
MARRAKECH – Latin Generation Panel Meeting
Sunday, March 06, 2016 – 17:00 to 18:00 WET
ICANN55 | Marrakech, Morocco

CHRIS DILLON: Is he going to go clockwise or counterclockwise? I think it is a clockwise day. I'll start by introducing myself. I am Chris Dillon, and I am Co-Chair of the Latin Generation Panel.

UNIDENTIFIED FEMALE: [inaudible] ICANN staff.

MEIKAL MUMIN: Hi. I'm Meikal Mumin. I'm not ICANN staff. I'm a community member. I represent academics mostly.

ANDREW SULLIVAN: I'm Andrew Sullivan, and I'm just watching.

GUSTAVO LOZANO IBARRA: Gustavo Lozano, ICANN staff.

MARC BLANCHET: Marc Blanchet, Viagenie and Integration Panel.

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.

AHMED BAKHT MASOOD: Ahmed Bakht from Pakistan and representing Arabic GP and a member of [inaudible].

HAZEM HEZZAH: Hazem Hezzah. I'm from Egypt, from The League of Arab States, and I'm a member of the Arabic Script Taskforce.

UNIDENTIFIED FEMALE: I'm [inaudible] from the Vietnam Internet Network Information Center and [membership] of the Latin community [inaudible]. Thanks.

[INAUDIBLE] DONG: [Inaudible] Dong. I'm from the Vietnam Internet Network Information Center.

CHRIS DILLON: Thank you very much. I shall just check in the chat room to see if there's anybody else attending remotely. Alireza Saleh, who is ICANN staff, and Gustavo Lozano are in the chat room.

Would you load up the presentation on the desktop? Thank you very much.

The idea of this is just to give a brief idea, especially for people who are new to the panels work, of where we are with this at the moment. It's effectively a status report. Oh, Lovely. Yes, thank you.

The current themes we have are the potential scope of the panel, so what's in scope, what isn't. Things on the edge of scope may be things like Romanization. The Latin script is used for many languages, but also used for the Romanization of languages, so what about the status of Romanization's?

Then I'll spend quite a long time on members, so actually who we have at the moment. We don't have very many people, and so that then naturally brings us into the next one, which is additional expertise. Then we're talking about the repertoire, which is basically the part of MSR-2, which is used for Latin script. Then what we do from now, and then we'll finish with questions.

This is an extract from a document created by Mirjana Tasic, and she had the herculean task of making a list of the languages which use the Latin script. This is just part of the list, and there may be quite a major difference between 185, which is EGIDS level four, and 186, which is EGIDS level five.

Because so far, one of the criteria for scopes seems to be the EGIDS Scale, which is a scale used on the Ethnologue website, and it measures how current languages are. So as a rule of thumb, if it's EGIDS one to four and it has a code point, then that will probably be okay. It's very early days, but probably. If it's EGIDS level five to nine, then it's probably not okay. But we are already aware of one exception to that rule, so we have found one EGIDS level five language, which we would like to include.

Sorry about the echo. This is a list of the current members of the panel. What I would like to draw your attention to is that there really aren't many people, and I actually use that word panel very loosely.

We hope to form a panel, but at the moment, we are really more a working group. We hope to be able to form in the near future, but we don't have enough expertise to form as a panel yet, and if I have time later, I will tell you where the gaps are in the expertise. In fact, I'm telling you rather sooner than I had expected.

So, it's a traffic light system. Looking at the members, the idea is red for stop, yellow or amber for maybe, and green for go. This is really our situation. The area we really don't have the expertise is actually technical community. This is really a simple stop.

Most of the others, we have some expertise, but possibly not enough.

Obviously, with something like this, we cannot possibly have experts for every language. This is just totally impossible. But at the moment, we don't even have experts for all the regions of the world, which is really not desirable. We have very few experts, particularly in Australasia. That sort of area, Central Asia. Those are the areas which we really don't cover at all, and I feel we need to.

Here is a picture of the repertoire as it is at the moment. We're really in the stage where we've moved from a situation where we could make very rapid progress. It was relatively easy to find languages looking in the CLDR, which is a part of the unicode website, or on Ethnologue, or on ScriptSource.

But the work is just getting harder and harder now, so to make one red character – red means we think we won't use it, black means we think we will adopt it. To make one red character black seems to require a lot of work at the moment. Certainly for generalists, and this is why we really welcome more specialists because it may well be that specialists look at this and say, "Don't be daft. This is used in our language." So that's really the situation as regards to that.

