

DEVAN REED:

Good morning, good afternoon, and good evening for the recording. This is Devon Reed. Welcome to the Latin Script Diacritics PDP taking place on Wednesday, 26 March 2025 at 13:00 UTC. We have apologies today from Prudence Malinki. 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 with your statement of interest, please email the GNSO Secretariat. All documentation and information can be found on the Latin Script Diacritics wiki space. Recordings will be posted shortly after the end of the call. Please remember to state your name before speaking for the transcript. Please note all chat sessions are being archived. As a reminder, participation in ICANN, including this session, is governed by the ICANN expected standards of behavior and the ICANN community anti-harassment policy. Thank you and over to our chair, Michael Bauland, please begin.

MICHAEL BAULAND:

Yes, hello and welcome. Could you share the slides please? So, this is our first meeting. I was a bit surprised about that because I thought we already had a meeting two weeks ago in Seattle, but obviously it seems that the meetings at ICANN meetings count differently from the ones we do remotely. So, this is our actual first remote meeting. Yeah, welcome everybody to the LD PDP, the Latin Script Diacritics without the S. Next slide, please. So, yeah, Bill says, like a true computer scientist, we should count with zero, start with zero, and then the first meeting is number zero. That could work too. So, yeah, we will go

---

*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.*

through this agenda today, some welcome and updates since the last meeting, and then we'll dive into the scope management discussions, and then we'll tell you about some next steps and finally have an AOB if there are any. Next slide, please. So, yeah, this one. We won't be showing the membership list each meeting, but since we're still at the start and we also have three new members since last meeting, we wanted to show you the list again. The new members, as far as I can see, are Harald Arcos, Edmond Chung, and Rudi Nolde. Just looking, but it seems like none of them is actually in this call right now. So, yeah, just members on the list so far, but maybe they will join later. Also, we have Sarmad and Pitinan as ICANN work staff, and Ariel, of course, not to forget her. It was always great to work with you on the IDN ePDP and lucky to have you on board for this PDP, too. Next slide, please. So, for the updates since the last ICANN meeting, we sent out the early input request to the community leaders on March the 20th and are requesting input by April 24th. As discussed in detail in Seattle, this is a voluntary input. The community groups are invited to provide any input they like, but it's not mandatory. All communities can also take part via their representatives and that's why this way provides input to the PDP. Next slide, please. So, then let's start with the scope and I'll hand over to John for this slide.

JOHN EMERY:

Thank you so much, Michael. Obviously, a big discussion today is and kind of going forward is going to be the scope of our work. So, we figured the best way to kick it off is to talk is to talk about what is in the PDP Working Group Charter. So, as a reminder, our objective is to determine the limited circumstances in which a base ASCII gTLD and the

Latin script diacritic version of the gTLD can be simultaneously delegated. So, we're going to go through in a bit more detail, but our scope, according to the charter, is that this PDP is limited to examining a single issue in circumstances where the base ASCII gTLD and the Latin script diacritic version of the gTLD are not variants of each other. What mechanism is needed in order to allow a single registry operator to simultaneously operate both gTLDs? So, a presumption for this issue is that the ASCII and Latin script diacritic have a non-negligible chance of being determined to be visually confusingly similar. So, for clarification, something not in scope, we will use the Latin RZLGR as one of the relevant baseline foundational documents, especially in discussing scope today. But we also must understand that this work is completed by the Latin generation panel, and we should understand the rationale and impact of this, but a goal is not to go in and kind of change or relitigate this. So, to kind of go into a bit more detail about each section, the base ASCII gTLD and Latin script diacritic version of the gTLD are not variants of each other. So, it is not within our scope to discuss the list variants, non-variants, evaluate, and amend the variant lists. So, in essence, the Latin RZLGR exists. And so, we must understand that work as much as possible and understand that rationale, but not necessarily relitigate what's going on there. So, Michael, I'll hand it back to you to talk through a little bit more about the scope and some examples from the Latin, the RZLGR.

MICHAEL BAULAND:

Thanks, John. Yeah. So, we have now several slides containing the existing variant sets of the root zone LGR for the Latin script. And we won't go into detail for all these relationships. You can always look them

up on the root zone LGR webpage yourself. Maybe Stephen or John can put a link to the root zone LGR webpage, but it can easily be found by just searching for ICANN and root zone LGR. I think it's the first hit for that. So, basically, we have several variant sets. And each of those characters in this set are variants of each other, since the variant relationship is transitive. So, if you have, for example, the first line that's A with the code 061 and the A with an accent with OOE1, if they are variants, and then we have the second line, that OO61 and O3AC are variants, as a consequence, automatically also this OOE1 and O3AC are variants. This is why, even though this variant set just has five members, the list is quite long because each of those is a variant of each other. Yeah. Next slide, please.

