Translation and Transliteration of Contact Information PDP Charter DT Meeting TRANSCRIPTION
Thursday 21 May 2015 at 1300 UTC

Note: The following is the output of transcribing from an audio recording of Translation and transliteration of Contact Information DT on the Thursday 21 May 2015 at 1300 UTC. Although the transcription is largely accurate, in some cases it is incomplete or inaccurate due to inaudible passages or transcription errors. It is posted as an aid to understanding the proceedings at the meeting, but should not be treated as an authoritative record.

The audio is also available at: http://audio.icann.org/gnso/gnso-transliteration-contact-21may15-en.mp3

On calendar: http://gnso.icann.org/calendar/

Attendees:
Chris Dillon – NCSG
Jim Galvin - RySG
Sara Bockey – RrSG
Justine Chew – Individual
Petter Rindforth – IPC
Peter Green (Zhang Zuan) – NCUC
Ubolthip Sethakaset – Individual
Rudi Vansnick – NPOC
Peter Dernbach – IPC
Wen Zhai - RySG

Apologies:
Amr Elsadr – NCUC

ICANN Staff
Julie Hedlund
Lars Hoffmann
Ozan Sahin
Terri Agnew

Coordinator: The recording has started. You may begin.

Terri Agnew: Thank you. Good morning, good afternoon and good evening. This is the Translation and Transliteration of Contact PDP Working Group call on the 21st of May, 2015.

On the call today we have Peter Green, Chris Dillon, Petter Rindforth, Jim Galvin, Sara Bockey, Ubolthip Sethakaset and Rudi Vansnick. We have apologies from Amr Elsadr. From staff we have Lars Hoffman, Julie Hedlund, Ozan Sahin and myself, Terri Agnew.

I would like to remind all participants to please state your name before speaking for transcription purposes. Thank you very much and back over to you, Chris.

Chris Dillon: Thank you very much. So let's just cover Point 3 on the menu which is the Statements of Interest. So officially I need to ask all of us whether anybody has had a change in the Statement of Interest since we last met? Hearing nothing and seeing nothing that means we can move into Agenda Point 4.

And for that I would like to hand over to Lars for a moment. Lars, are you on mute?

Lars Hoffman: Sorry, Chris.

((Crosstalk))

Lars Hoffman: Yes, I was on mute. Chris, due to a little situation is there any chance you can actually (unintelligible) right now but can you postpone me until the end? Would that be okay? I'm so sorry.

Chris Dillon: No problem. If I try to forget then hold me back under any other business.

Lars Hoffman: I'm so sorry.

Chris Dillon: That's all right, there's no problem. Okay, that means that we get to deal with Agenda Point 5 which is a - the second look at the second version of the draft final report.

It's a slight improvement on the version that we looked at last time because I went back and added things from our last meeting two weeks ago basically an attempt to respond to various comments that were made during that meeting. Now if it would be possible to have a look at that that would be good at this point.

And for those of you who have the document in front of you what I'm intending to do is to go back slightly in the document around about Page 12 and just pick up some points there.

And I think I might do that whilst we're waiting for it to display because I think we are familiar with this particular piece, we covered it last time. And so on Page 12 we have one of the arguments which is for mandatory transformation.

And this is - I'll just quote it so that it's in our minds. "Transformation would facilitate identification of and response to fraudulent use of legitimate data for domain names belonging to another registrant using reverse query on identity valid data."

And some questions were raised about that argument and in fact when Justine wrote her comments in the comment review tool she picked that up. So I'm wondering whether anybody could provide further information about this. I'm hoping Petter may be able to just add a little bit of information about this, anybody else but I think Petter is the most likely person.

Okay well it may just be that it needs to remain on my to-do list and I can get back to IPC and pick that up in that case. I'll just take a note about Page 12. Okay.

Okay and we've still got some - oh yes, now we are on Page 12 of the document. And it is this thing about reverse query. And we've had some feedback from Justine but also from I think other people just asking for some more information about this.

And we're struggling technically to display the document so I think what I shall do is just - well actually if you use Terri's link then the document is in there and so you can get it up on your screen anyway because of the technical problem. While colleagues do that it's always no bad idea to have paper copies in front of one on these occasions. So much for technology.

Okay well whether we're able to follow that up now or I think it looks more as if I shall need to do it after the meeting so that's okay. And that means that we get to go to Page 15.

Well, as usual, if you have comments on other things you're most welcome to make them but the next - oh now there's - and Rudi helpfully is just - has just pasted the link in there for those of us who don't have a copy of it in front of us.