As you can see, I've picked a screen which is on the edge. There's some black, but the Vietnamese colleagues here may be relieved to see that the Vietnamese is included in that, but plenty of red. Yes.

[off mic speaking]

CHRIS DILLON:

Well, we are interested, not just in Vietnamese itself, but in other languages spoken in Vietnam. I think we are particularly interested in those.

Okay, so where are we now? Add members, very much. Once we've filled the gaps – and they are quite substantial at the moment – then the next stage would be to apply it to ICANN to form the panel.

Possibly what I should have added here was that obviously, one way is to add member volunteers. Another way might be just to say to ICANN, "Look. We're not going to be able to do this with volunteers. What we actually need are advisors." So that decision actually hasn't been made. It's really add members or advisors.

Then once the panel is formed, analyze similar code points also in related scripts. So, as you were seeing a moment ago, we have made some progress with analyzing code points already. I mean, we've slightly got ahead of ourselves actually, but there didn't seem to be any reason why we couldn't at least start the work, so we did.

Then we start to move into the area of looking at related scripts. Related scripts obviously include Cyrillic because the Cyrillic script has many letters in common with Latin. Also, it obviously includes the Greek script.

What is less obvious is that it probably includes the Armenian script, as well, because Armenian includes at least two shared letters. It is most interesting that even two shared letters apparently means that it is not a separatable script. You cannot separate it officially. That is my current understanding. I would obviously be most interested in the opinions of others on this point. But that is quite interesting that just a couple of shared letters may mean that actually we need to be involved with that work.

The question then is really what is the best way of doing that? So should the Latin panel start making lists of things that we think are similar? But then if we do that are we not treading on the

toes of the colleagues in the other panel? So it's not totally clear how we proceed there, but we are open to ideas.

Once that is done, it's create a repertoire and any additional WLEs. These are the rules which are on strings. For example, the famous German sharp S. It's this letter which is used in the German language. That particular letter cannot be used at the beginning of the label. It can only be used in the middle or at the end, so we might want WLE to do that, to say actually it cannot occur in all positions. In fact, there may be other problems with that particular code point. But I'll move on briefly.

Beyond that, we're talking about creating the repertoire in XML. How on earth did it decide to do that? Then finally, it's write a report and submit the repertoire to ICANN. So that's really the summary of what we are looking at.

Any questions about any part of that? Meikal. We're not supposed to use more than two at the same time.

MEIKAL MUMIN:

I would just like to ask if there are currently any plans on how to expand the panel and see how to get for the members from certain communities?

CHRIS DILLON: Completely open to ideas. Yes, I mean, we can obviously make the best of opportunities such as this one to reach out. But really all members of the panel need to be networking really formally and informally. I am reaching out to other people at the moment just constantly.

One thought I've had, and I'd be quite happy to do it, is just literally stand up at the public forum and say, "Look. We suspect there may be people out there with expertise on the Latin script and some time. Would you consider joining us?" I'm quite happy to do that.

MEIKAL MUMIN: I understand that there should be a significant level of work amount on your shoulders as one of the co-chairs. I think it would be useful to actually go out there and motivate more people to join the group. There are different forums.

Since you say we should need further people from the academic community, we can look, for example, at different scientific conference formats. Largely at conference, you can go very globally or you can say, okay, a conference of linguistics, conference of African linguistics, Asian linguistics, [Amerindian] languages. You take your pick, or we could work through certain

books and see if there are specialists who have published on related fields.

I think if ICANN could provide some support for you in that direction to help you shift through these things and give you and other members possibly support to travel to such conferences, I think it would be easier to get further member. Because if we just look within the circle of ICANN and current ICANN members and volunteers, people of the community who are already engaged, I think there's a certain limit we hit by now, so if you wanted to carry that forward, we might make use of additional resources in such a sense. Thank you.

CHRIS DILLON:

Thank you for your suggestions. I have already gone to conferences like the Unicode Conference. I am just about to go to the Association of Asian Studies Conference, as well, so I'm doing a lot of that. But yeah, it's just really a matter of continuing to do this. I tend to find it easier to reach out to linguists. I think I know fewer technical people actually.

