ABU DHABI – IDN Latin Generation Panel Meeting
Tuesday, October 31, 2017 – 15:15 to 16:45 GST
ICANN60 | Abu Dhabi, United Arab Emirates

MATS DUFBERG: …are characters in the languages that are not part of MSR-2. And we both look for code points and code point sequences. And code point sequences are used when code point is a mark, which does not stand on its own. It depends on the code point before. So they create a combination.

So we collect code points when there are normal letters and we collect code point sequences when they consist of a normal letter plus one or more marks.

Right. I said already that we consider five, and that is in the case where it's on the border so that it's widespread use even though it's number five.

Code points may only be included if they have established use in one or more languages. And when we're registered in our database, or to say, spreadsheet, we note more than one language to be able to have a safety net if something has been incorrect. Next slide, please.

If a code point is a mark, then we don't include it as-is. We only include it in the context. That is, we look for the letter before in

---

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.

EN

that language and register the sequence of the letter plus the mark. So for the station of a mark that is as part of a sequence. So for every sequence, there must be some language using that sequence.

In some cases, a character might be constructed by multiple code point sequences. The order between the marks could create two different – I mean, if there are two marks, the order between those can create two different sequences. And we say that to include both of those, we must justify the languages. Of course, if two different sequences are with the same result, the same glyph are included, that must, of course, go to the variant group, too. So they must be treated as variants.

This is the fact selection and we might reduce later on if there are multiple sequences for the same character.

If there is a precomposed alternative, then that should be used instead of a sequence unless there is some evidence from a language that the non-precomposed is what is used for that language. And again, variant management must be [come in], and secondly, if the limitations within Unicode of the use and [IDNA] of the use of the sequence, then of course, we respect that.

When it comes to modifying letters, they are kind of something a little bit in-between. So we first, primarily, we will look for

sequences where a modified letter is part of a sequence and include that. But it might be that that some modified letters will be suggested to be included on its own.

MIRJANA TASIĆ: Mats, [Marcus] wants to [inaudible].

MATS DUFBERG: Yes, okay.

[MARCUS]:   This one, part of what you're talking about is related to NFC, which is applied, so in some ways, you don't have to care about anything before NFC. It's just what the output NFC gives, which is the thing you need to start working on for the repertoire because anything before applying NFC is essentially irrelevant.

UNIDENTIFIED MALE: [Heading] to that, that NFC would, in fact, dictate which sequence is authorized and which one is not authorized, so in fact, for a sequence of combining [marks], only one is valid. So the other one cannot even exist in NFC.

UNIDENTIFIED MALE: [Inaudible]

UNIDENTIFIED MALE: Only one sequence can exist under NFC. The other one is not valid, so you cannot ever justify having two different sequences because only one of them is valid in IDNA.

MATS DUFBERG: If it's already defined in Unicode, there's no problem.

UNIDENTIFIED MALE: It's not Unicode. It's IDNA. I mean, NFC. It's NFC.

MATS DUFBERG: NFC is defined in Unicode.

UNIDENTIFIED MALE: Yeah, basically NFC is the low end IDNA 2008, so there is no choices. Only one of them is valid.

MATS DUFBERG: I don't know. There might be cases which are not defined in Unicode as with the NFC [inaudible].

UNIDENTIFIED MALE: Well, by design, it is not. So if you find a mistake, you're welcome to find, but by design, NFC does not allow this kind of case, at least for Latin.

MATS DUFBERG: The composed characters added later.

UNIDENTIFIED MALE: That's the point. You cannot any precomposed anymore because of the NFC stability. That's a big deal between ITF on Unicode, I tell you that.

MATS DUFBERG: Yeah, but there was something added in Unicode 7.

UNIDENTIFIED MALE: It's a different matter. It's completely different. It has to do with Arabic and it's a completely different subject.

MATS DUFBERG: Okay.

UNIDENTIFIED MALE: But anyway, at least for Latin, it's not a concern.

MATS DUFBERG: Right. So I just want to make it clear that if anything is not NFC, we will not include it.

UNIDENTIFIED MALE: Okay, that would be a better statement than what you had on this other slide.

MATS DUFBERG: Pardon me?

UNIDENTIFIED MALE: That would be a better statement than what you had where you said, you would allow two combining sequences to be allowed from [inaudible].