And I'm now intending to go for a modification made on the bottom of Page 15 or possibly more usefully in Section 5.1.2. And again because we're having the technical problems I shall read it out slowly because that makes it easier.

So the section is entitled, Issue of Cost. And so it runs at the moment, "In its charter the working group was encouraged to discuss the issue of cost in the event of transforming contact information data into one single script. This section provides an overview of the discussion."

"In general those supporting mandatory transformation has argued that costs should be borne by those maintaining the data; registries, registrars, resellers. Those that have opposed mandatory transformation have stated that any transformation costs should be borne by those requesting the transformed data."

Now I think actually the next bit is probably key, the next two paragraphs in fact. "It is clear that blanket transformation of information data would incur large costs. It is likely that any manual transformation would cost a significant amount. Inquiries with ICANN's Translation Department show that transformations under 100 words currently cost a flat fee of between $25 and $75 US depending on the language script from which the transformation is sought."

There's actually - I've come away from the text for a moment. There's actually part of me that would like to do a rough calculation, just say that okay translation costs $50 which is in between the two extremes, and we then have a rough calculation of how many non-ASCII records there would be in the database and we multiply the number of non-ASCII records by $50. And the fear is that that could produce a mouth-wateringly large amount, possibly my being very slightly naughty. Anyway, back to our text.

"Such a blanket transformation at a significant cost would seem inappropriate also because only a small fraction of such contact information data is ever requested and an even smaller fraction would require transformation."

Okay, and we've got the - we've now got the document on the screen, which will mark things a little easier now. I'm hoping that if I scroll down to Page 16 we're all literally on the same page. I don't know how the document is set up at the moment. Page 16 is the answer anyway. Yeah, I fear we all have to do it separate - oh, no we don't, we're in the right place. Lovely.

Sorry, Petter, would you like to raise something?

Petter Rindforth: Yeah, thanks. And I still missed where we were on your initial question. But just a formal note from IPC point of view, and also from - well organizations and companies that needs to the initial search, so to speak.

When it comes to the costs what we want to find the solution for is a way that it's possible to easily make a global search to see - to get the initial contact information to identify the holder of the domain name. And then again as we have discussed during this work, when it comes to if you need to continue with specific legal matters, etcetera, etcetera, that's another thing. And then I can accept that it will be additional costs even for translation and transliteration to get this specific perfect contact information or even to use, of course, the national.

But I think it's - for me at least it's - seems a bit odd to say that anyone that wants to look at the Whois on the Internet pay for it. We have no response for you to have the Whois globally searchable. If you want it that's additional cost for you. I think it's more natural for me and from my view that at least the basic cost for that is for register and administer the Whois database.

Chris Dillon: Well, thank you. As long as the data are machine-readable it should be possible to contact the people providing it. So, you know, that, you know, I think even the original data would fulfill that role. Then what you might be looking at is the sort of ad hoc service which would then provide more detailed services in the event of something - some more substantial contact being necessary.

So, yes, I mean, you know, that would, you know, that would vastly reduce costs if that kind of approach were acceptable. And, Petter, would you like to pick something else up there?

Petter Rindforth: Yeah, I mean, thanks. Petter here. You said a very important word that I think we all can accept and agree upon and that's readable. And I notice that in our draft document here even if it's not - whatever if you choose voluntary or not, mandatory or not, it's somewhere written that it should be readable. And I think that's the important thing even if maybe no details are in, so to speak, in English transcript language but as long as it's readable so that you can actually compare with other contact information or script you have.

Chris Dillon: Yes. So I think we need to make sure that the machine readable aspect is part of the recommendations. And we will in fact later on have a brief look at the recommendations. But I should emphasize that it's only an initial look at them.

We will definitely need to come back and spend a lot more time with the recommendation so anything that we have time to do today is just a preliminary thing because apart from everything else it just brings the recommendations into our minds and gets us ready to do further work on them.

All right, now so that - we really need to remember to come back to that. Okay. And so I'll go on with Paragraph 4 of this section which is at the top of the page and starts with - should be comments I think. Comments, oh dear, oh dear, oh dear. And slightly - more than slightly lost. This should be a paragraph starting with the word "comments."

Right, I think possibly the easiest thing to do is just to read the whole thing and then come back. Okay. So this is actually covering the - that. So - oh sorry, it's actually on Page 15, not Page 16, that was what was causing the chaos.