HAZEM HEZZAH:

This way, I think you can only get linguistics, not technical people, and as you miss many technical people, I think you could reach out in more technical forums. Maybe Internet

governance forum or DNS forums and so on, you could find more technical people to interest in the issue.

CHRIS DILLON: Thank you very much.

AHMED BAKHT MASOOD: in your presentation, you have mentioned that we don't not have any expertise on DNS, so which level of DNS? Is it a requirement at the ISP level, registry level, or maybe at the root zone?

CHRIS DILLON: Actually, I think DNS is possibly not the biggest gap. It's more practical things. We have people who would be able to say something like [inaudible] moment again say, German sharp S cannot come at the beginning of a label. We know that we have people at that. What we don't have is the people who would actually be able to turn that into XML. That's probably one of the biggest gaps.

MICHEL SUIGNARD: Obviously I won't be part of that panel. But [inaudible] I'm part of the integration panel, and we have quite a bit of expertise on

the XML part. So if that's the only missing part, when you get to that point, I think the integration panel can help you formulate whatever repertoire you have unconstrained.

Even if you have rules that you want, we can for sure help you in creating those rules. And then certainly at the end, you own them. You have to own whatever. Even if you're being helped with creating something, at the end, you have to understand them well enough that you're comfortable with what you do or what was created.

But we can for sure help you in creating XML. We have been doing that for quite a few LGRs now. In fact, frankly, some that are way more complicated than the issue you have in the Latin NGR.

The worst part you may have to do is maybe with some sequences. If you have to use, for example, non-precomposed characters, you may have to use some combining sequence. For some of those, you may want to restrict your repertoire to only some given sequences, stuff like that.

But you're not going to have anything that complex here like you would find in South Asian repertoires or anything like that, or even the variant situation that CJK has. So it is, in fact, a much simpler case. I would not be scared by the XML side as long as

you define your repertoire and you're comfortable with it, the rest, all the XML cooking and all those things as, for sure, you can get help. You don't need to have a complicated advisory system. That can be done pretty simply.

CHRIS DILLON:

Thank you, Michel. That takes a huge weight off me. It is really something I have been very concerned about.

You mentioned as you were speaking, the possibility that we may discover some characters which are in use in a big language, which do not have a precomposed form. I think you may be interested to hear that we found one such character. We then investigated further and found that the language concerned was well down on the EGIDS Scale. So on that front, it is good news, so far, everything is precomposed. But obviously if that changes then that will be something I will be telling people about quite enthusiastically. It will be quite big news for us.

MICHEL SUIGNARD:

I have news for you. Montenegrin has apparently two characters that need sequences.

CHRIS DILLON: I'm not sure whether to say thank you. Yes, I do say thank you because if there is going to be bad news, then we'll need to have it as early as possible. It's not so much bad news as it is that means we are into an area of greater complexity than we had thought.

MICHEL SUIGNARD: Yeah. Just to be clear, the repertoire is not just defined in terms of code points. You can define repertoire in terms of code point sequences. Be clear that the LGR format. The XML format is a bit different than what you had in the old TXT format that people have been using for registries.

Now, you can also define sequences, so it's okay. It's perfectly okay to define your repertoire in terms of not just code points in numeration, but also in numeration of some sequences. You can perfectly define a base character followed by combining sequences, by combining marks as an element of your repertoire.

In fact, combining sequences are not creating any more issues at the regular code points because people come sometimes and they want to create new precomposers. That's not going to happen in Latin. We know that.

Especially for African, you may find in fact that quite a few of African languages will need sequences because they came too late to be encoded at precompose and because now we cannot add precompose to Unicode. They are never going to exist. So I would in fact expect when you do more work and especially on the African side, you're going to find quite a few that need combining sequences.

Also I'm part of the Unicode consortium, so maybe see if I can [inaudible]. I know a bit about that.

CHRIS DILLON:

I have started work on a technical document on African languages. It's called [inaudible]. It's very, very slow work. Thank you for that.

All right, if we could upload the next document. I think we may not be able to upload everything I'd hoped for, but we're covering most of the points I wanted to make. So this one is the proposal. Okay, this is really just to give you an idea of what the thing actually looks like. It's really a longer version of what I was saying before.

We are really keen to get technical people to read this because sometimes quite innocent things – I actually only realized today that apparently there is an RFC which defines things like letter

and character, and I think at the moment, these words are probably being used interchangeably, which is really not good. So it's really to encourage as many people as possible, really, to read it through.

