the first question we want to ask.

And the second question is related to Jeff's proposal and also related to what Michael said. Latin script is not the only one that could potentially cause confusion for two-character ASCII combinations, there could be other scripts as well. So Jeff provided some examples in Cyrillic, Malayalam, Hebrew, Ethiopic scripts, and Gujarati script. So these are some examples he provided in the e-mail. I very much appreciate that.

So basically he's proposing to the group to adopt a similar recommendation, like the one for single-character TLD, which is that, for the single-character TLD, it can only be allowed for limited script and language combinations, where the character is an ideograph and do not introduce confusion risks that rise above commonplace similarities. So he's wondering whether the group

should develop a recommendation for two-character labels as well—so basically only allow two-character applications for limited scripts and language that would not cause confusion risks with two-letter ASCII combinations. I think that's what he's getting at.

So basically there's two questions for the group. One is, should we develop a recommendation that explicitly prohibits applications for two-letter strings in the Latin script. That's the first one. And the second question is, should the group develop a recommendation to further limit the possibility of two-character applications—only limited to certain scripts and languages?

And that's all from me.

DONNA AUSTIN:

Thanks, Ariel.

So thoughts on Ariel's first question?

Sarmad?

SARMAD HUSSAIN:

So I'm actually just thinking if we have to list the scripts, for example. In the single-character case, the baseline was that they have to be ideographic characters. In this particular case, are we going to, for example, do a baseline—for example, say, all alphabetic scripts? Because there are different kinds of scripts, alphabetic versus [inaudible] and ideographic and syllabaries. Or are we going to arbitrarily set this list?

So just looking at the list provided in the previous slide, for example, just the first one, Cyrillic, along with the Greek and Armenian and possibly Georgian, those are the only alphabetic scripts here. Malayalam, Hebrew, Ethiopic, and Gujarati are not really alphabetic scripts. So if you're going into those other scripts—all kinds of scripts—then we are basically ... Then, what is the criteria for this selection? Is it arbitrary, or are we selecting based on some criteria? I guess I'm just trying to figure out what that criteria could be. And if there are no criteria then it generally extends to all the scripts.

The other thing which I wanted to note is that the string similarity will automatically take care of—or string confusion kinds of issues. And because of that, of the example with is shown for Gujarati or Ethiopic, some of those things are already addressed under the current, for example, recommendations. Thank you.

DONNA AUSTIN: Thanks, Sarmad.

So Michael and then Jeff.

MICHAEL BAULAND: Thanks. So to the question on page 13, I would say no, unless other scripts are included, too. Like Sarmad said, maybe extend that to all alphabetical script. Then I could agree with that. If it's just Latin, I don't see a reason why Latin should be handled differently. It contains the ASCII letters, but there are so many more letters which are quite different from the ASCII letters, and I don't see a reason to restrict them all just because some of the

letters are bad, so to say. Just let the string similarity review to handle all the cases and have no restriction in the first place. Thanks.

DONNA AUSTIN: Thanks, Michael.

Jeff?

JEFF NEUMAN: I kind of agree that we should never have any kind of role that applies to only one script. I think what we're trying to avoid is confusion with the ccTLDs. So that's the whole purpose for reserving the two-characters whether or not they are in use by the country codes.

Under the same kind of theory, though, is it possible in the future that country codes will be open for two-character IDNs? We don't know. Maybe at some point countries may want to indicate that they are two-character IDNs.

For me, applying rules uniformly seems like the way to go. So if you make the rule for single and two-characters the same, it's easier to implement, and it's easier for people to understand. But if you're not going to make them exactly the same, which is the proposal I had, then I agree that it's just confusingly similar and you don't need to single out any scripts because, as Sarmad and Michael said, the examples on the slide after this are pretty clear that they are or could be confusingly similar to ... And I used existing ccTLDs because there's an "OM," there's ... If you go to

the next slide, I think each of them correspond to existing ccTLDs. But perhaps not. Actually, the Hebrew one may not. I just thought that was ... Being Jewish all, seeing something like “oi” is pretty funny to me.

