Transcription ICANN Singapore
Translation and Transliteration Contact
Monday 24 March 2014

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/#mar

The recordings and transcriptions of the calls are posted on the GNSO Master Calendar page

Julie Hedlund: Good morning, everyone. As I said before, this is the meeting of the Policy Development Process Working Group on the issue of translation and transliteration of contact information.

And my name is Julie Hedlund. I'm ICANN - on the ICANN support staff on this meeting. We do have also form of participation in the Adobe Connect Room if you happen to be wishing to join remotely or if you want to follow along with the chat room, feel free to go in to the Adobe Connect Room where we also have the slides loaded as well.

So without further ado, I will turn things over to our co-chairs, Chris Dillon and Rudi Vansnick. Thank you.

Chris Dillon: Thank you very much, Julie.

And two thank yous. The immediate thank you is for coming out this early out to this meeting. The second thank you is for all of you who have sent in various contributions to this group over the last few months. They are very much appreciated.

Anyway. As we go along the roll call - I mean, if the people sitting in the white chairs would like to come and sit at the table, then please do so. It's a large room. We got plenty of space and it makes it easier to take part.

Anyway. Can we perhaps go round clockwise, please? Thank you.

Don Hollander: My name is Don Hollander from the APTLD. I'm the general manager of Asia Pacific Top Level Domain Association. This is a - we're not directly affected by this in the CC space but I think it's a very interesting area and one that the CCs are going to have look at going forward. So I'm keen to hear what you're doing. Thank you.

Chris Dillon: Thank you.

Jennifer Chung: I'm Jennifer Chung and part of DotAsia Organisation and also the Registry Stakeholder Group.

Chris Dillon: Thank you.

Petter Rindforth: Petter Rindforth, member of the GNSO Council but I am (unintelligible) Intellectual Property Constituency.

Chris Dillon: Thank you.

Jim Galvin: Jim Galvin from the Registry Stakeholder Group and with Afilias.

Chris Dillon: Thank you.

Julie Hedlund: Julie Hedlund, ICANN staff.

Rudi Vansnick: Rudi Vansnick, NPOC.

Chris Dillon: Chris Dillon, co-chair.

Peter Durrenbach:Peter Durrenbach, IPC.

Sarmad Hussain: Sarmad Hussain. I'm part of this working group but also part of the team which is conducting the study to evaluate the available solutions for submission and display of IRD.

Chris Dillon: Thank you.

Pitinan Koarmornpatna: Pitinan, ICANN staff.

Marc Blanchet: Marc Blanchet, (DSNE) and part of the study with Sarmad.

Lars Hoffman: Sorry. Lars Hoffman, ICANN staff.

Chris Dillon: Thank you very much.

                  Now, Julie, we'd move in the agenda charts if that's okay.

Julie Hedlund: Yes, there is.

Chris Dillon: Oh. Okay.

Julie Hedlund: I don't know if they can - their mic is - we have two people in the Adobe Connect Room. We have Amy Bivins and James Mitchell, ARI Registry, and then we have some staff, Nathalie Peregrine. Thank you.

Chris Dillon: Thank you very much.

                  As a formality, I need to ask whether members - whether there are any members of the working group whose statements of interest have changed recently.

Okay. Seeing none, that means we can move forward to Item 4 on the agenda which - well, there are two presentations. I'm not sure which one is going first. But as we go through the presentation, because it's rather tight agenda, ten minutes if you possibly can. Ten minutes each.

(Steve): Thank you, Chris.

My name is (Steve). I'm the staff support managing the study to evaluate available solutions for the submission and display of the internationalized contact data.

Today we have the study team to give an update. We have been given update to the PDP Working Group and this is also a study requested as part of the GNSO Council request.

So Sarmad will give the update and - Sarmad and Marc will give the update.

And followed by that, Jim from the WHOIS Review Team Internationalized Registration Data Working Group will provide an update. Thanks.

Chris Dillon: Thank you very much.

Sarmad Hussain: Chris, thank you. So shall I start?

So the study has been primarily focused on three different things. It is to document the submission and display practices of IRD, which exists today, in the ccTLD space, for example, or registrar space.

We are also looking at the availability and cost of open source and commercial solutions for transliterating and translating contact data and also evaluating the accuracy implications for transliteration and translation of this data.