MATS DUFBERG: We should clarify that.

UNIDENTIFIED MALE: Okay, thank you.

MATS DUFBERG: So there is no question that only NFC is permitted by IDNA and so anything which is NFD is ruled out. Yes, please.

UNIDENTIFIED MALE: Yeah, I agree. That's a better way to present it because if you start to consider the alternative available, then you have to consider in which order you consider them. So you may get into conflict because depending on the order you consider the sequence, then you may choose one alternative versus the other and people may prefer the other. So it's better to say that you use NFC.

MATS DUFBERG: If there is no question, if it's NFD, we cannot include it, period. We have not assumed that there is nothing in Unicode that can end up being alternatives on the order, but if there is no such thing, then there is no problem.

Yes, Mark?

[MARK]:   Maybe a final comment, but related to the conversations we had, which is, I guess I think you have enough work that you should start with PVALID and only NFC.

MATS DUFBERG: We are well aware of what is PVALID and what is not PVALID and we are well aware of what is MSR-2 and not. We have tools to verify if things are NFC or not NFC or if there is no such

relationship or not. So we will check that, so I can assure you that things that are not PVALID and are not NFC will not come to you. So you don't have to worry about that.

But when we created the principles, we had an open wording. Yes, so I think you have talked about this. Right? So we can go to the next slide.

Principles for exclusion, if a code point is allowed or unassigned, I mean, it's not PVALID, then it cannot be included. And if a code point presents a security stability issue, then it cannot be included. If the code point is either deprecated or not recommended by Unicode standards, then it should not be included unless there is no other alternative for that language. Next slide.

If a code point is no used for a contemporary, ordinary text, it should not be included and code point that predominantly is used as a formatting character mark, numeric digit, punctual mark, etc., then it should not be included either. Next slide, please.

So the resources we have are primarily the 180 languages and we have resources primarily on Internet that we look for, say, the alphabet of every language and written text by all languages to see what characters they use.

And the steps that we go about is that we, that every member, picks a language and finds resources and notes down for all characters that are to be attested, the information that we have found. At least, if there's a new character, not to test it by many already. I mean, there's no meaning of a testing characters that have been attested by several languages already. And A to Z, there is no reason to attest because that is, by default, included. Next step. Next, yes.

While we consult other sources and yes, a lot of text here. Yeah. And we also write down who has done the work so we can go back and talk to each other about any problems. Next slide, please.

This is very hard to read, but this is a list of the languages on the left and some information, and you see the EGIDS ranking. And we start with one there, and we have sorted by rank and alphabetic order. So next slide, please.

And here, we have an example of a testing. So the very left column is a note if it's in MSR. We keep track of that. So a cross means that it is in MSR and then we have the Unicode code, we have the glyph and we have the Unicode name. If it's a sequence, then it consists of two or more Unicode codes and the Unicode name is the combination of the parts. Then we note the

languages that use it and a link to a resource of that language and initial of who has discovered that.

So a question is if we should proceed with EGIDS 5 and what selection criteria we should have. We haven't decided that yet.

Sequences, well, we keep them as sequences. I don't know if it's so much, there's two examples here of sequences.

Many languages have digraphs or trigraphs as letters, but they are treated in their alphabet and their writing as one unit but they are composed by what we think are different letters. I don't think we should consider them as anything else but the letters they consist of. Click consonants are a little bit tricky because those letters are not always permitted by MSR.

Here is an example of something which is not found in MSR, 0264 Latin Small Letter Rams Horn, also apostrophe and those things we have to look at and come back to the integration panel.

UNIDENTIFIED MALE: Apostrophe is not PVALID.

MATS DUFBERG: At least, not in that form, no. But I think there is another form which is PVALID, same that is used in Ukraine.

UNIDENTIFIED MALE: [Inaudible]

MATS DUFBERG: Pardon me?

UNIDENTIFIED MALE: 2BC.

MATS DUFBERG: 2BC, yes. So apostrophe as such would not be. I mean, this is the raw material so I would assure you that anything not PVALID or context or [context J], but I guess we don't find, will not go to IP [inaudible].

UNIDENTIFIED MALE: Even 2BC, had in fact explicit ruling by IIB and not being allowed anywhere, so it's not even for us a done deal that you would never get anything that looks like an apostrophe in any shape or form in the root. That was, in fact, an IB decision, not even an IP decision. It's the Architecture Board decided against it.