It's a bit of a shame, really, that we can't release it for public comment. But until we form as a panel, we can't release things, so it's an unfortunate system, really.

ANDREW SULLIVAN:

I know I said I'm just listening, but this is not an official intervention. One thing that you can do since you wrote it is you can put it anywhere you like. It doesn't have to be an official ICANN infomatter that goes on it. If you're trying to get people interested in it – because this is actually the first I've heard of this document, and it could be that I would be able to tell you some of the things that you just asked about.

CHRIS DILLON:

Thank you very much for that offer. It's on the ICANN wiki at the moment.

One of the things we do on the panel is that we share. Everything we do is on the ICANN wiki, every meeting, everything. It's recorded, and there are notes, as well. We have some members

who are French speaking, so I write a French summary. It's written in school-boy French, so you'll have to excuse some rather peculiar nomenclature in French, I think.

I'm just going to show you this and briefly, whip through it. You may even spot things as we go. As I go, I shall raise various issues. If you spot anything you're interested in, just stop me. This is the usual way we work. Wait a minute. Have I gone too far? Yes, I have.

One of the things we're looking at is, what I'm calling, identical glyphs, and so there are things like schwa here, which looks exactly the same as turned E, so these things, lower case anyway, look identical. Although, strangely enough in the font I'm using, they look quite different and that's rather unfortunate. But in some fonts, they look very similar.

We're looking out for situations like this, and I think this particular one is actually not an issue at the moment because these may be variants. That's the reason.

UNIDENTIFIED MALE: [inaudible]

CHRIS DILLON:

Oh, yes. Just a moment. Sorry about that. Here we are with these very, very similar looking glyphs. This one actually so far is not an issue because I think we've only found one of them, and so it's actually not a problem.

What does seem to be a problem is the letter I because that is found as the letter I, but also, it's found in an African language in the italic font without a dot, so that's quite close. And also, it's found in Turkish without a dot and not italic, so at the moment, it's looking as if possibly that could be an in-script variance. I actually discovered this only after writing the document, so this probably will go into the next version.

As we come down, there's obviously all of this stuff about possible cross-script variance with Cyrillic and Greek. We mentioned that earlier on, so I will skip that, and we also talked about combining marks. I was going to skip that. I was going to bring that up, but we'll move on.

Oh, yes, and I'm fairly sure there will be people in the room who know the answer to this. We're looking at obviously the various ranges in MSR-2, and three of them were excluded. We know why Latin extended D was excluded, but we don't know – and let's be intelligent and actually use the mouse pointer – why Latin ligatures and fore with Latin letters are excluded, so if

anybody – and I suspect somebody probably does know the reason for that – then I would really like to know.

MICHEL SUIGNARD: Well, it's because they're not part of IDNA 2008 in the first place, so that excludes them right away. The sandbox we live with is what is called a P-valid character in IDN, so if it's not a P-valid character, it's gone, and those are not allowed as an IDNA 2008 character.

CHRIS DILLON: Many thanks. Sometimes when one expects a complicated answer, actually the reality is quite simple. Good, good.

MICHEL SUIGNARD: [inaudible] use a category detail [inaudible] script is because in fact what we call compatibility characters so they get excluded by having an [FKC] transform. So anything that is a compatibility character in terms of Unicode is by definition excluded from IDNA.

CHRIS DILLON: Thank you very much. Indeed, that'll be in the next version.

Then coming down, there's a bit here which is actually on the inclusion criteria, and I touched on this before. So actually, one of our big ones at the moment is – oh, I can never remember this – expanded graded intergenerational disruption scale. I don't know whether I'll ever remember that. EGIDS level one to four.

And then we use this concept of modern use, and so this might be current. Newspapers use the Latin script to write this language whatever it is, or the language is written in the Latin script and spoken by a large – needs to be defined – number of speakers.

At this point, I would like to bring up our exception language because the exception language so far, the EGIDS five language, is Esperanto because although, it's a level five language – and we don't normally include those – there are large numbers of people speaking it – about two million we think – and so, therefore, it has strangely various unusual code points, and so at the moment, we are making the case for those to be included.

Okay, I'll just come down a bit further, and we need to get to page six. There. How have I done that? Okay, so mouse pointer.