The methodology has been to - so we are looking at addressing these questions in multiple ways. As far as the practices are concerned, we've actually designed two separate surveys, one for registries and one for registrar. And that is now under circulation to these - two various registries and registrars. And we are now collecting data from them because it was also shared with this group for feedback.

We are also studying translation and transliteration requirements and metrics for languages and scripts for these entities and also some emergent online. And then finally, we are identifying tools which do the transliteration and translation and evaluating their availability, cost and accuracy.

Before moving on, so can we go to the previous slide, please?

So one of the things what we have found out and what we are also trying to encourage is use of common terminology I think is very necessary so that the same message can be shared.

And we've been sharing some of this terminology with other working groups who are active at this time. I've included some of the terminology in these slides as well. I'm not going go through that terminology but this will remain available to everybody to the presentation. But what we do suggest is that we formulate that common glossary of terms.

What we have been using has been published by the United Nations. And it is called the glossary of terms or standardization of geographic names. And these are - this is a small subset of terminology from that source which is reloading for us.

So there - can we move on to the next slide - in the next slide?

So those are three slides which have some relevant terminology. Just to, for example, refer to a couple of them, basically, the United Nations suggests

that we should be - we can lump together the terms "transliteration" and "translation" and call it "transformation." There is a concept of reversibility which means that it should be possible to - reversibility means it should be possible to get the same result back if you translate or transliterate - transform from one language to another and then back to the same language.

Next slide, please.

At this time, we are thinking - or we're considering three layers of accuracy that we consider as at least necessary to separate out which we're calling accurate transformation. This is the level of transformation, which should, for example, match requirements for legal documents, which are submitted in court of law or, for example, match passport information. Then there's a layer for consistent transformation that is not necessarily accurate but it is useful, for example, to find places on maps and so forth. And then there is obviously ad hoc transformation, which may or may not be based on anything which is reversible or accurate.

Next slide, please.

And here, this slide gives some examples. Accurate transformation is context dependent, not only context dependent in some cases, very arbitrary, and therefore, almost manual verification is needed for accurate transformation. And it does require a combination of translation and transcription and transliteration based on different cases.

Consistent transformation may not be accurate but it is always consistently predictable, and therefore, sometimes also easily - more easily reversible.

But then there are certain issues with transliteration, for example, in systems like Arabic, which are just consonantal-based transliteration would translate into just consonants in English and not readable, for example.

And for accurate example of accurate transformation, there's an example given in Chinese and it's the first letter of each of those words, either changes to (Jin) or (Kim), depending on whether that process is coming from China or Korea. So it's, for example, context dependent.

And ad hoc transformation obviously is ad hoc. So it's accurately unpredictable.

Next slide, please.

There are many organizations which are working in this area. The more significant ones we found include United Nations Group of Experts on Geographical Names. They have many documents out there already which talk about translations, transliterations and, as I said, glossary of terms and so on.

ISO, International Organization for Standardization, also has many transliterations, standards out between different languages and scripts. That list is available with us if that's needed.

Universal Postal Union has some recommendations on not only credentials but also processes.

And then Unicode Consortium, of course, has also some recommendations on transliteration.

Next slide, please.

Very interestingly, we found out that United Nations proposes an ultimate method of making data available from one language to another. And what this is not that data should be only available in a pivot language as has been the case in our discussions here. But they actually say that data can be - should be made available in all languages whatever the source of data is,

which means that any language data can be made in any of that language data.

And what I'm putting here is actually from the United Nations documents. And they say that Roman should be used as a pivot language. And if you have a reversible transformation form of language, you can actually go from any language to any language through them. And that is actually the model, which the United Nations GEGN group recommends.

Next slide, please.

And in many of these cases, there are transformation tables which are needed. However, it is not always possible to find transformation for transformation tables between two language pairs. So Unicode suggests that if the transformation pairs between two language pairs are not available to available standards, there are fallback options, which fallback giving privacy to target and then source and then the variant. So if you have a Russian-English transformation, you would want to use United Nations' tables for transliteration and do transliteration between Russian and English. If you in GEGN table is, for example, not available for a particular language pair at the table, one could fall back on them ultimate table, for example, the BGN table from the (micro viewers) or PCGN table standard from the (micro view) pair or so on.

And if that is also not available, one can fall back on script versus a script to language, so Cyrillic to, you know, English, and then, you know, and so you could progressively fall back until you actually fall back to script-to-script transliteration.