Okay so at the bottom of Page 15 we've got comments from both working group members during discussions and stakeholders through public comments. "Have pointed out that the costs of mandatory transformation are

likely to be passed on to registrants and in addition such costs could - would hit especially those registrants, registrars and registries in poorer regions in which cost can be a very significant market entry barrier."

"The need for creating new data fields to transform data and significantly overhauling the operational process to allow for transforming data and then verifying them would add to the financial burden of mandating transformation of contact information."

So I wonder if there is anything we want to pick up there or do we get to go a bit further into things? Okay, well hearing no criticism let's go into the next section and this is where my piece on - this is very much preliminary so, you know, we will be coming back here a lot. But we might as well have a first pass at doing it now.

And so this is drafted by me quite recently. Reliable automated transliteration is not available for non-alphabetic scripts and is unlikely to be available for a considerable time. Now what I've done quite often here is I've used the term like non-alphabetic script and then you need to look at the sub-note because that will then describe, you know, what's the scripts that we're talking about here.

And then we've got "See study to evaluate available solutions for the submission display of internationalized contact data, ICANN, IRD study team for further information." Oh, lovely, much bigger and easier to read.

"Many alphabetic scripts, again defined below, and syllabaries," syllabaries tend to be scripts like the sort of North Indian scripts or Japanese where you represent syllables by a letter and then various diacritics, "do not indicate all vowels or word boundaries and so cannot be loosely transliterated." So basically, you know, like with Arabic short vowels are not written. If you don't have your short vowels you end up with all sorts of strange things happening and you just have to guess what they are.

And then word boundaries, so here if you don't know where the word ends then you just end up with everything all in one word which can make comprehension quite difficult, also quite a major difficulty with the Thai language. It's a beautiful alphabet but there are no spaces between the words so for foreign people learning Thai that's one of the oddest things.

Okay. (Looselessly) transliterated, so yes, you can transliterate it but you're going to lose information on root vowels, other things. In Hindi, for example, you know, there's a word for chair which is (koursy), it means chair. But looking at it you can't tell whether it's (koursy) or (kouresy) with a vowel after the R. And in fact there are even some people who pronounce it (kouresy). I think that's not standard.

Anyway on we go. In all of these cases manual transliteration will be required. And then in brackets there being until, you know, whenever automatic transliteration becomes available, you know, if in fact it's ever going to be possible in some cases.

Transliteration of alphabetic scripts would not indicate, for example, and this is really going back to the Bangkok thing so we've been talking about Bangkok and Cairo recently but actually a bigger problem is streets, roads, buildings, so ideally what you don't want is the (unintelligible) which is the first insurance building, you know, (unintelligible) probably have it transliterated but (Bidu) really wanted his building rather than (Bidu) which is much less easy to understand.

"And then sophisticated transformation tools which know when to transliterate, so (unintelligible) you'd want it transliterating probably, and when to translate (Budu) as building, you see, do not exist and are unlikely to exist soon."

"Manual transformation could solve some of the problems outlined above but it is slow and expensive and should be conducted centrally to avoid consistency problems arising from transformation implemented in different ways by many actors." So basically, you know, in a library you could probably get really high quality transliteration happening or, you know, even a mix system transliteration and transformation because you've got a relatively small number of cataloguers and they're all transforming in the same way.

And they're probably rather grim-looking people who send things back to you saying, oh, you've made a mistake doing this and it's all beautifully done. It's possible but it requires quite considerable discipline, the sort of discipline that I sometimes think is disappearing into the past. You know, those sticklers in the library a few decades ago, I don't know how many of them are still non-extinct, perhaps some of them. I hope they're still alive and kicking but I wonder rather.

"Then as regards accessibility data in their original form, as long as they are machine readable, are easier and more consistently searchable." Okay well perhaps pause a few moments for us to digest that.

Before we go a little further down and now we really are into a preliminary part of this because we're into the recommendation section. So as I was saying, you know, we really are going to be coming back so anything that we do now is pretty tentative.

And okay and this is going to be - this is going to be fun. We're going to do the as many languages as possible. This has really haunted us. Yeah, okay, as many languages as possible.

So we've got rather stuck on this, that's the truth of the matter. And I was wondering whether may try and draft ourselves out of it. So if we were to redraft it and say something like instead of - just, yeah, so, "The working group could recommend that the field names of the Whois contact information

be translated into as many languages as possible." But apparently legally this is a seriously bad idea to phrase it like this.