There's a whole section on things which are not going to be included, and so I've got these old Gaelic consonants here. So in the old Irish orthography, they used to have lots of dots over

letters, but that orthography is no longer used. They now replaced the dots with Hs, so that's an example of something historic that won't come in.

At the moment, the idea is that Romanization – so things like Tokyo with the macrons – that Romanization is not a sufficient criterion for inclusion. However, I have to admit with you that I'm not particularly happy here because I think with something like Pinyin, it is used by so many people it feels to be on the edge of modern use to me. I haven't found any newspapers, but if somebody found a Pinyin newspaper, that's the kind of thing that might actually make a difference.

The other points I was going to make in the document are relatively minor, and I think in the interests of time, I'll move on, and we'll upload the next part of the agenda. Oh, yes. Thank you. Now, we've got to get this right. It's the first one. Many thanks.

So this is the document that I was talking about earlier. At the top of it, it starts off and everything's very black and there are very few problems, but as we go down, then we run into all sorts of issues. I have missed out. I haven't touched the LDH part, so specifically, the letter part of that.

I think we just presume there won't be any trouble there. These are only the decorated characters, and so what we've got is we have an example language. We like to have several languages if possible and we like to make them a bit more geographically diverse.

In the case of sharp S, then the only big language using it is German. There may be small languages using it, but actually, the big one is German.

But then coming down, this might be a relatively easy way for people to help if they're interested in getting into this to provide a language which is using this code point and is in another area of the world is actually quite a nice thing to do. It just makes it a bit stronger.

One of the issues that comes out of this is an open question really is whether the various demonstrations we've used. We've said, "We think this." I think the whole table should actually be regarded as tentative, so we say, "At the moment, in version eight, we think that this is the case," but that doesn't mean that we may get greater understanding and actually change the status of some of these code points. It's like a snapshot.

One of the difficult things is to work out the criteria for that – I mentioned it before – so we would very much welcome help

with refining those criteria, just really making them better. Also, people who would look at this and say, “This is adequate.” The fact that 203 million people use it, that in itself is enough or whatever those criteria are going to be.

I don't know. I think that reactions seem to differ to this sort of thing. This sort of thing for me is heaven. On one Saturday, I started work on it in the morning, and by the time I came to finish the work, I tried to stand up and found out that actually I'd been doing it for so long that I could no longer stand up. For some people, it's really enjoyable work, and for other people, it's a burden. I suppose that's something chemical.

I think that may actually be all I can immediately think of to say about it. I don't know whether anybody else would like to raise something, or perhaps, we just go further down our agenda. Okay, seeing no hands, let's go on.

Luckily, we actually don't have to share anything more. We could put the agenda up again, but that's all we need to do because we actually already spoke at some length about the volunteer/advisory issue and how best to recruit them, and there's no point really in continuing with that, but yes, we do just need to work on that obviously.

I brought this up already, but it's quite important. We really need to think more about what we do with the cross-script variants, at least with Cyrillic and Greek, and probably with Armenian, possibly with other scripts. I mean, there are just so many solutions that could be implemented here. It could be that we just have a quick look and we say, "Right. Well, we've already done our table and we can see things that look identical."

But actually, things that look identical may not be enough because there is a very important sentence in the procedure. I think Andrew Sullivan will remember this, but in the procedure, there is a very important sentence, which reads: "Finally, in investigating the possible variant relations, generation panels should ignore cases where the relation is based exclusively on aspects of visual similarity."

If we understand that sentence correctly, it leads us to the rather strange situation of a lot of these Russian letters we were looking at before. Let's take the example of Russian R, which looks exactly like the Latin A, so that would be a cross-script variant. But Russian [ER], this is the letter that looks like the English P. This is actually not a cross-script variant because it just happens to be a letter that looks the same as a Latin P. It's a slightly strange distinction to make, so I would be interested in colleagues comments about that.

Before leaving that one, one possibility is that the Latin Generation Panel does some analysis, and colleagues willing, it's not impossible. The other possibility is actually just to wait for the other panels to do it. From a work efficiency point of view, that would probably go a lot quicker, and we would then just say, "Oh, yes. This is what we think as well," or "No. We disagree with you," something like that.

As I say, it's some time away from where we are now, so I'm really open to suggestions on this.

HAZEM HEZZAH: I think the scope of the work is for the top level domains.

