TRANSCRIPT

IDN PDP Working Group 1 Meeting
Costa Rica 15 March 2012

Attendees:

Lyman Chapin, Technical Community
Edmon Chung, .asia
Hiro Hotta, .jp
Manal Ismail, GAC
Cheryl Langdon-Orr, ALAC
Vaggelis Segredakis, .gr
Giovanni Seppia, .eu

ICANN Staff:

Bart Boswinkel
Kristina Nordström
Gabriella Schittek

Apologies:

Jaap Akkehuis, Expert on Standardisation
Patrik Fältström, SSAC

Bart Boswinkel: What I'm trying to show is capture some of the elements we discussed on the call prior to leading up to the Costa Rica meeting. I've shared this with Chris but he wasn't able to respond. So what I'll do is run through -- what I've included are some major questions and actually -- so I'll scroll up and then you'll see. This is based on, say, changing a major role in -- (inaudible) Is this readable? That's good.

Unidentified Participant: (inaudible)

Bart Boswinkel: Can you read it this way? Yes?

Kristina Nordström: Hello? Sorry. Can I just remind everybody to speak closely into the mic because we have (inaudible) on the phone? Thank you.

Bart Boswinkel: The quarter leading into this meeting we discussed the possibility to replace the common goal in the past record which is captured in the overall policy as well which is some confusion exists (inaudible) resembles another vision that is likely to receive or cause confusion and those the major role in the past record which

was when we started the whole process of the IDN CCTLD overall policy was more or less captured. So, overtime what became clear is this was a very complex area and that's one of the good things of the process at least in my view, is that it will show reasonable limited environments. Now, if you're going to included into the overall policy you probably need to address some of the issues that have been that were encountered in the whole process.

So, what is proposed is a definition that is from the variant issue project because in the variant issue project there is also a definition that relates to visual complete ability and so the role that is a little bit additive but the core is itself is that the proposed rule will be as discussed visually confusable strings referred to two different strings of Unicode characters which apparently come from in small sizes, a typical string resolution. It's so visually close that people who are unfamiliar with this script easily confuse one or the other.

Now I don't know -- at least from my perspective as a non -- as an ASCII -- at least it makes it understandable what is meant by confusing similarities. In makes it clear in my view to see what is needed to be expressed. So, that was one of the reasons to try and include it -- and I think, I don't know to what extent it covers the same area of confusing similarities as and say with the current role.

And so what I've done is I've concluded in this as we discussed some questions related to this definition in order to understand whether this is indeed the role and a way for what that we can work on this proposed rule based on confuse ability. So, I'll go down to some of the questions and maybe in the discussion we had in the small group. it is recorded it will be transcribed that we shared again on the lists that at least now we can see a way forward.

So I think the first question that was discussed at earlier stages as well in small group which of the following cases of confused ability are considered relevant and should be addressed. And I've listed my suggestion to the group is that there are just specific cases of confuse ability in the form of what the people we refer to that we start to include explicitly in the overall role and this is where I think I've listed three types of user confusion or users who could be confused by the similarities. I want to say first of all is this listing, are these the logical categories? So, please read it and after the meeting I will send forwarded documents as well. (inaudible)

Edmon Chung: Edmon Chung speaking.   And curious actually to get a sense of what exactly is being asked. The reason why I want to clarify that is there's a lot of discussion right now about strings similarities with you and I guess that's a topic that was brought up (inaudible) in the VIP discussion. Related change to the IDN CCTLD field balsa pertained to this issue. My question with that is, is there any value for some consistency across the community I guess both for the new facilities and IDN CCTLDs on this issue and how we define it rather than if we create a certain definition ARRA description for and it's somewhat different than others. That may create -- sorry for use in the same words that basically some confusion around the community.

Bart Boswinkel: If you look at it the way we discussed -- this is been the discussion of first of all to a small group in the IDN working group and it's been going on since San Francisco. This is at this stage the major issue in the overall selection with the exception of the IDN variant. You will see the council excepted the recommendation of the working group not to move forward with variants because that has been sorted out. It moves forward. I think that there is -- how should I say? There is a responsibility of the working group to seek a possible solution based on the experience of the Fast Track Process that's varied with structure the way it was. They think again if the JIG has some (inaudible) advice because

there is this working group that has its own responsibilities regarding the IDN CCTLDs. That's one.

I think if you look across the board as a result of staff check process say the CCTLD community are people involved in rereading and encountered problems that do in fact due to the experimental nature of the staff check process that provides input into the overall policy. That's the way it was structured. And that's the reason and that's why the overall policy takes so long. You notice we've learned from what is happened in the staff check process.