And so there are possibilities available, more than one possibilities available to transliterate. However, it would be interesting to note that in the level at which the transliteration is done perhaps should be part of the data, which is available, so that people who are receiving that they can know whether it has

been Russian to English transliteration or Cyrillic to Latin transliteration or what level of accuracy that transliteration has been done.

Next slide, please.

As far as the submission and display practices are concerned, we are - actually if we have a survey out, that's a ways with you as well so you could look at details. But we are actually looking at how data is collected, stored and displayed by registries.

Next slide, please.

And we are in the process of collecting data from registries at this time.

Next slide, please.

We're also looking at evaluation of - looking at tools and evaluating them. As far as evaluation is concerned, we are looking at the breadth of languages we will cover then we're obviously - we're not looking at tools with just our pair of language because we just have, you know, a laboratory set of tools to work up, possibly a single tool that should cover all the languages and that will be the most convenient.

We are also looking at, so far, this tool. We're looking at a number of languages and scripts they cover, the standards they follow, the accuracy of various language pairs, licensing whether it has open or close, and reversibility of transformations they do.

Next slide, please.

And that's our final slide. So these are some of the tools we've identified. We are trying to get hold of as many of them as possible. We have a set of test cases we're developing for at least three scripts and multiple languages

and we are actually going to be testing each one of these tools on those test cases and reporting back some accuracy results.

And that's it. Thank you very much.

Maybe I'll pass on quickly to Marc and see if he has anything.

All right, thank you.

Chris Dillon: Thank you very much indeed.

Absolutely great to see that presentation. The working group has touched on various aspects of it and, you know, and just really great to see the detail.

Now any questions perhaps about that presentation?

Petter Rindforth: Hi. I'm Petter Rindforth.

I just have a question about your - the last page, the possible transliteration tools. I recognized some of them that I have used in other translation cases and got some quite interesting results. And that was not the positive, I think.

So can you say today that is there any of those that are more convenient, more - give more realistic results than the others? And can you see any developments on this translate - technical transliteration tools to be better in the very near future? Thanks.

Sarmad Hussain: So, yes, there is a lot of variation as far as accuracy is concerned across tools. Even within tools, there is a significant variation and accuracy across scripts and languages. And what we're doing - that's why what we're doing is we actually develop a set of test cases, developed 50 cases per script, and we're doing three scripts and multiple languages within each script to gauge the coverage and accuracy.

And, you know, at this time, I think I would probably not want to comment on the accuracy. I agree there are some tools which are not as good but there may be some which are better. The better ones, we're still not sure whether they meet the expectations. But we will obviously share the results and that eventually people who are the decision-makers on inspecting those tools to decide which tool fits the requirements. There are different tools different - doing different things. Some are doing translations. Some are doing transliterations. Some are doing a mix. So also it depends on what kind of use you want, whether you want accurate use or consistent use. So it really depends on a variety of things. Consistency may not require, for example, accuracy. And that's something which this particular group would probably be more interested in discussing.

Chris Dillon: Thank you very much.

Okay. Well - oh sorry, there's another question.

Rick Wesson: So a couple of points. My name is Rick Wesson. I'm on the Security and Stability Committee. I've done some work on internationalization, specifically around the WHOIS data.

So a couple of points. Getty Placenames Database was one that we looked at that does English and French and was quite interesting for me the postal addressing issues and being able to use place names for areas of the world that don't have highly accurate information for transferring or validating postal addresses.

Google Translate is a machine statistical transliteration of index words. And it's very interesting open source project called word2vec that a lot of that stuff was based on incident and excellent area analysis on how the insights of that work.

And so one of the issues that I have with using non-Roman, Roman, non-Roman transliteration is in the amount of intricacy that's added each time that you go through one of these processes.

And if you try and do several loops through, you find some very interesting and completely inaccurate gibberish that is produced. And so trying to understand the context of the goals that you're attempting to address and listening to ICANN contacts as well would be very useful.

Also, a review of the RWHOIS and WHOIS Plus templates that were developed in the ITF, there's RFPs on those, and they go into a great deal of analysis that's been done some ten years ago on being able to create templates that are internationalized, whereas the data that's presented isn't but it's a presentation technique. But it is centered around WHOIS data. Thanks.

Chris Dillon: Thank you.

Can we have other? Yes.

Man: Maybe just one - other one for housekeeping. Before you speak, please state your name and your affiliation so that we can - for the transcript easily discover who was speaking.