So, there's no need to really discuss all these variants. You should just take a look and see that they exist and they are, so to say, out of bounds for our PDP. These have been defined variants. We won't change any variant relationships here, and we will also not be able to add any variant relationships here. Next slide, please. Here is a bigger set which contains N and characters similar to N, both in script and cross script. You see in the comments here or on the previous slide that some of those are labeled as cross script variants and others are labeled, for example, required for integration. Those are cases that have not been considered variants by either of the generation panels, but due to transitivity, they had to have been added to the variant sets nevertheless. Next slide, please. So, this is the variant set continued as it's got eight members. It's quite large. On the bottom, you see a variant where the N with a tilde and the N with, unfortunately, I don't know what it's called, the straight line. They have been considered variants by

the latent generation panel because they are almost identical depending on the font size and the font used and your eyesight, you won't see a difference or hardly any difference. So, the confusion risk is really high. For that reason, the generation panel decided to make them variants. Next slide, please. Here are some more examples for the O letter and related characters. Next slide, please. Continued. Next slide.

And similarly, we have on the bottom, you see again the O with a tilde and the straight line, which are variants for the same reason. And above that, you see that we have an O with an accent and an O with an, again, I don't know. I don't remember what it's called. It looks like an accent, but it's not really one. It's got a bit at the bottom, which makes it a different code point, different character. But again, these are so similar to each other. It's very, very difficult to see any difference. So, they have also been made variants of each other. Next slide, please.

And now we go to one of the questions that has been briefly addressed during the Seattle meeting, and it was also being discussed on the mailing list. By the way, in case you have any questions or comments, just raise your hand and interrupt me. I don't have to do a monologue here. You're all welcome to contribute and ask questions or comment or point out when I say something wrong. So, the question of the scope of this PDP is, on the one hand, is what is a diacritic? This is, some might think it's not so clearly defined, and any letter that looks like an ASCII character is a diacritic of that ASCII character, but that's actually not the case if you look at the Unicode definitions. So, the code point 0301, this is basically just the extent from lower left to upper right, and it can be combined with basically any character, more or less. But here are some examples where it's combined to some capital letters, but also to the

---

small letter A. And when this is combined with an A, which got the code point 0061, then there's a composed form, which has a new code point, which is 00E1. And that is the one that is actually used, but it can be decomposed into this basic ASCII character A and the acute accent character of 0301. Yeah, Mark, please.

MARK DATYSGELD:

Thank you. So, Tapani asked in the chat, does that include things like A conjunction E, SZ, and so on? The objective answer, if we're looking into the diacritics as they are understood by Unicode and how we implement Punycode, is no, it does not include those letters, exactly because they are letters. And that's the difference, the core difference here that I believe we will look into, which this is a diacritic exactly because you have this combining sign that goes into the first one, while an SZ is a letter. It doesn't, in spite of us functionally having a replacement for it in ASCII, which in many cases is SS, it is still a letter being replaced by two letters. While the case of diacritics, we have the letter plus a combining sign that's augmented. So, they are, at their core, what those different code points are, are Unicode's expression of the ASCII code points. And, you know, I know Bill answered to me that this depth of technical discussion wouldn't be at the core, but I think it has to be. We need to ground ourselves in the technical part first and evolve from there to see what is an exception, what is an accommodation, right? We need to start from the technical so that we can grow towards what would be an accommodation. That way we'll know, right? That way we'll know, these are the technical, so that we can grow towards what would be an accommodation. That way we'll

---

know, right? This is what Unicode says, this is what we can do within the limits of Punycode and so on. Back to you, Michael.

MICHAEL BAULAND:

Yes, thanks, Mark. That's exactly right. I would like to quickly share my own screen and show you the Unicode table or one of the Unicode tables. This is a table in the range of 008, 0 to 00FF. And we have some interesting code points in this range. For example, we have this A with an accent or the other direction, which we just saw on the slide. And if we look at the definition of this code point or the explanation, this 00E1, which is below in this list, we then see here that this 00E1 is actually equivalent to 0061, which is the ASCII A, and the 0301, which is just the diacritic. And the same holds true for the other A with the grave accent, which is a composition of 0061 and 030. And in contrast to that, if we look, for example, at this character with the code point 00F8 and scroll to the definition here, 00F8, that's the Latin small letter O with a stroke. You can easily see that it's also called O slash, but there is no diacritic in here. There's no slash, which would be considered as a diacritic and thereby can be combined an O with this diacritic to be that letter. So for that reason, this character Latin small letter O with a stroke is not considered to be a diacritic according to the Unicode table. And for that reason, it's out of scope for our work. And the same is the case, for example, for the character with the code point 00E6. If we then scroll to the definition 00E6, we see it's called Latin small letter A or also Latin small ligature. I don't know how to pronounce it. Sorry. A or also ash from old English ash. And there are some usages here, but again, it's not a combination of an ASCII character with some kind of diacritic. And therefore, again, this one is out of scope, even though I personally