It's the other way around as well. If you look at the definitions as far as I know because I really understand the -- I've never really been directly involved in the African guidebook. I don't know what's happening in the African guidebook. In that sense the only thing I know is the use the same description of say the basic rule in the African guidebook is confusing similarities. It's the same basic role as in the staff check. And knowing that you have already these encountered issues in the staff check role and then it goes back to The responsibility of this working group and other people in the world to set to internal policy. There is this case. I would wait until we've gained experience of the first batch. That will take another one and a half years.

Or try to move forward because there will be public ramifications that affect this group again in order to sort out some of the issues and some of the good thing is that's why I'm very grateful that Liam in Frankfurt has kept his eye on this group. but because this is a small group and focusing on this particular issue we tried to create a bridge between the one side where'd it's happening in the policy and on the other side what is happening in the views of the second operational community. Because this is a really complex issue and to drill this down this is intractable and sorted for the overall IDN policy. I think it may be the other way around and looking forward and everybody's responsible and we have a way forward we have a other areas as well. So that was the answer to your shirt question.

Any other questions, yes?

Manal Ismail:	My name is Manal. Just a couple of things. First of all, the examples you're mentioning here are basically depending on similarity to the three scripts. I wondered where a case, for example, this is confusing review tools, some encoding issues with the Unicode. There are such cases in the Arabic script where we have three different Unicode but that are identical. So, where is the specific -- ?
Bart Boswinkel:	My understanding is -- please correct me if I'm wrong -- and I think I used the old language is the conclusion you're referring should be address with the IDN tables which Unicode characters do you allow to be registered. This is -- this looks into the confusion that what we have been discussing as examples of the lasts couple of months is on the one hand side, the application requests on Bulgaria and from Greece. In one case I think maybe I'm not allowed to say but I will quote the chairman of this working group by saying I think there was a general understanding or sense or acknowledgement that one of these strings in Cyrillic was confusingly similar with another one. And there was also an implicit sense that maybe the outcome of the review of the Greek string that went according to the criteria and according to the rule and according to the criteria ended up in the result that was unexpected. So, this is across scripts what we're looking in. I think if you go back to your question is that the whole issue of IDN rules, which characters do you allow is more within the script. But that's my understanding and I could be completely off track here.

Manal Ismail:	I'm sorry -- you have to excuse my confusion. I can take this offline if I'm confusing anyone else. Maybe my comment was more from an end user point of view trying to find a domain name and was not able to reach the domain despite the fact of he typed exactly what he saw. But when it comes to the language, I think this must be (inaudible) right?
Unidentified Participant:	(inaudible) Which characters are used -- that's my understanding. Which characters we allow.
Manal Ismail:	Yes. I completely allow only Arabic language Unicode for my registry. And then someone from Iran for example typed into the search engine which allows beyond the Unicode characters from the Arabic script and then he's not -- although I'm not allowing this wide range of scripts, again he doesn't know. He's typing exactly the domain name in front of his and he's not getting anywhere. Does this make any sense?
Unidentified Participant:	Not for me. Maybe Lyman? (inaudible)
Bart Boswinkel:	What you're referring to -- sorry. I almost have to bend over. (inaudible) this way. What I understand is -- the way I understand your question is of confusion within one script. And what we're doing here right now is more or less the confusion across scripts. So, between -- and this has to do with the string that has been requested by the Africans. And whether this is confusingly similar to an already existing string in another script and say in the case of two codes and that's a confusing one, is it's confusingly similar with all combinations of two ASCII codes because they are in principle eligible at CCTLDs. I think these are two separate issues. One of them, say what you're referring to, at least from where I sit is in the case of an IDN CCTLD, it is a -- what I would almost refer to as a local issuant and how you deal with, let's say the registration policy. But that's out of scope of the CCNSO. Does that make sense to you now?
Bart Boswinkel:	Okay. That's fine. I'm treading unexplored waters here. It's very difficult. But I think it would be very good if you put it in an email to the group because then at least we've captured this as well and we can say this is within or without of scope of the PDP and who should address this issue?
Mana I Ismail:	Okay. Just a quick follow-up. Maybe I should -- a complicated example or not so relevant example but basically I was trying to check whether you had a possibility for the examples we have provided with confusion within key scripts. This one is familiar because I see the three unique activities now. It's either we're not familiar with the script or we're not familiar with the example. So, is there a fourth script you can confuse within --
Bart Boswinkel:	Within the script?
Manal Ismail:	No, I mean two scripts you're familiar with both but again it's confusingly similar?
Bart Boswinkel:	It's --
Unidentified Participant:	There is no such thing?
Bart Boswinkel:	There could be. I could include it. But this is more -- logically that should be included. Let's put it this way. Then the question is which one becomes relevant or which one should the working group focus on? All of them.
Manal Ismail:	Yes. It's clear if I can see a representative point of view now. I'm not sure I have a concrete example.