So let's try and redraft. And one of the things I was considering suggesting would be, "The working group could recommend that the field names of the Whois contact information," or wait a minute, "that the new system facilitate the loading of field names in languages as required." And here we get away from the as many as possible.

But what we're really saying is that we would like, you know, there would be an advantage to having a system where it was relatively easy to change those field names in, you know, for the languages that people use in the database.

So something like, "The working group could recommend that the new system facilitate the loading of field names in languages as required." Well I think this week I am going to try redrafting that and that can then go in the version of this document circulated next week. And we can see whether that is better than the original wording. I think it's at least worth a try.

Then, okay, further at the end of this I think this last sentence here can probably come out. So, you know, "The community is strongly encouraged to supply its views on this issue regardless of whether they view mandatory translation and transliteration as recommended." So, oh, don't get me wrong, you know, I very much want to hear everybody's opinion on this but when we finish I think the sentence needs to come out.

All right, so that means we get to go perhaps about another page. The sort of greeny-yellowy text there. Now I've also done a redraft of the footnote but that's quite a cosmetic thing so I won't trouble you with that. But let's just have a look at the main new bit of text there.

"Following the publication of the initial report a public comment forum was opened that attracted 11 submission, a staff summary of which can be found here. Of these submissions eight were supportive of the draft recommendations and three opposed them favoring instead mandatory transformation of all contact information."

"The working group spent several weeks assessing all comments and discussing," oops, grammatical error there, "any new issues that were raised where appropriate they are included in this report. In addition, Annex B contains the comment review tool that was used by working group members to document its discussion on the public comments."

Thank you, Justine, for your comments in the chat room. So next time you see the document that sentence will go. And I should also explain that once we've dealt with agenda point 6, which are Justine's comments, then effectively, you know, the review tool is then frozen but there is one aspect in Justine's comments which I think requires very substantial debate and that's about scoping so we may need to spend some amount of time dealing with that.

And okay we're actually really near the end of this unless anybody's going to raise something under the one last point which is on the background. Now the background may need to be expanded to deal with other things which we have dealt with.

I don't know whether anybody is aware of something urgent that should be added to the background section but the way I'm seeing it it's not just the background of the project, it's actually other work that has been going on and things which we have studied as well. So the idea is to have this section up to date so it indicates where all of this work is coming from, what else exists at this time and that sort of aspect.

At the moment I'm not aware of any substantial addition. However, yeah, I'm fairly sure that there could be. Just try raising it and see if anybody happens to be aware of anything. Not seeing anything coming up in which case that's actually all I'm intending to say about this document today.

Now that means that we need to change documents and move into the comment review tool and ideally it is the version which Justine circulated about 10 days ago perhaps. And whilst that's loading I will - because a lot of - there are really probably two or three large things but I think I can perhaps raise some of them.

So for those of you who have a printout in front of you or can see an electronic version going to start with a comment on Page 5. And this is in Line 24 in the spreadsheet which is "BC supports this recommendation provided the transformation to ASCII is mandatory. We suggest that the language of the registrar service be used to determine the appropriate language." It's that comment.

And Justine picks up "Suggest that all references to out of scope of this PDP," and there are a lot of them, so we really need to get this right because we do slightly sound like a broken record. We've used the term of out of scope of this PDP several times, "be amended to say beyond the purview of this working group and all references to recommended the scope of another PDP be amended to say recommended for consideration through another PDP with proviso that the working group agrees that the subject matter is in fact beyond our scope."

So there are two quite big issues there. One is actually deciding how close to the, you know, the original questions - our two main questions we need to stay. And then it is, you know, whether we use beyond the purview of this working group and recommended for consideration through another PDP or some other form of words.

And actually whether we need to make a distinction between those two forms of words as well, you know, it may just be that for all of these occasions, you know, where we've said something like out of scope of this PDP or recommended the scope of another PDP, all we need to do is actually decide on one form of words unless there's a good case for having two.

Now before we start picking that up I notice Justine is just typing some comments in the chat room. And that goes - by the way, just for information, both links to the agenda posted on the ICANN wiki points to the same document of the review tool; the final report was not - oh, I'm very sorry.

There is a slight bug in that wiki and it means that when you upload a document it sometimes thinks that you're linking to the same document that you just uploaded a minute ago and actually you're linking somewhere else, and I forgot to check it so that's why all of that confusion broke loose.

Apologies. Well it's a very good wiki but all systems tend to have their shortcomings and that is Confluence's problem when one actually uses it heavily.

