Transcription ICANN Buenos Aires

Translation and Transliteration Contact Information PDP WG update: GNSO

session Saturday 20 June 2015

Note: The following is the output of transcribing from an audio. 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.

On page: http://gnso.icann.org/en/calendar/#fjun The recordings and transcriptions of the calls are posted on the GNSO Master Calendar page

(David):   And thank you Graeme for your report. I'm on this group. I know there's been

             an enormous amount of work and discussion gone into it. And there's still a

             fair bit to go as you can tell. So thank you for that and look forward to that

             session and hearing from (unintelligible).

             Okay. We can (read) that session and move on to the next - so end the

             recording and let us know when the recording start.

             Okay the next report is translation and transliteration of contact information

             and PDP final report from this working group.

             We have (Chris) and (Rudy) here to present who are the co-chairs of the

             working group so over to you.

(Chris):   Thank you very much. Okay could we have the next slide please?

             So you'll have to excuse me for I think a lot of people in the room are

             probably aware of a lot of this but for the benefit of those who aren't I'll be

             quite thorough.

So since December 2013 we've been looking at these two questions in our charter. So whether it's desirable to translate or transliterate contact information to a single common language or script and who should decide, who should bear the burden of transforming contact information to a single language or script.

We have used the word transformation rather than translation or transliteration because it's just such a mouthful basically right through all of our work.

You can see the timeline below so basically starting in December 2013. And recently we published the final report and that's just about to have the council vote.

This is an overview of our members. As you can see many constituencies covered. And then okay so here's a sort of summary of the arguments on both sides of the fence.

And they're not necessarily in order of sort of major arguments first and then smaller arguments later. It's more it's a logical progression as you go through them certainly on the right-hand side.

So the, you know, those who believe that contact information should be transformed, you know, they should be mandatory transformation of all contact information we're talking about allowing a transference accessible and more easily searchable database.

There was some amount of difference of opinion over the easily searchable bit. I think some people tended to think that actually one of the main things was that the data was just machine-readable. You know, if as long as it was machine-readable you could do a search. It didn't necessarily have to be in ASCII.

So also arguing, you know, the facilitation of communication among stakeholders not showing the same language. And so effectively this is just promoting English as being at least that to lingua franca.

Without the Internet and its benefits expound, you know, this is increasingly less the case. So you're looking at very, very large numbers of Internet users. For example in China and in India for which, you know, this really isn't the case.

So then there's this argument about flight by bad actors to the least translatable languages which to a great extent is actually a theoretical argument.

So at the moment statistics don't really bear this out. They - yes it does happen but it's a tiny, tiny phenomenon. It's not, you know, it's more of a theoretical issue than, you know, one that's actually happening.

So then, you know, when Whois results are cross-referenced it may be - sorry about that. When Whois results are cross-referenced it may be to ascertain whether this same registrant holds different domain names. Yes I think it has some weight actually.

And then, you know, transformation costs are part of doing business. So that's really the argument supporting the mandatory transformation on the other side.

So we have a situation for the next many years I would think where it would be near impossible to achieve consistent accuracy in transforming all scripts and languages.

And, you know, you're talking very often of proper nouns into a common script. I mean it's a very - this is actually a very complex phenomenon, I mean

we would really have to start talking about what the situation is for all the individual languages.

And, you know, some languages have several translation sorry, transliteration schemes,, some have none at all, some perhaps have one.

You know, the tools for translation are really pretty poor at the moment. You know, even from the iGoogle translator, you know, it works well but only in quite limited situations.

So you've also got things like situations where simple systems are going to take something like the original language versions of Cairo, places like Cairo and Bangkok. And they are going to produce transliteration. So you'll get things like Al Khair also Cairo. Excuse my very bad Arabic pronunciation and (Krumcap) for Bangkok.

So actually the problem here is that unless you have an intelligent tool which actually knows when to translate and went to transliterate you're in bad trouble because, you know, you end up with some really rather odd stuff happening as a result of that.

And in fact, you know, looking at that Arabic I mean some transliteration tools wouldn't even give you the vowels in that. You'd end up with a clump of consonants and perhaps a few suggestions about what the vowels might be unless it's a more intelligent tool in which case it may know what they are.

Then accurate translation then needs to be done manually really people doing it. And they will produce results with a high quality.

But as we all know if you got individuals translating or transliterating everybody does it just slightly differently.

So you'd really want to have standards and you'd also have to - you'd really have to enforce those standards very rigorously.

And so as a result of that it starts, you know, it's almost a secondary thing. It's starting to sound quite expensive at this stage.

Then we have a very heavy argument I think although it's quite far down the list. The financial burden could have negative impact on less developed regions which are often don't use the Latin script.