MATS DUFBERG: Do you know where that is stated? Do you have that?

UNIDENTIFIED MALE: I can find it, yes.

MATS DUFBERG: So, [Mark], you have it.

UNIDENTIFIED MALE: I think there is a reference to that in RFC 6912 where we can go and check again.

MATS DUFBERG: You can check it out just for curiosity. So it means that these will be removed, of course.

MARK: But really, 0027 Apostrophe, there is no discussion about. Less of time.

MATS DUFBERG: No. So I can assure you that anything, I've said it many times that anything not PVALID would not be sent to IP. So this is raw material. This has not been analyzed, so no problem. So we know how to find out if it's PVALID or not. So we have no problem with that.

EN

UNIDENTIFIED MALE: So there is some discussion on 02BC in RFC 6912.

MATS DUFBERG: Okay, thank you. Yes, so there is what to do when a symbol or a character used in the language but not appeared to have Unicode Code Point. Well, we cannot create it. So which one? Do you know, Mirjana?

MIRJANA TASIĆ: Bill found it.

BILL JOURIS: [inaudible].

MIRJANA TASIĆ: In [Tuban] language.

BILL JOURIS: Between the one after the [G].

MATS DUFBERG: Okay. Have you drawn that?

BILL JOURIS: That was taken from Unicode [inaudible] or from [inaudible].

ICANN 60 ANNUAL GENERAL
ABU DHABI
28 October–3 November 2017

MATS DUFBERG: But if it's code here, written, then it must be Unicode. Yes, but I don't think you can write it if it's not the picture. It might be a picture.

BILL JOURIS: [Inaudible]. I don't do this slide.

MIRJANA TASIĆ: Raw material.

BILL JOURIS: [Again], it's not PVALID.

MATS DUFBERG: No, it's not PVALID. So it can't be included. Next slide, please.

So it's over to you, Bill. Yes, anymore questions on this material? Please realize that we are just in the end of the fact-finding, the digging, and we have put up as much as possible. We have not processed material. We will filter it through what is allowed by IDNA and also if it's ruled out by IAB, then there is no possibility, and then, okay, MSR-2, is there anything that don't go through MSR-2, then we have to look at it and consider requesting IP for an extension of MSR with those code points.

BILL JOURIS: Okay, the Variant Analysis Group has started out looking at variants for the various related scripts. We specifically started with Cyrillic. We are about halfway through verifying the variants that we have found so far. We have the great advantage over the Cyrillic panel that we have a very small set of characters on their side to look at, so when we find something, if it isn't in their character set, it's easy. If they found something, they would have had the 180 characters to try and dig through to find it. So we've made fairly good progress on that. We have not yet started on Greek or Armenian. We are open to looking at other scripts if we come across one that does seem to have a significant number of variants, but to date, nobody has found an additional member.

Once we have done that and the Repertoire Group has wrapped up their work, then we can start to analyze variants within the Latin script. We are, as previously noted, restricting ourselves to MSR-2, although if something is found that wasn't in MSR-2, we would at least look and see if there are variants of those as a precaution against possible changes.

What we mean here by "sufficiently universal" is there are cases, for instance, an O with a diaeresis, which can alternatively be written as an O followed by an E. But that happens in some languages and not in others, and our inclination is if it only

EN

happens in one language and not in others, we're probably not going to consider that as an alternative. Next slide, please.

This is an example of how we are going through the analysis. The visually identical items are relatively easy. For example, the Latin small letter E (e) and the Cyrillic small letter IE (e) are identical. The Latin small letter E (e) is not identical with a Latin small letter E with a diaeresis (ë). We may question the wisdom of that choice, but that is the instruction that we've been given.

When we come to orthographic considerations, the German small letter S, if you have an S set, there are some cases where that can be replaced by a double S. There are other cases where a double S cannot be an S set, and that has issues with the mandate in the RFC that variants be transitive, the word that they go both ways.

And finally, we have the issue that Mats raised earlier about if we have combining code points, and so there are multiple sequences that will get you to the same place. For instance, if there are two diacritic marks on a letter, clearly, you could do those in either order and if there exists a Unicode combining mark already for each of them, then you could have either of those combined marks plus the second combining diacritic mark, etc.