Okay so personally I have no objection to changing the various references to any of the four possibilities. I think I have a preference for a form of words which is basically along the lines of recommended for consideration through another PDP because that means that there are things here which perhaps we have talked in this PDP to some extent but really it's outside our immediate scope. And ideally it would be picked up by another PDP.

So as far as I'm concerned I would really go for - the whole lot going for something like recommended for consideration through another PDP. So seeing no comments on that I could simply try redrafting it and seeing if that is acceptable to people as a final version of the review tool which would be one of that annexes of the final report.

Okay, try doing that and see what kind of effect that has. Okay and that means we get to go to Page 7 - oh sorry, not this document, it's the other one I'm afraid, review tool.

And that is saying okay so this is actually in response to Comment Number 38. And Comment Number 38 runs that dot(Shabaka), it's the Arabic registry operator, recommends further community discussion to understand better how the PDP's effort and the effort of other Whois related will fit together.

And Justine writes there, "See Response Number 27," is a little bit strange since NCSG isn't endorsing this recommendation and Response Number 37 starts off with this is out of scope. I think to some extent that is just because we sort of put the out of scope thing in a standard position. And, you know, within the response from the working group. That may be ironed out when we change the wording.

Okay and then I have a feeling it's only one last place, that's on Page 14 of the review tool. Sorry, it is difficult, I realize, if you don't have it in front of you this may be difficult to follow. Okay oh and actually we've actually already dealt with this, this is the reverse query thing coming up so and actually we could return to it because this is picked up on the screen so if I can - I think it's on Page 12 of this document so let us return to it because it's really haunting us in two different places.

There it is in all its glory. "Transformation would facilitate identification of and response to fraudulent use of legitimate data the domain names belonging to another registrant using reverse query on identity validator." There it is. I was rather hoping I might be able to get that one out of my (unintelligible) but it is not going to happen. Oh well, one cannot win every battle.

Well in that case, as far as I know, we have now covered both the existing drafts of the - so that's the initial - this is the second draft of the final report and heavens, it's about the ninth draft of the comment review tool. I will

release new versions of those based on decisions taken during the - or preliminary decisions taken during this meeting.

And that means that we're getting to Agenda Point 7, any other business. Lars, would you like to pick up the thing on the plausible necessity for additional meetings?

Lars Hoffman: Sure. Thank you, Chris, this is Lars for the record. So there's two things. I think, Chris, you might want to speak about the remaining timetable for - in principle we have three meetings left before we would have to submit the document that could be considered in Buenos Aires. And what I would suggest is that to make sure that we meet the deadline and to schedule via Doodle two additional meetings potentially on the Tuesday so not this coming Tuesday but the one after and the one after that, that we can cancel if need be but that we have options in case we need to prolong our discussions that would still allow us to meet the deadline.

And I think in addition the idea was, Chris, to focus the discussion during the call on the recommendations that we'll update later this week and have editing on that during the calls and discussions and obviously anything else that comes up can be dealt with on the list, you know, nothing is obviously closed in any way, but to have our calls.

And then in three weeks' time whether it's the third meeting of the fifth, depending on whether we use the (unintelligible) meetings, that concludes the group with the formal consensus call. And, Chris, if you want to add anything to that but I think that would be the plan in order to meet the BA deadline.

Chris Dillon: Yes, I have very little to add to that. I mean, the only thing is that, you know, we need to take some time thinking about exactly how that consensus call works because I think it's very important that we have a system for it so that, you know, basically all members votes one way or the other and perhaps we

need to consider exactly how best to do that and give people plenty of warning of whatever system it is that we set up for that.

I mean, certainly what we wouldn't want to do would be, you know, just the people who happen to turn up on the call on that day, you know, if we were unlucky and we had small numbers then we could end up getting a misrepresentative result so, you know, that's the only additional thing I would add to that.

All right, well we're still under any other business so we - so I can ask is there anything else apart from that? If not I shall release new versions of both documents soon reflecting the suggestions that we've had for various changes today. And we meet a week from today with the idea being that we focus on recommendations but, you know, if you see something in a document and you really don't like it that's fine, that can also be raised next week or the week after but basically the focus is on recommendations.

And unusually it looks as if we finish a quarter of an hour early that I was not expecting. All right, well thank you very much for today's meeting. And we will meet again one week from today. Thank you.

Lars Hoffman: Thank you, Chris.

((Crosstalk))

Petter Rindforth: Thanks.

Woman: Thanks, everyone. Bye-bye.

Chris Dillon: Good-bye.

Terri Agnew: (Nick), if you can please stop the recordings.

END