Jim Galvin: Thank you. Jim Galvin. I speak to here now as the chair of this WHOIS Review Team Internationalized Registration Data Expert Working Group update.

Next slide, please.

We've been working for a while. We were chartered. I - and we expect to actually have an interim report that we want to put out the soon after Singapore so we can pull this together.

We have encountered a couple of interesting questions and we have an approach and we want to put those out to the community and, you know, get some input for ourselves to help our deliberations and provide a more useful, hopefully, report out to the community as something to use.

We're not a PDP working group. So it's just, you know, set of recommendations that hopefully will feed into other activities and ideally into this activity, too, with the translation and transliteration working group does.

So next slide, please.

We have two particular deliverables that we are actively seeking. First, of course, is specifying requirements for internationalized registration data. So what that means, what the expectations will be of, you know, the domain names, the registration system, what expectations will be of the main parts of that system and supporting internationalized registration data, and then following that, a data model to use for registries and registrars in support of all of the services and applications that go with that.

Next slide, please.

I want to highlight a couple of things right now that you'll get to see when the interim report comes out. So this could be the teaser or a heads up. If anybody has any immediate comments about some of this, that would be helpful and useful.

On this slide, I want to focus on the group data by categories. One of the things that was interesting for us in the early discussions is what data elements do we think about and which ones do we specify.

And if you look at the systems today, there's not a 100% consistent set of data elements that are used. Different registries have different reasons for

having extra specialized elements. I mean, there is a certain core that you'll find at all of them obviously. I mean, contact information is there, registrar ID, for example, certain dates and stuff.

And as we started to realize that we also noticed then that it would be helpful, we thought, to rather than looking at list of data elements to think about categories of data elements, we believe that we would have better luck defining a list of categories of elements rather than a list of elements and then we could focus on some of the elements in those to help us drive the requirements that we want.

So next slide.

One of the other things is similar to the terminology issue that Sarmad was talking about for the work that he's doing, we also had our context and terminology issues that we thought about. And so one of the things that's interesting is to distinguish between localization versus internationalization of the data.

Localization, of course, would be making the data useful to a natural user. Okay? Where at the point where the user is involved, that would be where the data is localized. And separating that out from the internationalization requirements.

And my view, my presentation here, I would suggest, that what happens is the transition from localization to internationalization is where the translation and transliteration step comes in and making that separation. So these two terms represent the endpoints of what translation and transliteration is doing. Okay?

Next slide.

And this is an example of a localization example. Okay? This is a real world example actually, pulled off of a JPRS WHOIS element. But it's also a look at what - currently it shows a mixed view. But in principle, if you were localizing registration data and you had a directory service and you had made a query about a particular domain name, you'd like to see something like this if you were localizing the data for the purposes of the user that was asking. Okay? And that, of course, would be independent of the internationalized representation of it which would be what the registry or the registrar would have stored or the way in which that they might interact with their specific user.

Next slide, please.

And just give you a moment to look at that slide. So the goal of internationalization is to ensure that you can localize it. So the localization endpoint is - there comes the submission side - making choices for the particular requirements for categories of registration data.

So the first principle we decided at our current census as the most important thing is what we're calling the user capability principle.

The idea here is that a user who wants to provide registration data on behalf of registering a domain name should not be obligated to do anything more than use whatever language or script in which they are most skilled. Okay? You shouldn't put the translation and transliteration requirement on the user who wants to get the domain name. So you want to make sure that, you know, the user is the primary constraining factor. You don't want to put any burden on them in particular.

And the next most important principle is simplicity and reusability. So for the most part, we want to try to use the work of others obviously. Sarmad has also already pointed out that there are the standards out there for all kinds of things and we've been talking about some of the things that he had

highlighted there in our group and we will continue to do that. And the ITF has some documents that are useful. There's some other things around in ICANN that have been done and they'll all be useful.

And then, of course, where there's conflicts -- or not necessarily conflicts but competing choices -- you know, go for the thing which is much simpler rather than more complex.

And so those are, you know, principles that are guiding what we're up to.

Next slide, please.

So we have a set of requirements. We have some high-level requirements. And two of those would be, you know, requirements where the name and address information and then requirements for other data categories. Okay?

So moving to the next slide.

High-level requirement, the first thing I've really already kind of stated it draws directly out of the user capability principle. A registrant, you know, should only have to use the language or script in which they're skilled. Okay? And, you know, we're recommending and suggesting that that's kind of important. That's the best way to, you know, globalize or internationalize for the user community at large.