UNIDENTIFIED MALE: Not possibly NFC, again. So I don't, that case does not exist in NFC, at least for Latin. There are some more complicated cases in Indian languages, but there is, in fact, a mechanism to avoid that. Each [inaudible] has a weight, a value, on when you do [inaudible], they get stacking according to some value. So NFC guarantees that that does not happen.

BILL JOURIS: The people in the Variants Working Group were not sufficiently expert to be aware of that particular piece of information. Next slide, please.

As I said, we have gone through approximately half of the Latin code point repertoire compared to Cyrillic language. We have found some 17 variants. About half of them were variants that the Cyrillic group had already identified. We have communicated with them about the additional ones we've found, and Mats, I believe you said that they had confirmed.

UNIDENTIFIED MALE: We were checking – me and [Dema] – about those characters. It's possible that we made some mistakes during the process so I will wait until the end of your work and probably the end of the public comment period, and then we will combine everything into a final document.

BILL JOURIS: As I say, we have it much easier because we only have to compare against eh much smaller repertoire that you have established and our repertoire is still a work in progress, so you were working under a handicap. Okay, next slide, please.

So our next steps, we need to finish processing the Russian language and we have to then do the Greek and Armenian and once the Latin repertoire is completed, then we can look at variants within the repertoire here. Next slide, please.

We do have a number of issues and questions that we will need some feedback on. Next slide, please.

Early this summer when we first started looking at the variants with Cyrillic, we found these two glyphs and actually discussed them at our face-to-face meeting in the early summer and went – yes, those are indeed sufficiently identical that they would be variants.

A week and a half ago when we were looking at the – we're doing a double check and we got the glyph shown further down. You will notice that Unicode 00E7 has changed. And so we would appreciate some intelligence if anyone knows of what other changes in the Unicode cable glyphs might be coming along so that we can take account of that. Next slide, please.

UNIDENTIFIED MALE: I don't know what [inaudible].

UNIDENTIFIED MALE: What different sources of [inaudible]?

BILL JOURIS: Those were all cut and paste out of the Unicode table.

UNIDENTIFIED MALE: But no [inaudible]. I am in fact [inaudible] producer for Unicode, so I do produce Unicode [inaudible], as well as the ISO [10646] and I can guarantee you that did not change. I mean, you may have [font] that changed but the representative glyph has basically defined by Unicode Consortium on the ISO standard did not change. That's not correct. In my opinion, I won't [inaudible] but I'm producing the thing so I should know I guess.

UNIDENTIFIED MALE: It's [inaudible].

UNIDENTIFIED MALE: It's a font issue. It is probably you change your printing systems or whatever something happen or font and then it's –

BILL JOURIS: Possible but it was, as I say, it was strictly cut and paste from the Unicode table.

UNIDENTIFIED MALE: What do you mean [inaudible]?

UNIDENTIFIED MALE: Because when you do cut and paste, you actually – you're probably – the destination file or Word processor or whatever is actually choosing a font for you and then that font probably has changed or you had a different font. Anyway, it's a font problem.

BILL JOURIS: So what you're saying is even though domain names are independent of fonts, the only fonts we can consider is the one and only font used in the Unicode table. Is that what you're telling us?

UNIDENTIFIED MALE: Come on, I mean, that's –

UNIDENTIFIED MALE: That's [inaudible].

UNIDENTIFIED MALE: [inaudible] of a long story on the complicated story on – that's due in fact to previous unification between Romanian and Turkish, and it would took forever to – at some point, they wanted to differentiate those on some of those characters where this unified between a Romanian usage and Turkish usage. I mean, we could spend two hours going for that.

And so that explain why some fonts turn to be ambiguous and they're wondering that they can stay in between the comma, below, on the [inaudible] below but that's nothing to do with the definition. Whatever font do, it has nothing to do, [I'll never say] you have a risk. After all, in the platform if you control the font that it's showing the address bar, you are controlling everything because you can change completely any character to show whatever you want based on the font that you have installed in your machine to display the domain name.

So you have to rely on some degrees that the font are doing the right thing but I can assure you that there is no difference between early September or mid-October but the [position] of E7 is still the same. On E7, it is defined with [inaudible], not a [inaudible].

BILL JOURIS: Okay. Next slide, please.