---

would have liked it to be in scope, but that's unfortunately not the case. And we are very strictly bound by this scope given to us by the GNSO. If you look at the next character, which is the 00E7, it's a Latin small letter C with Cedillia. Here again, we have decomposition form, which is the 0063, which is the C combined with the 0327, which is the Cedillia below. And here again, this is in scope because, again, we have a basic ASCII character C combined with a diacritic. So this character is considered to be a diacritic. And this way we could basically go through all the characters to find out whether they are in scope or not. We don't need to do this because we most likely won't have to write a recommendation for every diacritic letter. But we hopefully will have recommendations that cover all diacritics letters together. And this Unicode table is also publicly available on the Unicode page. I'm not sure if I have the link available right now. But there are also other similar tables in Unicode for similar letters. So just look at this chat, which I've not been able to look at while presenting this. So, yeah, to Tapani's impression, we treat this A differently from the R with two dots. And also the O with a strikethrough is different from the O with two dots, because in both of these cases, the former one is not a diacritic. And therefore we are, so to say, not allowed to look at them, whereas the latter one is a diacritic. Sebastian?

SEBASTIEN DUCOS:

I'll put my camera on. So this is Sebastien Ducos, who early on with Michael and others sort of helped putting this on this track. I just wanted to go back to the reason why we call it a diacritic. And it's actually a fortuitous reason we just use diacritic, because that's the first word that came to my mind when I looked at a dictionary for a



definition of an E with an accent. I don't know. So I really appreciate the idea of a tightly scoped PDP. And I think that we should, for many reasons, keep to it. There's been strong discussions within the GNSO of tightly scoped PDP, particularly with DNS abuse. And if we can demonstrate that something like that can be done quickly and super efficiently, I think it will be an excellent signal for the future on many other topics. So far from me, the idea of deriving too much from it. But again, the word diacritic, I don't know. We may have put a lot of science around it afterwards. I don't know that it was chosen that scientifically in the beginning. And I don't know that the GNSO is that bound to it in the sense that we found, again, linguistic reasons around it. I don't know. Yeah, I don't know that the GNSO meant it to define such a hard border. But again, I'm in two ways, because I do support the idea of a tightly and efficient PDP to demonstrate, again, that it's feasible.

MICHAEL BAULAND:

OK, thanks, Seb. Yeah, I agree. And I also see what Tapani said there, that if we are required to restrict this to actual diacritics as they are defined in the Unicode, then this could or even will very likely lead to other people not being happy that their case is not addressed here. For example, Norwegian and Danes, which have characters like the O with a strikethrough. Seb said maybe it was not even intended by the GNSO to restrict this PDP this strictly. I don't know how to continue here. Maybe we'll have to have a discussion with the GNSO rep and prudence, unfortunately not available today, and leadership to see whether it is in the interest of this PDP to include these characters, which are not diacritics but behave in a very, very similar way, or whether that's not the case. Theo, please?

AMADEU ABRIL:

Hello. Well, I think that Discussing too much what's a diacritic in different contexts will not lead us that far, in the sense that it means different things, even in linguistic contexts in different languages. Now, my recommendation would be that we stay within the clear boundaries of what's diacritics as you defined them, Michael, because we all agree that this is the basic notion of diacritics, which does not prevent them later being applied to things like a ligature, etc. What I mean is, if we step outside of the clear boundaries, probably the result will be that whatever we do will not be taken into account for, let's say, constitutional reasons. If we do what we have to do within the boundaries, we have to be fixed. But we recommend that the same solution is applied somewhere else by GNSO. I think this will be more efficient than trying to solve all the problems, simply because we see problems beyond the scope that we have strictly being allocated. I don't know whether this makes some sense.

MICHAEL BAULAND:

Yeah, thanks, Amadeu. I think it makes sense in the sense that when we try to extend it to other characters, it's always problematic to decide which characters are in, which are out. And yeah, exactly, the new boundary with just stating diacritics, it's very clear. We have a way by the Unicode chart to check whether a character is in or out of scope. And so we'll have people that will apply for TLDs and will want to rely on the results of this PDP. They can also easily check whether this case is covered or not. If we say, get diacritics and literature and possibly a strike through and possibly a bit, then it gets a bit difficult to define. So

maybe, actually, the way forward is to just find a solution now for the diacritics. And if we have that and found it, then maybe through either another PDP, I don't know, or through some other mechanisms in ICANN, this result could be extended to further characters. Any other thoughts here? Comments? Yeah, well, then I think we can go back to the slides, please. Okay.

So yeah, we at least have the scope as it was defined in the scope of work document with diacritics clearly defined. Or if anybody is having questions regarding what is a diacritic or whether a certain character counts as a diacritic or not. If there are still some unclearness there, please let us know. Okay, so yeah, we agree what a diacritic is. The next step is, the next step is, do we want to stick with this strict definition of our scope or do you think it's required to extend this, which would most likely require us to have a discussion with GNSO? Marc, yes, please.