So, anyway, I think the point is you should never have a rule that just singles out a particular script. Thanks.

DONNA AUSTIN:

Thanks, Jeff.

Thoughts from others?

Dennis?

DENNIS TAN:

Thank you, Donna. I think I’m going to side with Jeff and Michael. I think I’m going to echo what has been discussed here. I want to take us back a little bit to the single-character analogy. I don’t think that’s the right way to see it. Unlike single-characters, that’s something new/novel—[inaudible] to put into the root zone. Also, being a single character, there’s less context to compare to. So I think we are okay with putting more restrictions, taking the conservative approach.

Talking about two-character IDN labels, we have other things to consider, but, yes, I agree that any policy should not single out a single script. The example showing Latin, Cyrillic, Greek, all might be confusingly similar. So I think, in this case, I’m inclined to go in through the simplest approach, which could be to not single out or

add more complexity to the rules but go by the confusing similarity evaluation process. And that should take care of the potential problems on string similarity. So I'll leave it there. Thank you.

DONNA AUSTIN: Thanks, Dennis.

Ariel?

ARIEL LIANG: I see Jeff still has his hand up. I wonder whether that's a new hand.

JEFF NEUMAN: No, sorry. Old hand.

ARIEL LIANG: Oh. Thanks, Jeff. And so this is just some staff observations I'd like to share, also similar to what other folks have mentioned. So basically, the current string similarity review rule is going to take care of any two-character or two-letter applications because they will be compared against all the ASCII string combination, no matter what script the applications are in.

So basically, that could take care of the case. And if the group does develop a recommendation based on Jeff's proposal, basically the benefit I would see is that the applicant would be able to get clarity on what is allowed or not allowed. And then, if some applicants wants to apply for some two-letter application,

then they have to think a little further to consider whether that could cause any confusion similarity with two-letter ASCII combinations. So maybe they will be able to save the time and energy and money not to submit that kind of application for the program. So that will be the benefit of this recommendation.

But if we don't have this recommendation, it's still an okay option because the string similarity review standards or criteria is going to take care of this kind confusion similarity and not allow those applications to pass. So it's no harm. But of course, there may be some applicant that may waste money and time submitting an application that will never have a chance to succeed.

So that's my personal observation. And maybe that could help the group decide what to do in terms of the proposal from Jeff.

DONNA AUSTIN:

Thanks, Ariel. Ariel, can we just go back to your previous slide, please? So can I—and this is something I don't have clear in my mind ... But is there a distinction here between a two-letter gTLD string and a two-character gTLD string? And if there is, what is it, and how do we deal with that? So I just want to be clear in my mind that, when [inaudible] confusing terms here and whether it matters that we are confusing terms. So I just want to be clear on that. So, Ariel, did you want to respond to that?

ARIEL LIANG:

Yeah, I'd like to, but I know there are linguists and the people like Sarmad and Michael and others. They may be better qualified to

answer this question, but I did do some research on my end to understand the distinction. So I'd like to share my findings.

So, basically, character (the term) means symbols. So it's a broader term that encompasses alphabets and also non-alphabetic symbols, like the ideographic symbols, like in Han and other scripts. So it's a very broad term that encompasses almost everything.

But then, for letters, that's really just related to alphabets. So the alphabetic-based language use letters, and we refer to them as letters instead of characters.

So basically, all of the letters are characters, but not all of the characters are letters.

And if we use a two-character string, that encompasses two-letter strings, but it also includes two-character strings in Chinese, Japanese, and other ideographic language, for example.

So that's my understand and indeed what Justine captured. Character is being than letter but encompasses all the letters.

And I see Sarmad has put a link in the chat. And I welcome other experts to chime in on this terminology matter.

DONNA AUSTIN:

Okay. Thanks, Ariel.

Jeff?