Okay. As everybody in ICANN is very aware, domain names are only done in lowercase script. And so if you'll look at those, if you'll go to a user that is not in ICANN, that is not in IETF, he has decades of experience that tell him uppercase or lowercase are totally interchangeable.

The fact is that the uppercase all gets translated to lowercase before it gets to the DNS lookup but he doesn't know and he doesn't care. So he would look at those and say, "Yup, they're all three .com just like I'm used to." The first one is lowercase, the second one is uppercase, the third one is also uppercase, just a slightly smaller font. It's all the same.

And except of course the last one is not uppercase, it's lowercase Cyrillic. And so, the question arises, are the – [called] third character in the first and third items variants because they are going to be considered identical by all of our users or are they not because they are visually different even though nobody will notice the difference.

MARK: You're completely right that this is a big problem. As been discussed at [inaudible] for years, that IDNA doesn't solve the

EN

problem at the higher level thing but again, that's somewhat out of scope of our work, which is a different issue [inaudible].
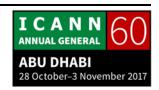
UNIDENTIFIED MALE: However, if we say those two –the Cyrillic M and the Latin M – are variants for our purposes, then we don't have the problem of saying, "Okay. You can do a Cyrillic .com as a TLD because it's not a variant." No. There's a difference. They're not variants. So we're all good.

And within minutes after that new TLD comes live, the domain names of every financial institution on the planet have been registered with the Cyrillic .com and basically we've enabled phishing on a massive scale. So if we don't say they're variants, then we're asking for trouble. Yeah, the sensible approach would seem to be to say yes, they are variants for these purposes but we need some clarification from the IPs to whether we're allowed to do the right thing on this. And we don't need any answer today but we do need any answer at some point.

UNIDENTIFIED MALE: I would just like to add to this discussion that in the TLD evaluation process, there are actually two different steps. So, after the DNS stability interview, there is also a similarity review check, which is outside, I guess analysis as they're looking at

variants itself. So that is also another place. This actually could be looked at as well.

MATS DUFBERG: Just a comment that the work that we do here will probably also have influence on next level, what they consider at being variants. So I think choosing not to consider this case has an effect on the security and stability of Internet.

BILL JOURIS: Okay. Next slide, please.

Okay. This is another example. And the second one is obviously a font with serifs and the other two are sans serif and even people who don't know what a serif is can look at them and see the difference. The question in our mind becomes will someone who isn't seeing this next to each other when they see the first one, realize that it isn't the same as the second one. In fact, of course, the first one is Cyrillic but the difference between the first characters is visible if you see them side-by-side and totally unnoticeable if you don't see them side-by-side. So that is something of a question as well. Next slide, please.

And this is think is what all of our questions come down to is when we have distinguished between variants and things that are really confusable, what is the real-world problem we're

trying to solve here by making that distinction? It's pretty clear that it would be important if you're writing an academic paper in linguistics. It's not clear that if you're doing something that is trying to adhere to the least astonishment principle when it comes to the Internet using community that there's any real value here. Next slide.

Back to you.

UNIDENTIFIED MALE: [Inaudible], I mean, in Latin, it's a vast subject and we can go all of the place on this. I tend to think that we need the first line of defense on this mechanism, that is what we see and there's also a level at the ICANN level where they can block basically two different root levels that will be too similar will create confusion. So we are not going to do solve everything at all levels, so it's just the first line.

On the way that in fact even in ASCII you have the same issue really. Within ASCII, you can create in fact levels that are very confusable like between the [au] compared to the M or the L, versus the 1. And there's all sort of things you can do to trick users even in ASCII, the plus digits.

So, it's not going to be perfect and so we have to agree on what we want to do. I agree on that. We tend to be – what I would say

EN

for example is an absolute need when you have a variant situation where a typical font using, you know, you can take your favorite vendor and say use exactly the same character, the same glyph, there's not even a difference. You can even go sometime and the font, if you have a font inspection tool, you will see in fact that two Unicode code point refer in fact to the same glyph ID. Well, that's very common like between Latin on Cyrillic. So use in fact the same glyph, it's not even repeated in the font. It is basically mapping to the same. So that's how we call [homoglyph] sometimes.