And, you know, one of the things that's interesting is we observed that, well, the Internet in some ways knows no boundaries. And so you have a registrar who could put up a Web site and they might have someone from anywhere who might come to that Web site. And so you do run into an interesting situation. And I'll get to this in a minute in the example. I'll show specifically what we're talking about here.

But a registrar, it's entirely possible a registrar might not be able to serve a particular user who comes to their Web site. That might be a consequence of how one does this. A user could come and they want to use a particular language or script and that registrar doesn't support it. You know, they simply don't have a mechanism for inputting it and allowing that data to come in.

Although we have served their registries, obviously do have - or rather are motivated to incentivize registrars to want to provide what they need to support the target community. So these become the business issues that go with internationalization. These will surface and they will simply be there and, you know, our current thinking at the moment is that a market will have to sort that out. We're not suggesting that we're going to make requirements to fix that problem.

Next slide, please.

Another important consideration in internationalizing data is all data elements would need to always be tagged with the language or script that's relevant for that data element that, you know, as an implementation, we're not talking yet. We will - in the data model, we'll probably make some suggestions about how that's managed. But as principle, it's important that if you store the data, you have to know what you've got and you can't localize it if you don't know what you have and moving it around. So that just becomes an important principle for all of the data.

Next slide, please.

So the thing to highlight here is the - this now begins the slides are a set of requirements and I'm not going to go through all of these. I want to highlight one in particular for discussion point.

Again, this falls into category of heads up because we'll be describing this in more detail in the document in our interim report and we're specifically looking for comments from the community. I know the folks want to say here, too.

In looking at the issue of the address of a registrant and then, of course, there's also technical administrative contact points and contact information, the question comes up as to what should the requirements be on an address that's entered by a user?

And we have three proposals right now in our group which we're currently discussing and we're interested in getting some comments from the community on what they think of each of the separate ones.

You can imagine the easiest one being you simply allow free-form text. So in some way, the registrar has to figure out the language or script that's being put in and you allow the user to put in anything and then there's no other requirements on that data.

However, if you think of it about data validation as a new requirement that's coming out for registrars, the problem with a free-form text version is the registrar then really has an open-ended data validation or data verification problem. They have no context in which to execute on that requirement in the contract.

So another possibility would be to add a little bit of a constraining factor and that is it can be free-form text but the language or script that's used for the address that's entered should be appropriate for the region in which it's located. So if you're putting in the US address, you know, you would allow yourself to - you would require, you know, English and US-ASCII. If you're putting your French address, you know, you would allow for the French language and French script set of characters that you could use to enter your address.

But, of course, the history that you run into there is that in many cases in some regions, you are allowed mixed script and Arabic numerals or Roman numerals are sort of the common thing, the easy one to put out even if you have - for any kind of language you often will see a mixed sudden insertion of Roman numerals that are put in there.

And so you do have to allow that you could have mixed scripts and - in use or mixed languages in use in the entering of an address. And so you would need to be able to accommodate that and registrars have to be able to support that.

The third proposal down here is put here - in times that you have to account for completeness. One of the questions that's in front of this working group is whether there should be a single language or script in which all registration data appears and/or if there should be translation and transliteration requirements at different points in the system.

And the third proposal is intended to accommodate waiting for the results of this group, you know, should you somehow require that the data somehow has to get into a US-ASCII form perhaps, if there's a single language or script, or you could imagine PLDs often will only support drawn from what they allow for the domain name. They allow a domain name to be entered in certain languages or certain scripts, right? You could buy your domain name. Perhaps you want to suggest that the contact information needs to be in that same language or script, too.

So those are sort of the three things that we're thinking about with respect to address that we really do need some, you know, comments and advice from the community about. And when you see the interim reports, you'll have that in the comment period and we'd like to get people thinking about that now.

If you go to the next slide, you can just pause here for, you know, a few seconds and give people a chance to look at that.

Most of the rest of the categories of data elements have fairly straightforward requirements, at least we hope that they do from our point of view and the working group.

Next slide, please.

We really didn't have too much contention or discussions of this. We very quickly got to consensus on the rest of these requirements. But again, the community will have a good chance to look at those and tell us what they think when you look at the interim report.

Next slide, please.