JEFF NEUMAN:

I think also, in SubPro, when we were talking about two letters (letter-letter), we envisioned brands or others being able to apply for dot-3M, for example, because that's their brand. So we didn't say two-characters. We said letter-letter combinations.

Now, I'd understand that, for whatever reason, domain names can't end in a number, I think, so I don't think you could have dot-02.

But, anyway, the point was we didn't want to necessarily have the rule being that all two-characters are reserved for country codes. It's that just two letter-letter combinations are, in case you can use numbers or whatever.

But the other point I have too is—and I have to look back—I know I just said that two-characters are reserved, but is that the terminology we use? Or do we say that two-letter (letter-letter) can't be applied for? And that makes a difference because, when you do string similarity, you're only checking it against the reserved terms. So if two-characters aren't reserved, then it's not part of the string similarity. It's an important distinction.

So I think we need to be very careful as to how we say things because it's not automatically ... If you go back to the last slide, for example, with the Hebrew example I put on there, if we say, "OY," for example (letter-letter) you can't apply for, then that Hebrew application for what looks like a "Y," which is really actual Hebrew letters, which I think is [inaudible] ... That won't checked in a string similarity. So we need to be very careful as to what the words actually say.

DONNA AUSTIN: So, Jeff, just so I understand what you're saying, two (letter-letter) for a Latin script we probably shouldn't allow, but did you just suggest that, for the Hebrew script, that might not hold? So the two (letter-letter) doesn't hold for Hebrew? I'm not sure that clearly understand.

JEFF NEUMAN: No, no. Sorry. Two different things. So the first concept I want to go back to is I'm not sure it's a safe assumption that everything will be caught up in a string similarity review unless we use the terms: "all two-character ASCII letter-letter are reserved." If that says reserved, then it is checked in a string similarity. If it just says you cannot apply for two-character ASCII strings, then that may not be in the string similarity review. So I just wanted to double-check that.

But in SubPro, I also want to make the point, as Emily posts here, that I think we say letter-letter combinations ... What does it say there? Okay, so it is reserved.

So it's not two-characters. I don't want to use the term "two-characters" for anything because someone could apply for, like is said, dot-3M or something like that. We don't want [play] with that. We don't want to be too broad.

DONNA AUSTIN: Okay. So there's a fair amount of stuff going in chat. Okay, thanks, Jeff.

So, Steve?

STEVE CHAN:

Thanks, Donna. Just trying to catch up on the chat and make sure what I was about to say isn't already covered. I don't think it is. So to the question about whether ... I'll say it differently. So it says in the 2012 Applicant Guidebook that any two-character IDN strings will not be approved if they are visually similar to any possible two-character ASCII combination.

And so what I was trying to say in the Zoom chat a little while ago is that the existing string similarity in the 2012 Applicant Guidebook is actually more expansive than the question we're looking at. So that is inclusive of Latin but also inclusive of all the scripts that we're looking at on the screen right now, as well as any other two-character IDN strings that would be confusingly similar with two-character ASCII combinations.

So I think the question that the group . Actually, one other thing is that standard for review—any two-character IDN strings that are visually similar—was reaffirmed by SubPro. So that standard that I just mentioned from 2012 should continue in the future. And so that's a string similarity standard.

And so I think the question that is before the group now is whether or not it needs to go further and potentially disallow the application entirely. I think that's the question that was on the previous slide. So it's not just reliant on string similarity to prevent the delegation of IDN two-character strings that are too similar to two-character ASCII. Is it enough to just rely on string similarity, but does there

need to be also some sort of ban, which I think is, like I said, on the previous slide?

So I hope that helps. So there is a standard in the AGB. It is reaffirmed by SubPro. Thanks.

DONNA AUSTIN: Steve, sorry. Can you just say what that standard from the 2012 Applicant Guidebook is again, please?

STEVE CHAN: Sure. So it says, "Applied-for gTLD strings and IDN scripts must be composed of two or more visually distinct characters in the script as appropriate. Note, however, that a two-character IDN string will not be approved if it is visually similar to any one-character label in any script." Or the second sub-bullet is: "It is visually similar to any possible two-character ASCII combination." So that's the existing 2012 AGB language. And then the SubPro final report reaffirms that same standard.