MARK DATYSGELD:

Marc speaking, just as a group member. Probably it's best that we stick to this definition because conceptually we will arrive at the proof of what can be done with this. And nothing really stops anybody from bringing to council, let's say, the Latin letters PDP idea and make an issue report around that. But at least we'll be solving, let's say, good part of the problem. And that's way better than we have today, even though I fully agree with Tapani when he mentions, you know, the impact this will have. But from what I remember in council, there wasn't really a champion for those concepts at the time. It's more so that nobody was really asking those questions at the time and the scope got consolidated in the way that it is over months under that pretense. But

one thing that I think is very possible that we do is that whatever report we put out already comes with what we see as the exceptions. What did we find that's relevant and important that did not fall under the scope of this? Because seeing this from the council side, that would be actually super useful to get something started fast, right? Instead of having to relitigate this over and over again. So, I agree to the methodology, but also pointing out that we don't need to discard those conclusions. We can actually roll them into our conclusions for what would be next steps or what is useful.

MICHAEL BAULAND:

Yes. Thanks, Mark. I think these are very good suggestions. And yeah, I guess we should now just stick with this unless there are any objections. I see none so far. Yeah, exactly. That's what I need. Let's try to make good rules for this very narrow scoped PDP. And then we or somebody else can then use these rules to extend this in a sensible way. Yeah. So, I guess we have one part of the scope clarified, which I think is nice. We could go to the next slide, please. Oh, Ariel, please.

ARIEL LIANG:

Thanks, Michael. And I hope my intervention is appropriate. It's just from ICANN org perspective, like to understand future implementation of this. I think I understand if this PDP develops some kind of rules applied to a diacritic slash base ASCII labels, it could be expanded to similar cases. But I wonder whether it will be beneficial for the group to clarify what those similar cases could be. I understand one of the examples is like the Norwegian Deng ladders. They're not strictly

---

diacritics, but could be confusingly similar. But I think just to understand the boundary of the expansion, it will be very helpful to clarify that and so that we know what other cases those rules can be applied to. But maybe it's not something that can be solved right now, but maybe something that can be a solved right now, but maybe something that can be a table and discussed by the group later to kind of limit the boundary of applicable cases.

MICHAEL BAULAND:

Thanks, Ariel. Yeah, the question would be, how do we do this? Do we list all characters, which is essentially possible because we are just looking at the repertoire of the Latin root zone LGR, which if I remember correctly, has something like less than 300 characters. So it's a lot of characters, but it's not like looking at thousands or millions of cases. And so that might... It might be an idea to go through that repertoire and mark the characters which we think could be included in a second approach, second round. Bill, please.

BILL JOURIS:

I think we really, really want to come up with something more like principles for how we approach these rather than going on a case-by-case basis. Not least because it's entirely conceivable that someone will come along from a language which uses the Latin script but isn't among the 200 out of 1,000 that the Latin generation panel looked at and go through the process to get some of the symbols in their language uses. And some of those have the same issues we're talking about here. I think we need to have a principle of if there is a diacritic, etc., etc., then

X is the exception process rather than just going through and saying, this one is an exception, we handle this way, this other one is an exception, we handle this other way, possibly not the same way. To my mind, that way lies madness.

MICHAEL BAULAND: Okay. So, just to make sure I understood you correctly, it's your suggestion that if we follow Ariel's suggestion to also at least list those other cases, that it might be problematic if we just go through the Latin repertoire because at some point the repertoire could be extended by some other process and then it's not defined whether those additionally added characters fall into this category or not?

BILL JOURIS: Yes, exactly. There's no problem with listing the current examples if we want to, but I think we need to establish a principle of how we deal with it rather than just handing back a list of specific cases. Thank you.

MICHAEL BAULAND: Okay, yeah, good point. Even though it's not very likely to happen soon, in theory it's definitely possible and we should probably also cover those cases here. Thanks. Amadio, please.

AMADEU ABRIL: Okay, I think that I am unmuted. I think we should shift a little bit the focus of how we approach this. Our subject matter is not Unicode code points or diacritics. Our subject matter is how to prevent user confusion

in case of simultaneous applications or TLD strings existing and proposed that a. are already considered confusingly similar in some way and b. are based in a pair or more than two base Latin script and some diacritics. What we need is a solution to solve the problem of whether these can coexist or not with the main idea of whether to avoid user confusion. Now, if we stick to each concrete diacritic and want to find a solution for each one of them, I think the only solution to say is that there is no solution. Unless we find some kind of good solution, then this good solution could work for others. It doesn't matter whether it leads to this other case of confusion, ligature, or even confusion between a Latin script and a Cyrillic or Greek script proposed TLD that may look very similar but certainly are not part of our scope. So the solution we may find or not may be applied to other areas, but I don't think we should focus ourselves on what other cases could be applied. Even similar and plural could perhaps benefit from the solutions we find. But our scope is just to take as granted that there is some cases in which a base Latin character and a diacritic, even narrowly defined, are considered to be confusingly similar and find a solution there. And the most, you know, usable solution for different cases that we can find. Not to find all the possible other cases. These will come somewhere else because, as Osiru has said, we cannot set in stone today all the other possible cases and certainly they are not limited to Latin script. But our scope is the one we have. The solution may be applied elsewhere. Our work has to be focused on what we have as a scope. Thanks.