And you can see here, as you look down the left-hand column, I probably should have pointed that out on the first one, these are the categories that the elements that we have developed so far based on a survey of what data elements are out there. We, you know, divided things up into the categories that you see on the left.

And next slide.

So our next steps here are to have an interim report about the requirements hopefully soon after the Singapore meeting here as we can pull together our text. And then we'll begin work on the data mile while we rate and then incorporate comments from the community. And hopefully, you know, if all goes well, we expect that we'll actually be done and have our two work products completed by London, unless something especially contentious happens from the comments on our interim report. I guess we'll see how that goes.

And next slide is the end. Thank you.

Chris Dillon: Thank you very much, Jim. Another very fine, useful presentation.

Perhaps we should start with the question that you asked about the address fields, whether anybody actually have any comments on the three options. Would it be helpful to display those, that slide again?

Rudi Vansnick: Rudi Vansnick speaking for the transcript.

Jim, you mentioned the Proposal 2 when you say it's appropriate for the region. Is that at national level, country level or is it even in the country to different language levels that are officially recognized?

Jim Galvin: Jim Galvin for the transcript.

And that is exactly one of the questions that is in front of our working group, too. And we had some discussion about that.

One of the things that I'll point out from that discussion - and there are a couple of other people in the room here from the working group. So if they want to add anything to this, that would be fine. I don't - I wouldn't say that we have complete consensus on this point. But we observed that the UPU requirements for postal addresses all have a - the country field.

They suggest that the postal services for a particular sovereign nation, they can have their own requirements for what they want postal addresses to look like. So one could determine that region could mean whatever the postal requirements to find it out and you could fall back on that as a possibility and use it.

But one of the reasons for, you know, putting this question out there is, in fact, to see what the community has to say about what region might be.

We, in particular, have a couple of, you know, registrars obviously in our working group and, you know, an important question is, well, what's the region and if I'm a registrar, am I now supposed to somehow know what all the regions are and then know what all the requirements are in those regions and how exactly am I going to do that and what does that mean?

But this goes back to the business rules that might apply. You know, you're either targeting a particular area or you're not. So if you're targeting an area, then you'll know all the rules because you'll figure them out. You just have the problem of, well, I might have a presence in an area because I'm on the Internet. You can't really hide because there's no boundaries. So suddenly, you're in a situation where you might - a customer might try to come to you and you're going to have to turn them away because you simply don't know how to deal with them.

So these are some of the issues that are there. No solid consensus but see what the community says, too. Thank you.

Chris Dillon: Thank you. This is Chris Dillon for the transcript.

Now there are two questions I think (unintelligible) and then this one right at the table. Oh, and one over here as well.

Man: Well, first of all, it seems that they have made a fantastic work. And when I heard your presentation, I wondered what's left for us to do in this case.

But I have a question around these three proposals. And as I understand, when I freely read it, Proposal 2 seems to be quite a practical one. And you mentioned that - and I guess that was the Proposal 2, the translation of that could fit for the open postal office.

And without going into any details, it seems quite good possibility for me because looking at it on the legal view, as long as you can actually find where the local post or when the police search for a specific address, you can actually identify it even if there are, as we have seen, five or six or seven different language variations of that specific address. But if it's one of these addresses that's officially registered and translated, you know, that's one that also locally can be found in that way. I think that's a good proposal.

And sorry I have quick additional question. You mentioned that you would be - the final report in London. Was that correct? So then I turn to what shall we do before London to cooperate in this better?

Man: Actually to considerable extent, there is cooperation going on. I mean, anybody who gets the weekly meetings almost faster, there's a lot of communication going on.

And then I think…

Rick Wesson: My name is Rick Wesson.

And to speak to the points that Jim was raising around Proposal 3 that if there was the capability for free address - original address templates so that all registrars could leverage such information to analyze the information that is posted by a registrant, then I think Proposal 3 would be extremely - is it Proposal 3? That it would be valuable to have that done. It seems like a reasonable capability that ICANN could leverage some either their technical prowess, their members with IPU, Postal Union, and make this available. It could be quite useful.

And so I believe that it is a capability that does exist although it doesn't appear that that's widely known.

Chris Dillon: Okay. Thank you.

Jim, have you any response to that? Okay.

Peter Durrenbach: Peter Durrenbach from IPC for the transcript.

First of all, I'd like to thank Jim for the excellent report and Sarmad for the excellent report. I think both of these are extremely helpful for our work.