DONNA AUSTIN: Okay, thanks, Steve. I don't know. In my mind, it seems like that standard would hold for what we're trying to wrap our heads around here.

Jeff?

JEFF NEUMAN: Sorry. Old hand.

DONNA AUSTIN: Oh, okay. I think we're—Edmon, go ahead.

EDMON CHUNG: Sorry. This is something different from the active discussion here. I don't know whether I should bring it up now. We probably won't have time to talk about it, but I'll quickly bring it up before I forget.

This is related ... Sorry. I'm speaking in a personal capacity. I suppose many of us might remember the extended evaluation process for the IDN ccTLDs when the dot-Bulgaria [situation] came up and it was similar to dot-br. For completeness' sake, I think this might need to at least take a look at it and make a note saying whether, for gTLDs, such an extended evaluation will be acceptable or will not be acceptable.

So I think we probably ran out of time to talk about this time, but this is something that should be on our radar screen.

DONNA AUSTIN: Thanks, Edmon. And just to confirm, that was something that happened in the IDN fast-track process, not part of the 2012 gTLD [process]?

EDMON CHUNG: Correct.

DONNA AUSTIN: Okay. Alright.

Jeff, go ahead.

JEFF NEUMAN: On that, Edmon, we did review it in SubPro, and we decided that, because the cc process is so different and because there are no means of filing objections and there's lots of other parts in our policy, we didn't think it could easily apply, although I think there were a couple things we may have taken out of it in some of our recommendations. But I do remember we absolutely reviewed it. But because of the differences in the processes, we just didn't think we could use that.

DONNA AUSTIN: Okay, thanks, Jeff.

So thanks for the discussion, everybody. It seems to me that the language from the 2012 Applicant Guidebook and confirmation by the SubPro work probably covers off what we're trying to address here. So I guess I need a little bit more time to look at it, but my sense is that we already have a recommendation or recommendations from SubPro that we can just confirm in response to this question—that the two-letter Latin script will be addressed by those recommendations. So if that seems like a reasonable approach, if people agree with that ... And, similarly, I think there's a sense that the possibility of confusion will be picked up in the string similarity review anyway. So I don't know that we really need to create a recommendation that's any different from what already exists from SubPro.

So does that seem reasonable? We can draft up some language around that. But I think there are policy recommendations that already exist that we can just adopt.

Okay. So I think there's support for that. Obviously, until you see the language, it's always difficult to confirm. So we'll go ahead and draft language along those lines to address this charter question, but I think we're good with this conversation.

So do we have anything else, Ariel? I see we're almost at time, anyway.

ARIEL LIANG:

Well, we don't have anything. That's all.

And I just want to confirm that staff could draft an answer to the charter question, but it seems like we don't necessarily need to develop a specific recommendation because SubPro already affirmed the standard for a string similarity review, which should take care of any two-character applications that can cause confusing similarity to ASCII labels. Is that the correct understanding?

DONNA AUSTIN:

I guess it's a distinction that we may not ... So the policy recommendation may just be that the IDN EPDP reconfirms Recommendation X from SubPro. And that addresses his question. So we can do it that way. I think we need to call it out explicitly. But whether it's a recommendation or whether we call it something else I don't know.

And I see there's lots of chat about whether digits are allowed in top-level domains.

Okay. Something that I was remiss about was at the beginning of the meeting with changes to SOIs. So some of you already know that I am leaving GoDaddy or I have left GoDaddy. It doesn't mean that I'm leaving this group or my chairing responsibilities. I will stay on until you guys don't want me anymore. So that won't change my chairing of this group. I have spoken to Sebastien Ducos, who is Chair of the GNSO Council, and I have the support the council leadership to continue in this role, which I intend to do.

So with that, I think we can end the recording, Devan.

[END OF TRANSCRIPTION]