When it's really exactly the same, then there is no doubt that when [inaudible] have to be put in the variant relationship. When you go beyond that, there's some judgment call on – the only thing I can say we're going to help you at some point to finalize that and to say, "Well, you're going too far. You're not going far enough." I don't think there's a perfect answer to that. There's some judgment to be done and the Latin is probably one of the worse case so you will have to kind of make your own way.

BILL JOURIS: Perhaps what we need is to get some perceptual psychologist in here to give us a little expert opinion on what people can tell the difference between… people who are not linguists, people who are not Unicode experts, what will they notice and what will they

ICANN 60 ANNUAL GENERAL
ABU DHABI
28 October–3 November 2017

not notice because that's what the least astonishment principle is all about, right? It's making sure that people will notice that there is a difference and not be confused.

MARK: [Inaudible]. I would suggest that you restrict your work on PVALID and MSR and NFC that would list the starting point. Because, for example, the Greek [clutters] that you were talking about are not PVALID, so it's like it doesn't exist a domain name.

Even PVALID and NFC as we've been discussing will actually help you a lot for your work because you will restrict the actual list of code points you will be studying. So, I really suggest that you start with that smaller sandbox.

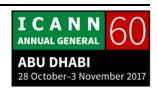MIRJANA TASIĆ: Okay. Any other questions, suggestions, comments, thoughts?

So, thank you very much for attend –

UNIDENTIFIED MALE: You know what, there is one.

MIRJANA TASIĆ: Oh, [Sarmad], sorry, I didn't –

UNIDENTIFIED MALE: I'm not [Sarmad].

MIRJANA TASIĆ: Not [Sarmad]. Sorry, I didn't see you. Sorry. Please proceed.

[JOVEN]:   I'm [Joven] from a [Mara] nation in Bolivia in South America. And so my language is based on Latin [reading]. So we have some like two dots above the vowel and we also use a lot upper [inaudible]. And so I was wondering, this is my first time in ICANN and so also I'm not connected to technology stuffs in my country. I belong to the social scientist branch. I want to know how we can get on domain names or – especially on domain names those kind of writing, reading, like with two dots, you understand, right? Two dots above the vowel and also upper [inaudible]. I didn't see included in the Spanish.

MATS DUFBERG: So the two dots over diaeresis of vowels is quite common in Latin script. So it's used by several languages like German and Swedish, my mother tongue. And so, that is quite uncomplicated. But the user depends on – I mean, if you are to register a domain name, the TLD that you want to register domain name under must support that, so you have to ask the TLD.

The apostrophe is more complicated because it's probably less accepted to be used. And it might be accepted under a TLD, so it depends on the TLD. So what TLD are you thinking of? The scope for this discussion we've had today is, I'm sorry, is the creation of new TLDs and it's more restricted for new TLDs.

For example, digits are not allowed at all in TLDs. But if you register a domain name under a TLD, normally, you can include digits in the domain name. We'll have a digit-only domain name.

So what TLD do you want to use?

[JOVEN]:   Actually, it's a [inaudible].

MATS DUFBERG: Yeah, but what TLD?

[JOVEN]:   [Inaudible].

MATS DUFBERG: Yeah, but top-level domain.

[JOVEN]:   It's like a comma.

MATS DUFBERG: Right.

[JOVEN]:   [Inaudible], I mean, you're going to [inaudible].

MATS DUFBERG: Turn a microphone on again. So what top-level domain do you want to register or do you want to create the top-level domain?

[JOVEN]:   Yeah. Actually, it's like comma, the apostrophe, like comma first of all. And so, I'm not sure but that could be in both level of domains I think.

MATS DUFBERG: Yeah. The top-level domain of your country, then you should ask them to include the characters that you need for your language. So what country do you come from?

[JOVEN]:   Bolivia.

MATS DUFBERG: Bolivia, yeah. So, I don't know then the TLD of Bolivia but that [TLD], that is quite free of creating or adding code points as long as they are according to the standards. It's nothing that goes through ICANN. They decide themselves because it's as a ccTLD. But the comma here, the apostrophe is shared by Ukraine. They also use that.

UNIDENTIFIED MALE: Just that in Belarusian, Belarus language they are using on second level, so we are – let me explain. We are not talking about NBA.com, so you can ask about N apostrophe BA.com. We are talking about .com. This is top-level domain name. I suppose what you were asking is how can you get apostrophe in domain names like N apostrophe BA.com? Am I correct?