My question is really a procedural question, which is, for this expert working group, I've seen the charter of the expert working group but do you have a wiki online where the resources that the expert working group is using are shared because I do think that as we do our work between now and London, in addition to the interim report, which will be out shortly, some of the materials and references that you've been using might be very helpful into that sharing of information. So I just haven't been able to find it. It might already be there.

Man: I can answer that. Yes, it's online. And I can send the link to a wiki to the working group.

Chris Dillon: Thank you very much.

Man: Marc, if you…

Chris Dillon: Oh yes. Marc.

Marc Blanchet: Marc Blanchet speaking as individual.

Jim, have you discussed, looked at or, you know, whatever, about having multiple versions of those bills such as, for example, two scripts are, you know, the address being, you know, written in two different scripts or languages or either user contributed or meshing translated or, you know,

there was - as you build that this possibility or is that part of the requirements or…

Jim Galvin: So Jim Galvin for the transcript.

I mean, actually that's why they each begin with a phrase "Each of those proposals" say free-form text. Okay? So the actual format that's in there the user - so to walk through an explicit example for how one could make this work and this is not necessarily we don't necessarily have consensus on this point but I'll just - for explanatory purposes, I mean, the user could enter the data in whatever form is convenient for the user. I mean, that's sort of the model that we're headed towards here. So they will just type it in.

It would be up to the registrar who would take it in and would have to know the language that's being - or in principle, the language being used and the script and be able to figure that out in looking at the free-form text and then encode it in an appropriate way in which to store it. And so you would store it encoded so that you could then flag it and would have -- always have -- the language and script tags to go with it for the data going forward. That's what meant by free-form text here.

Marc Blanchet: Okay. But, for example, it's typical in various countries and - that people are discouraged with, you know, a local script, you know, version of the postal address and then Romanized, for example, or other script. So my question is, so they effectively have two addresses. So that was my point. Are you - have you discussed the fact that people could have more than one address field? And most likely in different scripts.

Jim Galvin: Thank you. I apologize. I didn't quite catch the question. But I get that now.

Actually that, at least in my view, the question has been raised in our group but we really haven't talked about it too much, except that we're aware of the fact that, yes, if you had two forms of the address for whatever reason, if you

- whether the registrar creates a transformation of some sort or the user creates the transformation and happens to provide you with both as part of the registration process, you do suddenly have a synchronization issue and you now have a priority issue, which one matters if they're conflict, and what you do about that.

And, in fact, we kind of left it at that just as open questions at the moment. But I would observe that I think that's also a question this group has to address and maybe that's an area for our two groups to, you know, collaborate and talk about. I mean, if you're going to translate or transliterate the information and decide that somebody is going to be burdened with that requirement, you're going to have to say something about which one gets priority if they should ever become in conflict and/or in some way indicate which was the one entered by the user and which was the one that you've created, something like that. So it's an open question in our group.

I remember putting the question out there in the early days. But we haven't gotten to focusing on that particular issue.

Chris Dillon: Thank you. Thank you very much.

Now I'm not sure what the - oh yes, okay.

Man: If I may have a subquestion. Continuing on that thread, the way - it's not necessarily script related but the way an address is structured, for example, the street name and stuff, is usually in different order depending on the script or the completion or, you know, various reasons.

Therefore, a free-form text means that, you know, for one entry, it's there, and the other entry is there and it's kind of - so if we are contemplating any kind of message that, you know, trend something, it might be even more difficult. So as you move that the drawbacks of free-form text and there the

drawbacks are all related to handling various machine, you know, transformations.

Jim Galvin: So thank you for that question.

The only - I can only give you a partial response. In fact, one of the reasons why we're sort of stuck on this address field is because - at least I think because of issues like you're raising. As we look at examples of what addresses look like, you start to realize that, you know, structure is an interesting question, shall we say.

The only observation that you can make about it is the UPU, you know, does say that the individual countries or sovereign, you know, regions that have postal regulations each of those areas specify their own requirements for what's a valid address and what it's supposed to look like. And, of course, even that has lots of options and sort of the - which makes it tough on a registrar who wants to target different communities in different regions.

I don't know how to deal with that and our working group hasn't figure out how to deal with it, either. But I suspect we'll - you know, our goal here is to try and document this question and make that part of the discussion here, especially in the interim report.