MICHAEL BAULAND:

Okay. Thanks, Amadeu. I think I agree with you that we should find a solution for this limited scope and then see how this can possibly be

used for other cases as well. And as I said, maybe it even automatically extends to other cases. Or could extend to those cases not in our scope because our solution is so versatile that it could cover similar cases too. Yeah. Thanks for that. Okay. So, continuing with the scope slides here. We have some examples here. Like the Deja with or without diacritics. The Spanish Español and Espanol. Or the Portuguese Violao. Sorry, I don't speak Portuguese so I can't pronounce this correctly. Thanks. Yeah. But I guess this has already been covered by the previous discussion that we now exactly know what diacritics are. And so these examples that were also presented in the final issue report are certainly in scope. And they are definitely not exhaustive as there are other diacritics forms too. Next slide please. So I will hand over to John again to this slide. Thanks.

JOHN EMERY:

Thanks so much. So taking us through the next kind of details of the scope. What mechanism is needed in order to allow a single registry operator to simultaneously operate both gTLDs? So this part of the scope aligns with Charter Question 1. So under what circumstances should a base ASCII gTLD and the Latin script diacritic version of the gTLD be simultaneously delegated, if any? So this is one of our main questions in the Charter. If such circumstances exist, what measures should be put into place in order to mitigate potential for end user confusion? So what we have here is the potential considerations to allow or limit instances of ASCII and diacritic versions. So this is, as we noted earlier, definitely an exceptions process. And there is a question from Tapani in the chat. Charter Question 1 seems to allow simultaneous delegation to two different applicants, or do I misread it?



So we're going to be getting to that in a little bit, or Michael or Mark, if you wanted to chime in, because we do have that, whether you are an existing gTLD applying for the diacritic version, or if two are simultaneously applying, it would go through the normal process. But Mark or Michael, if you have anything to add to that.

MICHAEL BAULAND:

Yes, thanks. I think what Tapani meant is not that whether it's an existing versus one applicant or whether it's two applications, but it's that if two different applicants, so if the entity, two different entities could apply for this, and yeah, I'm not sure, I don't know if Steve maybe could chime in and say something about that.

STEVE CHAN:

Sure, thanks Michael, this is Steve from staff, and it's a good helpful question, especially when we're early, this early in the process. If you look at the way that the sentence is phrased in the in-scope section, you can see that it actually specifically calls out a single registry operator to simultaneously operate both gTLDs. So I think the presumption when drafting the charter was that these exceptions that would allow the simultaneous delegation of both the ASCII and Diacritic version would be to a single registry operator.

JOHN EMERY:

Great, thank you Steve, and thanks for raising the question as well, that's really helpful. And so some considerations to keep in mind when we're managing the scope, and this is all from the final issue report, to

kind of help us talk through some of these considerations as we go through this. Do we limit the same entity operating the ASCII and Diacritic versions? The limit to applied for IDN strings of existing ASCII gTLDs, where the existing ASCII gTLD is a workaround for the proper IDN string, and or vice versa. Limit to applied for ASCII strings that act as a workaround for existing IDN gTLDs. And if that happens to be the case, what do we mean by a workaround? These are all things we're going to have to discuss throughout the PDP process. For instance, other than for an absence or inclusion of Diacritic marks, the strings would otherwise be identical, like the Deja and Deja example. Also, this could mean that only fully decorated and fully ASCII versions are in scope, and that the solution would be limited to a maximum of two strings. That's something we have to consider as we go on and well. So this would exclude linguistic characteristics, like singular plurals, or alternative spellings like color and colored. These are things that have kind of already been established and not something that we would be concerned with in this PDP. And we also need to understand and talk through to include new gTLD applications in the future. So a single applicant begin to operate both ASCII and Diacritic versions, and also consider whether additional limitations are needed. For instance, does it only apply to certain application types? So community, geographic, dot brand, etc.

So the first case is those who have already registered an ASCII gTLD and wish to have the same letters with one or more Diacritics added. And for case one, theoretically, we could simply say they have one and they could get an additional gTLD, provided it uses exactly the same letters with Diacritics. And case two, if you had two applicants, those who

---

currently don't have an ASCII gTLD and you had two applicants, it would go through the new gTLD process, and it doesn't matter whether the Diacritic version has ever been used, and you would only get one. So in that case two, we would break it down a bit to where if an applicant wants both the ASCII and Diacritic version and nobody else is applying, we would treat it the same as case one. If there are two applicants that want to register Diacritic versions, then these would go into the contention sets within the new gTLD program. So, you know, overall these are a few considerations to think through as we go through what comes next. So these are some examples. This is very much a non-exhaustive list, but Michael, if you wanted to take us through this a little bit, these are some examples that were presented to GNSO Council from ICANN 78.