Yes. Personally I think that's the heaviest argument but, you know, everybody probably has slightly different opinions about how they should be weighted.

Now okay so imagine we've actually succeeded and we've successfully transformed our data. It may actually not be very much use if we have because, you know, we may be trying to communicate with people, you know, in transform the English data. And actually they don't speak English so that's the end of that story.

Next slide please. Now I've actually covered quite a few of these but these are just the sort of more detailed information about the mandatory arguments.

So, you know, just a little bit more detail about for example, you know, how much things might cost. And you're actually starting to see conceivable figures coming up there. I don't know $25 a translation. There could be quite a lot of things needing translating.

And then there are very comments so this thing about regular cost of doing business. So, you know, bearing in mind, you know, the likely frequency of the usage of this stuff it's, you know, does seem to be quite a large mallet to crack a very small nut.

And then we've done the usability so the next slide please. And now we're on to the non-mandatory arguments.

So actually submitted data are likely to be consistent, yes because they - the original data nobody's fiddled with them. Nobody's transformed them so they're going to be consistent, reliable.

And actually the more consistent they are the better searchable they are as long as they're machine-readable. That's important.

So equal costs and opportunities for registrars and registrants regardless of their linguistic and script backgrounds.

And then so just starting to head towards recommendations here so language and script should be easily identifiable to facilitate transformation if necessary.

And that might involve some kind of marking. You know, so you might be marking a field as being Arabic script but Urdu language -- something like that.

And also you might be needing to mark each field rather than whole records because you could get records where you've got an Arabic company but actually with an address in Pakistan so you've got an Urdu address. So definitely on the field you actually have a different language and script.

And then so consumers and then the consumers those requesting the data should carry the burden. You know, that sort of argument usually goes on the non-mandatory side of the fence.

Next slide please.

And so out of this come these fixed recommendations. So the main one really it is not desirable to make transformation context information mandatory.

And the burden of voluntary transformation lies with the requester of information.

Data fields are stored and displayed in a way that allows for easy identification of what individual data entities represent. This is really what I was just as saying a moment ago and what languages or scripts have been used.

Languages and scripts supported for registrants that submit that context data may be chosen accordance with gTLD provided business models.

So the idea here is that we're not suddenly - this is perhaps a mistaken impression I gave the last time I spoke to the council. We're not suddenly expecting all the business, provide the business models to support all language of overnight -- absolutely not. You know, they can choose which they're going to do.

Then Recommendation 4 regardless of the language use the fields must be consistent, data verified, scripts and language easily identifiable is really going back to what I was saying a moment ago.

Number 5, so if a replacement system -- and we're saying any. It's actually that's what's a is capable of having multiple data sets. So this would be the original data in one data sent and then transform data in another data set if somebody wants to transform it.

Transformed data should be marked as such and presented as additional fields.

Last but not least any replacement system or, you know, of that is actually a protocol must allow for new scripts languages to be added and expanded to linguistic script capacity.

So you might start off with running a system with a relatively small number of languages but it would be important that it, you know, it would have the flexibility to allow the addition of other scripts and languages in the future. And next slide please.

So what happened when we had our consensus call recently? So Recommendations 2 to 6 received full consensus and Recommendation 1 received consensus with one member not able to support it.

Over the long term of the working group there was a tremendous movement in one direction. But, you know, it did not become a simple picture.

And I think for a considerable time the situation that you see in front of you on this screen this is really what we had for, you know, for a long time. And so I feel it is a representative expression of the situation that we're in.

If, you know, if we continued discussing, you know, what things have changed actually I don't think so. You know, it seems to be quite a sort of stable situation.

And I'll just read out what the minority statement is. And that is working group member that reinforced with line the position taken by its constituency. The IPC recommends mandatory translation and/or transliteration transformation of contact information in all gTLDs.

There are a number of situations where a global Whois search providing access to data in as uniform a fashion as possible is necessary for the data registration service to achieve its goals of providing transparency and accountability in the DNS.

Last slide please, and this is basically further information. So you've got the initial report. You've got the final report. You've got a report which indicates the differences. You've got public comments.

And you've got the wiki page which is an extremely detailed picture of everything that we've done. And, you know, that has included reading many, many relevant reports created by other groups.

I guess over the months that that probably goes up into the thousands of pages. And it's actually makes it sound rather dull stuff.

And believe you me that's the last thing it ever was. I actually had to watch myself that I wasn't over using the word interesting during the meetings.

(David):   Thank you very much. Rudy would you like to add something?

Rudy Vansnick: No I can only admit that the work that (Chris) has been doing and his professionalism in - and expertise in language has been really, really helpful because otherwise I think we would have stranded quite early and not been able to give a good answer. So I would like to congratulate (Chris) with the work he has been doing.