CHRIS DILLON Yes. There is also work going on at the second level, but I think there may be colleagues in the room who know more about the details of that work than I do, so I shall not tread on their toes.

HAZEM HEZZAH: Maybe someone can correct me. In the top level domains, when there is a registration for a domain, it accompanies the script, which is the script of the level. It can't be a mixed-script level.

CHRIS DILLON: No. Not at the top level.

HAZEM HEZZAH: Okay, so this cross-script variant case is only for the second level not for the first level.

CHRIS DILLON: No. Cross-script variance could be. Well, when the Armenian panel did their draft it was deferred, and I'm fairly sure that that included cross-script variance, so I presume from that, that we may need to.

MICHEL SUIGNARD: It basically means that you can create a whole script version that looks exactly like a string from another script. Even if it is purely Armenian, you could basically create two or three letters, in fact, more than two in my opinion. But you can create basically something that is totally Armenian but looks exactly like a Latin string. There is quite a bit of example, in fact, of Russian that did create Latin in .com that in fact, looks Like Russian because they have exactly the same spelling if you want in terms of lower case characters. Between Latin and Cyrillic, even if they mean something totally different.

HAZEM HEZZAH: In this case, if someone types a Latin label, it goes somewhere else? It's not the same label which is registered as Armenian?

CHRIS DILLON: Possibly. We'll have to see in the longer term what happens on these occasions, yes.

AHMED BAKHT MASOOD: Actually, it depends on the keyboard. It will not be possible to just press one language's keyboard and go to another language. It may be through some link, but not through the keyboard.

CHRIS DILLON: Yes. If there are no other issues on the cross-script variance thing, that takes us into any other business. Does anybody have something they would like to raise? Yes.

UNIDENTIFIED MALE: Good evening, everyone. I am [inaudible] from Nairobi, Kenya. I'm a [inaudible] support, and I'm [inaudible] this committee. I came here to pass my regrets to you, and then say how I'm happy because of being selected to work with you people.

CHRIS DILLON:

You're most welcome here. We're especially keen to welcome Africans, and particularly Africans who are using languages with unusual accents and diacritics because we are very short of that sort of expertise at the moment. Thank you.

All right. I think I have one or two things here, which I may not have mentioned earlier, so let me just check my notes and we'll try and raise some of them.

There is an additional category of code points whose status is uncertain. We became aware of it because in the Polish language, they have lots of zeds, so they use zeds in several different ways. I think they have a zed with a dot on top, and I think it's maybe zed with an accent. Oh, or is that Czech? I'm starting to get them mixed up. I don't know Slavonic languages very well.

But the main thing is that I believe that for the zed with the dot on top, it's quite common for people to write zed with a bar through the middle of it. There is a police service in Poland that actually writes their name using that in an official situation, and so of course, that then raises the question well, what is the status of such things? You would really like to call them unofficial orthographies.

There are other orthographies which are even less official, so things like Arabic chat might be even less official than that, and so the question immediately comes up, what is our stance on these unofficial orthographies?

At the moment, the stance is the same as for romanizations, so we're saying, "At the moment, probably that is not a sufficient criterion to get the code point turned from red to black." But it's the sort of thing that could conceivably change. Perhaps it's not very likely, but I thought I would raise it for you. Yes?

AHMED BAKHT MASOOD: In that case, Like say, in Poland, they have a cross zed. What is the official status there if it is being used? The police may have some official status for that, so if they have some official status, then it should be included.

CHRIS DILLON: Yeah, I think that's a good point to make. I believe the situation is that the official orthography is the zed with the dot, but that particular police service very definitely has a badge which uses the unofficial variant. I'm pretty sure we're safe to call it unofficial, but all of these things on the edges conceivably could change. That's really the point. Meikal.

MEIKAL MUMIN:

I would just like to point out that in such a particular case, it might also depend on the input methods and aka the keyboards used in this region. Because if they have no way of typing that logo in the form it is commonly written on such labels – such as the official emblem – then it might actually be considered some kind of orthographical variant.

But I think it would be very relevant to see what IPs position is toward such cases. So it really depends, in my view, on input methods or not if they are present for this particular code point, but in the end, someone else will be the judge of that. Thank you.

CHRIS DILLON:

As far as I know, that particular code point is really rather difficult to input unless one happens to know its Unicode code point number, so one has to be quite enthusiastic to use that. Yes.