MICHAEL BAULAND:

Yeah, thanks, John. I'm just reading the chat. One second. You said, can we consider cases where only the Diacritic version is desirable, e.g. if the ASCII version has negative connotations, as it were. I'm not sure I understand that question, because if there is only a single version, then there is no contention at all and no rejection possible either. Could you maybe elaborate? Mark, do you want to respond to that?

MARK DATYSGELD:

Thank you, Michael. So, to Tapani's point, at that point, if one of the strings is, let's say, undesirable for any reason, then it's just a normal application of an IDN, right? You're just applying for a string in IDN, and unless you or somebody else wants to have that undesired string, it falls

---

under normal procedures. One thing that I have to remember myself, and I've been looking at this a lot, is that the normal way will still exist just fine. You can still just apply for an IDN. The thing that we're trying to look into here is, what if both are desirable? So, to answer to your question, then that would be a normal application, but yeah, it's something that I'm trying to put into my head as well. Thank you.

MICHAEL BAULAND:

Yeah, thanks, Mark. So, yeah, if you're looking at the slide here, if you would just apply for the lancome with this O and this, I should learn those critic names, I guess, this hat kind of thing, then, of course, you could apply for this and get it if the lancome just in ASCII does not exist yet. This is just a standard case that's already covered by the current rules, and there's no problem, and we don't need to make any recommendation or rules here in our PDP for that. The problem is really only if, looking at these examples, if both versions, the green background one and the yellow background one, so the only ASCII and the ones with their critics, they are both applied for either in the same round or in consecutive rounds, whereas if it's in consecutive rounds, then, of course, the first one will be allocated since there's no contention yet, and if then in a consecutive round the other one is applied for, then we would have to check what we do and what rules we want to apply for this to make that possible, and if it's in the same round, then, similarly, our rules would have to make sure that those two strings don't go into a contention set, which would result only one of them to be allocatable. So, yeah, here's just three examples of existing ASCII gTLDs and their potential applied for strings with their critics. There are possibly more, but these are just showing the French

examples as this list was introduced to the GNSO council. Yeah, next slide, please. And I hand over to John again.

JOHN EMERY:

Yeah. Thanks so much, Michael, and just pointing to Steve's discussion, The central presumption for this issue is that the ASCII and Latin script diacritics may have a non-negligible chance to be determined to be visually confusingly similar. So this is the issue of concern because in the non-variant scenario, even if the Latin GP deems the IDN and the ASCII to be distinguishable, they may still be determined to be visually confusingly similar during the string similarity review. As the diacritics only add a small change to the ASCII letter shape, this could obviously lead to end-user confusion. So since many existing gTLDs omitted the diacritics to adapt to the DNS, their correctly spelled IDN version, if applied for, will likely face such a challenge in the string similarity review, and it might be ineligible to proceed. So again, the kind of key for us here is that this is an exceptions process, so we really need to determine what that looks like throughout the process of this PDP. And I see, Bill, you have your hand raised. Bill, go ahead.

BILL JOURIS:

Perhaps someone has a better handle than I do on the string similarity review process, but does that process establish a general case of these two code points are confusingly similar, or do they just deal with the specific gTLD that is in question?

---

MICHAEL BAULAND:

I can try to answer that. So Jeff got some grasp of this process, as I'm also involved in part of that process. So it works in a way that the string similarity review panel will look at labels, TLD labels, and since the list of potential labels, especially with variants, since they also have to look at variants applied for strings. So it's not that important for Latin, but for other scripts maybe a bit more. So even if two applied for TLDs are not looking similar at all to each other, there might be the case that one of those TLDs has a variant which looks just like the other applied for TLD. Even if they do not apply for that variant, this might cause a problem, and the panel has to look at this. And for that reason, since the number of comparisons is potentially really high, there's a pre-process, so to say, that for each script, all code points have been classified into categories one through four or five, which tells how similar those code points to each other are, or the glyphs of the code points are. And this data will then be used in a software tool, which will help the review panel to pre-select potentially similar labels. And out of these labels, the panel will then decide whether it's actually potential to cause confusion or not, if that helps. Bill?

BILL JOURIS:

Yeah. Actually, I'm afraid it didn't. What I'm trying to ask is, say, for example, the panel looks at the two .Quebec cases and says, these are visually confusingly similar. Does that mean that E and E with a Q are now effectively blocked variants for anybody else who applies? Or does it apply just to that one case? That was what I was trying to ask.