Bart Boswinkel:	I can include it because then you'll have at least something, to know that's the way I draft this stuff. At least you have a list of all the different cases and then you can start thinking about -- and I know with (inaudible) on the call instead of this small group, we discuss these cases and now you can check which one is relevant for the confusing similarity and I would suggest, as I said, I would suggest focus in the rules and the overall policy on which side of confusing similarities, in which cases it should limit. Because if it's overall, then it can be as broad as possible. It's hardly any -- it's not discriminating anymore. And that was one of the experiences I think we had from the past. I will include your cases coming up. I have to write it down otherwise I'll forget it.
	I think as well that's a good point. If you send your email saying what you just described within a script, at least what we should do -- again, this is making clear this policy, the overall policy is not about these issues that should be taken care of by whoever will operate the IDN CCTLD. So, it sets the boundaries of this confusing similarity rule. And there are some provisions in the overall policy that deal with it. For instance, that section on the IDN tables. But it strengthens the need for them as well.
	Thank you. Okay. Moving forward, I think this group is too small -- it's good that we start including these other examples and then I'll send it to the list and that can be done very quickly. Now, this is a particular one and it's unfortunate that people with a more technical background are not -- maybe Hiro -- are not in the room. It is, what is the impact of case mapping on criteria if we agree that we move forward with the visual similarity rule as the new one, the question is common forms in small sizes. It's -- yes? If you go just a little bit --
Unidentified Participant:	(inaudible)
Bart Boswinkel:	What we're doing right now, say we have another half hour, is I'm using this document I sent to you and just checking whether all the cases are included and then we have a conference call based on the updated version of this and then we organize a conference call in two or three weeks and run through it as questions -- but now we capture what we need to do.
Bart Boswinkel:	So, my second question is -- what I've included -- if you go back and I hope you don't get seasick -- if you look at the proposed rule it includes the section in common forms in small sizes. Now, one of the at least on clarities of matters or room for interpretation is if we accept this rule which I just showed, then we talk about common forms in small sizes and the question that was posed on the prep call is what does in small sizes mean? Does it refer to say lower-case, upper-case characters or does small sizes refer to the font size? So, if you have a font size of nine then it's very difficult or does it refer to a font size and what is the limit? So, that is a question. What I've included at the suggestion of idealists is a section from ROC5894 which is not an internet standards specification but it's published in the information of the (inaudible) because Galen's suggestion was this is relevant to interpret the small sizes. Maybe we should or should not include the reference to lower and upper-case as well and whether this works. So, that's why I've included it and I've included the text of a particular section in that ROC on case mapping and related issues. Yes? Manal and then (inaudible)?
Manal Ismail:	Hi. This is Manal again. If I recall correctly, small size was referring to the font size because actually the view points dealing with the specific script and they don't even all have upper-case and lower-case and what the group was trying to make is stripping all variables and keeping just one variable which is the confusing -- because some forms are confusingly similar and this implies or existing resolution -- I think they were just trying to fix everything and then just looks how confusing it is being.