MICHEL SUIGNARD:

I can't really say too much as IP here because I'm not really supposed to say too much. But typically, there's [inaudible] policy matter basically.

Block variants are typically not an issue. You can put as many block variants as you want. Allocatable variants are always more problematic, so if you start to play with allocatable variants, you may have to be way more careful typically. But block variants, you can, if you want to, restrict your root-level Latin envelope and create variants that are blocking each other. That's your decision to some degree, not us. We're just looking at it.

But allocatable variants, especially creating levels that that can code generate a lot of allocatable variants is typically a very bad idea. That's in the procedure. We don't even have to say anything. We can just point to the procedure to say, "You can't do that," and we'll be not shy of saying that. But that's more like a guideline. You have to go back to the official Bible, and we have the master here.

CHRIS DILLON:

Many thanks for that. At the moment, we have a binary distinction between black, it will probably be allocatable actually, and red, it will probably not be. But we actually don't have a concept of blocked variants, and so that might be rather a nice refinement for our system.

MICHEL SUIGNARD: Let me say it straight. We don't even expect the Latin to create allocatable variants. That's not really expected frankly. We are not going to prevent you obviously because, again, we're not in charge here. We're just basically on the next step if you want.

But typically, I would not expect anyone to want to create allocatable variants on the Latin side. That's my expectation. But you see, you can always create some, and then you'd see the reaction to it. But we're not expecting that.

CHRIS DILLON: I fear that the I case I was mentioning earlier could be problematic here. It looks like an in-script variant.

I would be the usual I, which would be the I with the dot. Then you've got the Turkish one which is the I without the dot. Then you've got the [inaudible] one, which is the African language, which is the I without the dot, and it looks italic. At the moment, unfortunately, all three of those are black. Andrew.

ANDREW SULLIVAN: So this is the second time I've violated my vow of silence. There really is an important distinction between allocatable variants and blocking variants. I urge people to go back to the procedure

and read those principles because that conservatism principle is a really important one.

A variant that blocks is an ultra-conservative thing. What that does is it says, “If you allocate this thing, then you must not do this other thing,” and you just block it. The advantage of that is instead of it creating combinatorial problems, what it does is it reduces the space under which something can happen, so it is a mechanism of conservatism.

The problem with allocatable variants is that it actually is a mechanism of liberalism. It expands the number of things that you can do. Remember that there isn’t a good way for people to know how this thing works.

So in the case of the Turkic dotless I and the lower case other I , which has the additional problem of course that the normal upper casing and lower casing functions don't work the way a non-Turkic user works, there is a problem there.

But there are two things to point out. The first thing to point out is that upper case doesn’t work in IDNA at all, so the fact that you can’t have upper case, to a certain extent, saves you because of folding issues. Although, certainly people are going to input those kinds of things and you're going to have mapping problems.

But more importantly, the ability to say, “This is a blocking variant,” means that for writers who use a Turkic convention, they’re going to know if the one I doesn’t work, we’ll try the other one. Whereas, people who are not familiar with that will not, and so you won’t have these kinds of collisions. So that’s part of the purpose of the blocking versus non-blocking stuff.

I encourage you really strongly to think about blocking as your main mechanism because I believe what Michel said is probably correct. If you go back to the original Latin issues panel report, it basically said, “No way, no how, no variants,” and at the time, we were not making the active blocked distinction.

I think that it’s going to be very difficult for any panel to explain itself if it goes against those recommendations. You’re going to have a whole lot more work to do. Just in terms of eventually shipping something, you might want to think about finding a way that block variants are your right answer. I’m sorry. I’ve talked too long. We’re over time.

CHRIS DILLON:

No. That was most welcome.

I should have explained at the beginning of this that we are well aware that the original report forbade those variants. Yes, I will revisit the procedure and give it a slow reading in this area and

probably badger as many colleagues as I can manage to find on it.

There are so many documents relevant to this. It is actually extremely difficult to keep the whole thing in your head correctly at any time. That's really the issue.

All right. That was actually the last of the points that I wanted to raise, so I think that may mean that we are into date of the next meeting. They are basically every two weeks, and they are Tuesdays at 1700 UTC. Now, I've forgotten which Tuesday is the next one, but anyway, on the list, that will be announced as usual, and you are all most welcome to attend either as members or just to hear about the sorts of things that we discuss. Many thanks.

[END OF TRANSCRIPTION]