---

MICHAEL BAULAND:

OK. It is just for that one case. It's very, very, very likely that this will also apply to other cases. And there's also an objection procedure possible. So if, for example, the panel decided this Quebec case to be confusingly similar and decided the cafe case, which also has the same E, not to be confusingly similar, it's most likely the case that you could raise an objection to that and have the panel re-evaluate. But it's not defined that they have to decide the same, at least that's what I know about the panel. I'm looking for individual cases.

AMADEU ABRIL:

Just to reiterate something that I explained in a reply to Bill in the mailing list. We need to take into account that the string evaluation evaluates strings, as the name says, not characters. The question is not whether an I and an I with an acute accent are similar per se, are variants, or are confusing per se, it's whether within a character set that composes a label that wants to be TLD are confusing. And take the example of one of the contention sets in 2012, Unicom and Unicorn RN. The problem is that, you know, it's not the same to look at one character individually or within, you know, a group of 12 characters where the other 11 are identical. So the likability that this is confusing to a consumer, I'm not saying that I agree with that decision. I just explained the decision was made that Unicom and Unicorn would not coexist. I personally disagree, but it doesn't matter. The question is that in a string evaluation, it's not just character by character, it's label against label. And the longer the label and the more similar the other characters, the more identical are the other characters and the positions, the more likelihood that they are similar. But indeed, the panel never said that RN and M are variants, that would be completely

stupid. They are not. They only said that in that case, those two strings were incompatible to coexist in the DNS. So this is what is judged. That concrete label against that other concrete label, no other general conclusion is taken from there.

MICHAEL BAULAND:

Yeah. Thanks, Amadeu. That's exactly the case. So we don't know how the panel will decide. And even if we know one decision, we do not know whether similar decisions will be taken in the similar way or not. Okay. Yeah. So here we can have some potentially visually confusingly similar strings. And as you see, this also is not necessarily language dependent. For example, the last one, this Haagendazs is just an invented word. It does not exist in any language or it exists in all languages, however you would like to see that. But still, it might be the case that they would like to have the ASCII string version and the one with an A. And if they apply for this, it's not unlikely or even likely that the string similarity review panel would mark those as being too confusing to coexist and thereby reject one of those. So, yeah. Next slide, please. I think we are done with this part of the scope finding mission for today. John, would you like to proceed here?

JOHN EMERY:

Sure. So kind of next steps, we'll go through the kind of timeline coming up as well. We will have our GNSO council liaison chat and the working group to also remind their respective groups to provide any timely feedback for the call for early input. If they have any, again, remember this is voluntary. So just a quick overview of our updated and proposed



project plan timeline. Right now we're seeking the early input review at the April council meeting where we hope that the GNSO council will confirm our project plan. We don't see any issues there. And then aiming for February 2026 to publish the initial report and then go through the additional PDP procedures. So Sebastian, I see you have your hand raised.

SEBASTIEN DUCOS:

So at the council wrap-up session, Prudence presented the idea of having two phases, including a very short one in the beginning, in order to essentially accept applications in the next round that would qualify for what we're doing here, but not having a fully developed and implemented policy that we wouldn't be able to process, but to accept these applications and park them until the policy that we're going to recommend is fully implemented. Is this still part of the program? Is this still something that we're talking about? Is this something I missed somewhere? I didn't quite understand where it came out of, and I know that that will be questioned in the next council meeting in April.

MICHAEL BAULAND:

So that was an idea that came up during the Seattle ICANN meeting, and it was discussed by ICANN staff, ICANN Legal, and several other entities. And while in the beginning, the idea sounded interesting and approachable, someone I don't exactly know, was it John or Steve, did some actual best case timeline review of this to – even if we do this phase one separation and are as fast as possible without any problems, it would still turn out that this report would only become policy by

---

November, I think it was. And that would still be too late for us to include it in the applicant guidebook. So the current approach is to look at this a different way and try to get something similar into the applicant guidebook, but not via this policy development process, but via some different mechanism. But the general idea is still that we try to have something in the applicant guidebook that allows parties to apply for their TLD in the next round, even if our policy development process has not finished by then, but that they will be sure to not get rejected right away, but they will have the chance to wait for the PDP outcome and make that applicable to them. Is there anything you would like to add, John?

JOHN EMERY: No, thanks.

SEBASTIEN DUCOS: Thank you very much for the explanation.