This is the [inaudible] of your local guys from your registry and they need to submit a table according to certain rules, they know better which rules, etc., etc. and you will get apostrophe inside.

MATS DUFBERG: Actually, the ccTLDs, they decide by themselves so they don't have to go through ICANN.

UNIDENTIFIED MALE: To IANA, they need to submit the table. So there is a process of –

MIRJANA TASIĆ: Yeah. [Inaudible] giving some [inaudible].

UNIDENTIFIED MALE: Yeah. And that there is a – believe me.

MATS DUFBERG: They should probably but they [don't must].

UNIDENTIFIED MALE: No, they must.

MIRJANA TASIĆ: No, they must.

UNIDENTIFIED MALE: They must. They must, believe me.

UNIDENTIFIED MALE: [Nobody cares].

UNIDENTIFIED MALE: At least.

UNIDENTIFIED MALE: No, no, no, no, no.

MATS DUFBERG: I doubt that my own ccTLD .se accepts that they must submit. So I don't think that that [inaudible] but that's a political issue. And so –

UNIDENTIFIED MALE: No, it's not a political issue. This is technical issue and as me, Mirjana and [Dema], we are coming originally from ccTLDs. We all needed to send a table.

MATS DUFBERG: What happens otherwise?

UNIDENTIFIED MALE: To IANA.

MATS DUFBERG: What happens?

UNIDENTIFIED MALE: You don't have a ccTLD.

MATS DUFBERG: Pardon me.

UNIDENTIFIED MALE: If you don't send to IANA a proper table, then you don't have ccTLD.

MIRJANA TASIĆ: IDN ccTLD. He's talking about IDN ccTLDs.

UNIDENTIFIED MALE: This is IDN. This is not –

MATS DUFBERG: I know. But –

UNIDENTIFIED MALE: And you have IDN Fast Track Process and you'll –

MATS DUFBERG: Okay. Okay. We're talking about different things.

MIRJANA TASIĆ: Different things. Yes.

MATS DUFBERG: Under the ccTLD, you decide by yourself. I'm not talking about applying for a ccTLD. If under the Bolivian ccTLD you can register, you can decide by yourself.

UNIDENTIFIED FEMALE: [Dušan], please, [Sarmad] wants to say something. Maybe he will [inaudible].

[SARMAD HUSSAIN]: This is [Sarmad]. So, yes, it is correct that if you apply an IDN ccTLD to the Fast Track Process, the Fast Track Process as for submitting an IDN table but it is also correct that that information is used within the Fast Track application process. There's a separate process to submit that table for publication at IANA and the ccTLDs are requested to do it and encouraged to do it but I think that's the way the process stops. But if you're interested, I can also check back with IANA and see how. But I can check with IANA and see whether that is a requirement or not.

BETTINA GERTUM BECKER: Our online comments.

MIRJANA TASIĆ: So we can have one comment in the chat room. Please, Bettina.
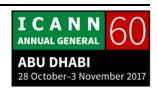
BETTINA GERTUM BECKER: Okay. It's a comment from Dennis Tan. The Bolivian registry domain names were using A to Z and 0 to 9 and hyphens, and he put the reference, you can check in the chat room. And so they would need to implement IDNA to allow for N and [the sign above], accent and diaeresis.

UNIDENTIFIED MALE: I think [he] should differentiate between the ASCII domain names and the IDN domain names. It's not possible to include the special characters in the ASCII domain name. And the ccTLD is a normal one. But in the IDN maybe, it's possible to include them.

MATS DUFBERG: I mean, IDN works on a label level. So, under an ASCII ccTLD as .se, so on the second level, you can create whatever you want as long as it's IDNA valid. So under .se, we have some Latin cold points but we also have Hebrew script, part of Hebrew script permitted. They're not mixed because [that's not] but either or.

MIRJANA TASIĆ: We are almost at the end of our session. We have a few more minutes. I don't know – I am not in the possibility to wrap up this session and make some conclusions. But I think that – are we exposed here and everything what was said, it will be somehow – is it recorded and we can listen to it afterwards and take whatever we need from this, so from this recording.

Is there any other questions or comments or addings to what we have said up to now?

Now, I want to thank you all very much for attending this meeting and that's all for today I think. Thank you.

[END OF TRANSCRIPTION]