Bart Boswinkel:	So, moving forward that means, say, I heard similar voices from ICANN staff involved in the variant issue project is that small sizes does refer to the font size then I would say that I think that results in two questions. The first one is what arsenal sizes, what is the bandwidth? And secondly, should a reference to lower-upper-case for some of the -- and as related in ROC5894, be included, yes or not? But this is a separate question that's been -- then it becomes clearer. Now Hiro?
Hiro Hotta:	Yes. I thought that -- again, my feeling was (inaudible) so I wonder if this concept of lower-case or upper-case in many languages or scripts and maybe there's some concept of upper or lower-cases in other languages is correct -- it's not easy to define lower or upper-case matters and I thought that the small size was the font size.
Bart Boswinkel:	Okay. I'll clarify this and say just for confirmation send it to the group but then the question becomes whether or not say for those scripts which make a distinction between upper and lower case should there be a special provision, yes or no? Yes? Manal?
Manal Ismail:	Just very quick for our founders, the sense of the discussion of the font size was finding a font size that was on the bulletin boards, down the stream, and that's why we did not really get into the details of how big is the font size because we were talking about extreme issues.
Bart Boswinkel:	You could also refer to it to make it a bit more specific on font sizes, for instance, you use in emails and browsers, in the top bar of browsers, because there is a certain limit to it.
Vaggelis Segredakis:	Can I add something? Sorry to interrupt over the phone. Can you see me?
Unidentified Participant:	Sorry? Vaggelis is trying to speak?
Vaggelis Segredakis:	Yes. Can you hear me?
Bart Boswinkel:	Yes, Vaggelis, sorry.
Vaggelis Segredakis:	Hello. What I was trying to add to this discussion is that the reference to the information on RFC that you have added are more general than this discussion because I can understand that small fonts might refer to small characters and lower or small size but the reference that RFC also points out that in certain cases the DNS will not support upper-case characters. So, that's an incremental and very specific part in this discussion.
	I'll include that element as well if you can send -- because what I've done here is I've included the discussion you refer to in your email.
Vaggelis Segredakis:	Yes.
Bart Boswinkel:	So, if you send this as well, I can include that one bit as well in this document. Then we have probably a very good overview of this issue regarding the upper and lower-case and font size.
Vaggelis Segredakis:	I had some trouble following all the discussion because there is some bus thing over the voices and it's difficult for me. So, I'll check on the transcript exactly what was discussed and I'll send you whatever I can.
Bart Boswinkel:	Okay. Thank you. Okay. So, that was -- we got to font sizes and upper and lower-case where it is relevant. So, I've included case-mapping, et cetera. And

again, another -- if you go back to the visual similarity rule as included in the VIP report, there is a reference to common fonts. Now, the question is can this concept be -- what are common fonts? Can that be determined? I'm not sure what -- Windings is probably not one of those common fonts although everyone uses it as an example. Can this be more -- should this -- let me put a phrase in, rephrase it. Should this be more qualified, more specific? What are common fonts? Because what is a common font for one person is not a common font for another person and is there, say -- I don't know whether across scripts there are differences in common fonts as well? So, if you just say common fonts you have no criteria to determine what is a common font. Then you'll have introduced a concept which is very fluffy and you end up with the same issues as we have. So, that's more or less the purpose of this question. Should it remain or should it be rephrased or any other ideas about this font?

Vaggelis Segredakis: I think it would be nice if at least some common fonts were proposed because there are so many of them.

Bart Boswinkel: Yes. But this is literally from the definition from the project.

Vaggelis Segredakis::   I know.

Bart Boswinkel: What I could do is go back -- I could go back to ICANN staff and ask (inaudible) Yes. So, what we discussed and what does it refer to?

Manal Ismail: I think it's going to be easy to give some examples of what are common fonts but would it be possible that we have an exhaustive list or a concrete -- this is the issue. Again, if we give examples, then it is still open and flexible. I mean, the decision is are we going to have a concrete, exhaustive list on the common fonts?

Unidentified Participant: No.

Manal Ismail: It could be just some guiding examples?

Bart Boswinkel: Yes. We always have agitators, unfortunately. But at least it is something that you don't have to look at the very extreme fonts. I think if that is excluded, you could have a discussion about it, whether this is a common font. If that's the issues, it's only -- that's what we noticed in the examples we were discussing in the working group. It is sometimes the very edge cases regarding the fonts. So, you end up with some fonts where they were confusingly similar which are not regularly used. So, look -- I'm not absolutely an expert. There were one or two fonts I've never heard of. So, I've included this question and I think this is more -- I'll show you what I'm referring to. This is from the implementation plan and fast record implementation plan.

This is a more description -- say, a test basis of what you should look for. It should not be confusingly similar in the case of IDN CCTLDs with say two basic ISO -- that's the ISO 316 list existing in order to ASCII characters combinations which are in order to protect everything under the CCTLDs for CCTLDs and broader because the ISO list is used for a lot of purposes besides the DNS. So, just to avoid future confusions. Then these are the criteria that were in the implementation plan. So, I've copied this and secondly what I've included and just explaining this and then I'd go back to my question. Currently the DNS stability planner who does the review uses this as a test to determine whether a string is confusingly similar. And I think one of the issues that has arisen is that this test was only made clear in a blog posting. It's not included in the implementation plan. It has been developed after, say, in parallel to the implementation plan. But it is in fact a test to determine whether a string is confusingly similar according to the rule.

Now I go back to my question here. Should a test like this, because it's almost very deep into implementation, yes or no, but should a test like this be included in the policy itself -- yes or no? And if we think it should be included, maybe at a higher level, is this test as we use it right now, is that the correct one? Or should, say developing such a test be part of the implementation of the IDN CCTLD overall policy? And on the call we had leading into this meeting, there were some suggested that some implementation aspects should in fact be included in either the processes bit or in the criteria if possible in order to avoid the experience of the new CCTLD process. So, that's a question again. We don't have to answer it but my question in fact for you is, is this a relevant question? Yes or no?