JOHN EMERY: Yeah, Sebastian. So essentially, the phased approach would be a moonshot and still impossible at that. So we're sticking with the conventional project plan now and working behind the scenes as well to determine kind of best steps forward through the applicant guidebook. So just a quick timeline, our more tactical, that was the big overview. Our next meeting, we'll continue to discuss scope management and understanding the topic a little bit more and deliberate that and discuss Charter Question 1. Just as an FYI, we will not be meeting the week of

the 16th. I'm sure you saw that in your calendar invites that you got the invites and then this was canceled due to the Easter holiday. So considerations, again, for everyone, this is available on the wiki space, and I'll go ahead and copy that into the chat for everyone. And also the slides are available on the wiki space as well. Some background documents for everyone to review. And we'll see as well if we can get the Unicode that Michael shared with us as well, either on list or on the wiki space, because I think that table was especially helpful for next time. And what we want to do next for AOB is just have Steve Chan take us through a little bit and kind of make this a normal procedure for the ICANN staff taking notes throughout the process, any kind of decisions or action items that came up today, just to remind us what to do. So, Steve, the floor is yours.

STEVE CHAN:

Thanks, John, this is Steve from staff, and this is a new process we put in place, which is partly why I actually forgot I suggested to do this, but I think we want to get into the good practice of getting agreement on the action items at the end of each of our meetings. It's just, I think, good housekeeping for us to make sure we're all on the same page. So I think there were a couple of outcomes, and then I think the next steps are really what the action items are, but in terms of outcomes, it seemed that there was general agreement to adhere to the charter and define Latin script diacritics per the Unicode, although there are some other ideas circulating about how it can be potentially made more flexible. And I think that was really the only outcome, and as you noted in the next steps, there's really some homework, I think, in reference to the ccTLD fast track, I think we would like to do a little bit of make that

---

homework easier for all of you to point you to the relevant parts of the ccTLD fast track, because at least from the staff perspective, I don't think all of it is necessarily relevant for our case. So I think that's all I had on my end. Thanks, John.

JOHN EMERY: All right, any questions, Michael? Anything you'd like to add at the end?

MICHAEL BAULAND: No, I think just to maybe answer Anil's question, so the next meeting will be next week, Wednesday, same time, though for the ones in Europe, it most likely feels like being one hour later because of the change to the summer time. Philippe.

PHILIPPE FOUQUART: I have a question. I didn't know where it belonged, but if we could go back to slide 26, I think, where you alluded to Haagen-Dazs, it's just a comment, not a suggestion, but maybe the thought for the AGB shortcut that you mentioned or even the policy moving forward. It seems to me that if you look at these examples, and I appreciate the number of them are in French, thank you. The first one, which is a common name, is indeed to be looked at. The other ones, like brand names or city names, to me, at least, are a no brainer, really. So maybe a way to simplify this and to avoid litigation, etc., would be to single out the former, the brand names and the city names to have a policy that would say just what you said earlier, i.e. that the registry would have to be the same to avoid confusion for the Internet user. Just wanted to

---

share this because I didn't have the opportunity to say this without derailing the meeting. So I'll stop here. And back to you, Michael.

MICHAEL BAULAND: I'm not quite sure I understand your comment. Do you want to handle words which are kind of real-life dictionary language words differently from like fantasy brand names or city names?

PHILIPPE FOUQUART: That's another way to put it. I was singling out proper nouns because for proper nouns, I don't think there's an issue. For city names, I don't think there's an issue there. It's confusing for internet users that it would have to be the same registry. For that, that would be another way to put it.

MARK DATYSGELD: I get the point that Philippe is talking about. And that is mentioned previously as a possible consideration in this very slide deck. What would be additional considerations? Let's call it that. There it says perhaps only certain application types. But it could well be that there's some nuance there as in if we were to have talks to the GAC, for example, and come to the conclusion that if a geographic location just wants to assert both forms in which it's mainly expressed, then that should be fine. That's the sort of thing that at least in my mind is within scope. But that is something that remains to be seen, of course. But in my head, it is within scope for us to discuss that. So at least that's how I see it.

AMADEU ABRIL:

Very short regarding this question. I don't know how this can be handled. Because the string evaluation comes at the very beginning of the evaluation. And it looks at the string. So it doesn't look at whether this has a brand behind or a community TLD, et cetera. If we want that, then we will need to ask that lots of things in the application guidebook are changed. And at the end, what they are talking here about the strings. Indeed, the context is important. The usage or the real confusion. We may take that into account. But if we want to single out something that proper names, what do you do with coin names? I mean, com is not a proper noun. It's just an abbreviation of something or like org. Could be organization or orgasm. Who knows? Right? So it doesn't mean something completely. So sorry. I think that we need to stick to applications for strings that want to become TLDs.

MICHAEL BAULAND:

I think I tend to agree with you. I'm not sure it makes sense to distinguish labels by the way they are used. But we can certainly have a follow up discussion on the next call. I think with just three minutes left, we shouldn't open up this point. But we'll take it into consideration and maybe revisit it in the next or the next calls. That's okay. Okay. Any other business? Any last questions for the first meeting? If not, you'll have more chances in the next meeting. So yeah. Thank you everybody for taking part and looking forward to working with you during the rest of the PDP and talk to you next week and have a nice morning, afternoon, evening, night. Thanks.

EN

---

[END OF TRANSCRIPTION]