It was a great session during. And in fact it has been kicked off in Buenos Aires two years ago. So we are celebrating the second year.

And we are happy that we are able to drop it now on the table for the GNSO counsel to vote on it. So we hope this is going to be a successful step.

(Chris):   Well that's very kind of you to say so Rudy. But really without your knowledge and understanding of how all of this stuff works I think we'd have gotten nowhere fast. So I'd really like to say some other things to you -- much appreciated.

(David):   I have a question from - (Olivier) you'd like to - you have a question?

(Olivier):   Yes. First of all let me congratulate you on the excellent working that you've done in my opinion in coming to a interesting or a very good solution to many of the problems that were put before you in a realistic and good timeframe I think.

I do have however two questions which relate to Slide 7 that's now before us. The first is to Recommendation 4 which is (consistent) language which is of concern to registrars. It uses the words verified.

That's not what the current requirements of the registration agreement towards Whois data states. Is this misnomership be validated instead?

Was there a discussion of validation versus verification? Is this just using the best word at hand or is this actually intending to verify data that is currently not required to be verified?

So then we probably would have to move for an amendment from the side of the registrars.

Second question, has the question of transfers been addressed intra-registrar transfers would indicate of internationalized Whois data be more difficult, maybe even impossible for some registrations if the data in there is in IDM format and the receiving registrar does not support that format?

Has this been addressed in any form and if so what was the proposed solution for this?

(Chris):   Thank you. I suspect the use of verified in that point may be incorrect because I don't think that we are suggesting a change. That's the first one.

As regards a transfer it wasn't a topic that came up frequently during our discussions I think is the best way of putting that.

(David): Thank you. Question at the...

Jim Galvin: So Jim Galvin. And speaking as a participant in this particular working group I'll add just a little bit of context to the question about transfers of data.

Reflecting back on DNS SEC and, you know, we first put out there and its existing today you in fact have the issue today that a registrant who wants to move their registration from one registrar to another if that gaining registrar doesn't support DNS SEC that would be an issue.

And in fact the transition, the transfer while there's multiple paths here but in essence something fails, either the registration fails or, you know, perhaps the DNS SEC suddenly has to get pulled out and it fails.

So I simply observe that this situation is no different. You know, we're creating an environment in which these kinds of things can happen.

I don't remember this discussion really coming up in any detail in the working group. So I suspect it's just kind of left alone for the market to figure out so just like it was with DNS SEC. I just make that observation. Thanks.

(David): Sorry Amr if you're online can you...

Amr Elsadr: Yes. Thanks (David). This is Amr. Can you hear me?

(David): We can hear you.

Amr Elsadr: All right great, thanks. I just wanted to know to Volker's first question that the mix up between verified and validate is probably my fault.

I was - I think I mixed up the two words in terms of what is actually required in the RAA. There was certainly no intent on the Working Group's part to create any new requirements at all. And this probably should be fixed and I'm not sure what the procedure to do this would be at this point so we probably should figure it out.

I think the absolute intent is not to create new requirements and certainly have no requirements at all on any contact information of the Whois that is transformed by any party.

So all RA requirements should be to whatever data set is submitted by the registrant to the registrant kind of registration and no other form at all and certainly not to create any new requirements at all. Thanks and apologies for that.

(David): Thank you. Do we have any further comments or questions for (Chris) and Rudy?

If there's no other maybe just one suggestion for clarification of that language and I'm also not sure about how the procedure should look like because the motion before us does not of course contain the language of the recommendation. It only refers to it.

But essentially the only thing that would be changed here would be changing verified to validated/verified and then have a remark in accordance with the existing policy or requirements.

That would probably be sufficient to diffuse the language that's in there right now which might - which would be contentious.

(Chris): Thank you very much for that suggestion.

(David):   Thank you. And I'm sure we can organize that before they vote on Wednesday.

Okay so if that's okay we will thank (Chris) and Rudy for their report. And of course we will continue to discuss this item later in the week or in response to the motion about it.

Thank you all. And we will now move on to the next...

Amr Elsadr: And (David) this is Amr again.

(David):   Yes.

Amr Elsadr: If I may...

(David):   Go ahead.

Amr Elsadr: I was looking at the final report now and Recommendation Number 4 actually reads the Working Group recommends that regardless of the language the script uses to ensure the data fields are consistent to standards in the RAA.

Yes. And then it does add verified later on so I just want to note that this does need to probably be changed at some point and would welcome input from policy staff or advice on how this could be done. Thanks.

(David):   Yes thank you. So yes, I'm sure we'll deal with that before the matter comes to a vote on Wednesday.

Okay if that's okay then we'll close this item and move onto the next. And thank you (Chris) and Rudy for your very useful report.

(Chris):   Thank you.