Giovanni Seppia: Hi. Thanks for this. I've been -- the (inaudible) .EU has been quite deeply involved in this compatibility matter and we understand that there is at least some criteria for determining if two characters are visually similar with other two characters. However, what we have been objecting to since two years, I'm talking about some technical points of view, is that these visual similarities should be based on scientific grounds. Currently it's based on something that has no scientific ground in the sense that when we started to look into these, we had these covered work, the work of neuroscientists and there are I believe more than 1,000 researches developed authentically clear to prove that characters are visually similar to other characters when people look at them.

So, these are scientific-based versus the basis of certain opinions. I do believe this kind of scientific approach has been included in this process and what can we visually -- I can give you an example. There was an article which was published in one of the most famous medical journals in the States which says what is visually similar for an American person with an American culture background is not visually similar for somebody who has been educated in Asia or somewhere else in the world. That's my $0.02 on this that whatever kind of test or research criteria, there should be a scientific background. I believe it is time for ICANN to move to this level, to have something really scientific. Because if you're just very -- okay, this is up to the people on the deciding panel, there are no neuroscientists in this panel. It's a bit difficult to motivate that two characters are visually similar to others on the basis of a lead that has no scientific reference.

Bart Boswinkel: Could you be so kind as to suggest this to the list maybe with some literature or overall criteria because then if I understand you correctly, the way I've done it -- and I will adjust this -- there should be in your view a reference to the test itself in the overall policy? But whether this is the correct test, that's not a question? So, that means we need to discuss the level or give reference to testing and then the question is to what level do we need to go? We don't need to read all the scientific articles. That's a matter of implementation.

Giovanni Seppia: That's correct. I can give you another example which is probably my last, but basically what's happened is I had this meeting with this famous neuroscientist, a French lady. She put me in front of some Arabic characters. To me they were absolutely not visually identical. She smiled and said 100% of Arabic people, they are visually identical. But for me, they are not. That made me think a lot. Because of course we have a different interpretation of what you have around you.

Bart Boswinkel: I think -- I'll rephrase this in another way. It should be included. The level is one. And there should be a reference to the type of test? And then it's a matter of implementation to really work on it and set it up. Preferably, as you said, scientifically or at least with a high level of scientific evidence behind it. Yes?

Manal Ismail:	Just seeking clarification. We are talking here about things that are confusingly similar, meaning that you can't differentiate between one and another and not used interchangeably by the community, right? This is -- yes. Exactly. I'm just -- thank you.
Hiro Hotta:	I guess I have a related question. Maybe this question is relevant the similarity definition at the top of the paper Bart is presenting. My question is what is the case we define similarity, for example cross gate. Users inspect between a string appearance, a close, careful look or the second case, the users see the string appearance as a very careless plan. For example am I right to say strings are similar at a glance but looks different when we inspect them carefully, they are not defined as visually similar.
Bart Boswinkel:	I'll tell you, it's -- I think it's one of the best things to do, to check initially with the working group. And again it determines the level of testing and it relates also probably in my view -- that's just me speaking -- to what Giovanni just suggested. What is the level, how does perception work of similarities. Yes?
Manal Ismail:	(inaudible) Thinking over again what Giovanni has just suggested, I think it's very important that we decide which category of the ones you proposed this morning are our target because I mean the perception of someone who knows this script is different from someone who doesn't. If you don't know the script you're going to be very cautious and more thorough. But if you know, if you try to recall the knowledge you already have you might make some compromises.
Bart Boswinkel:	That's why I've included it because at the end of the day it came up in another discussion. It clarifies the discussion in my view of the similarity. So, for whom are you protecting, from what? That's the question. Maybe I'm too cynical. We can't protect the whole world from everything.

These were the four questions I have included and thought of. Is there in your view something else that needs to be included? Maybe that's a way forward because you see the document now for the first time and it's a bit of a brainstorming session in that sense. I'll update this one based on the comments from today and from (inaudible). I'll send it around and before we move forward I'll ask the working group whether there should be other elements included, yes or no? And then we can have -- then you'll have an overview of where it leads to. Then we have, I think, a very sound basis to finally really go into depth of the visual similarity based on the experience and say new language. Yes? Then I thank you very much for your attendance. Thank you. |

Unidentified Participant: Thank you.

Unidentified Participant: Bye-bye.

Unidentified Participant: Bye.