And it may be that becomes a question that we even leave in the final report because at some point, you know, I would expect our work product is going to be input to some PDP process somewhere along the way, which we'll have to decide. And if there are decisions to be made about choices, it may be left to that root to make the final choice. We'll just lay out the issues and leave it at that.

Man: Thank you.

Chris Dillon: Okay. Thank you.

We're starting to time out. But two quick comments, one from (Edmon) and one from that gentleman standing at the mic.

Man: Thank you. So yes, two quick comments actually in response to I think Jim's mentioning about the transliteration and Marc's suggestion there or comment there.

I think I hear from Sarmad's presentation earlier, Jim, I think we should bring it back to our group is that part of recommendation in that suggestion was to actually tag it and say what kind of transformation happened in the resulting sort of text, whatever.

So I think that's probably one way to go, which is to tag not only what the transliteration form is and in terms of language or script but also what transformation that took us there. So I think that's a pretty good point that we should take back to our group.

On the other point about the free-form situation and, you know, how the data is (swarmed), I think we've talked a little bit about the - in the group that our focus is more about the internationalization of it, whether the - of course, we need to think about the validity and those kind of things. But what, Marc, you are talking about is probably more about the WHOIS as general case because right now, it is in free-form text. So that's, you know, that level of data structure should probably not be the IRD kind of - the internationalization part but the overall WHOIS discussion that we're having at ICANN in general.

Chris Dillon: Thank you.

David Conrad: Is this one?

Sarmad, David Conrad representing (unintelligible), I guess, also on (FSAC).

I guess I have a question that is, you know, probably is - probably resolved already but I just wanted to clarify.

Based on your question, Marc, it sort of implied that there was some thoughts about the actual use of this - of, in particular, the address field. In a previous life, I was in a position to try to deal with WHOIS data, registration data throughout the Asia Pacific region. And it quickly became apparent that the - how that data was to be recorded within the registration database depended very much on how it was actually going to be used since we at AP didn't really have the particular use for that data. We didn't place any restrictions on how it was to be encoded by the registrant, allowing, you know, basically making it impossible to parse, you know, in a machine sense.

So I guess the question that I have is, you know, how the actual use of these fields then identified in the context of, you know, is it WHOIS or is it some other purpose, you know, statistical, you know, gathering or is it for identification of end users for serving legal process or that sort of thing.

Chris Dillon: Okay. Thank you very much.

Jim Galvin: I'll make a comment. One of the problems is - one of the reasons why - I'm sorry. This is Jim Galvin for the transcript.

Let's start again. I think that one of the motivations for moving towards categories rather than data elements was, in part, because the question that David is asking is somewhat difficult to answer for our purposes. It would be nice if the directory services expert working group was, in fact, done and had decided what that was supposed to look like because, in some sense, we would know then what the data elements are that we're dealing with. That group, in fact, is defining purposes of registration data and that's driving in some sense the set of elements and, you know, maybe we can move towards having a little more standardization and consistency with respect to all of that.

You know, we had that problem at least in the working group that I'm chairing in our expert working group if you just don't know what data elements you're working with. And so it's a little difficult to talk about the purposes of the data. And that's why we sort of drifted, at least in my mind, towards categories of data elements and figured, "Okay, someone else is going to fix that problem. Let's, you know, approach this from a place that we can and provide some usefulness."

So I know that that's probably not a satisfactory answer for you. You know, I feel like we at least are aware of the question but we don't know how to answer it. So we chose an approach that gets us close. That would be my response for the moment. Thank you.

Chris Dillon: Okay. Thank you very much.

I think we're just about to go over the time. So I'll just wind up very quickly.

Just to say we do have two replies from SAs and ACs, (ALAC) and CSG. We haven't set time today but very interesting (unintelligible) on our next call.

We've also got a lot of talking to do about the variability, you know, should it's transliteration or translation and not likely to be a matrix. And that will also be covered in the near future. (Rick) has been doing work on that.

I think that's, more or less, it. Thank you then. I turn it over to Rudi.

Rudi Vansnick: Yes, Rudi Vansnick speaking.

We are also trying to figure out some answer to some of the questions about what's next and then what are we going to do with the data if it comes late for us.

The planning is that for April 24, we will try to do the - this update and then try to find out where we are, what data we have and what we are missing so that we can start creating a review of where we are and what we have - still have to do.

And another deadline in the agenda is June 25th, the London meeting, where we will have again the face-to-face working group meeting and where we will try to present the work we achieved up to those - that moment.

It is clear that it is not going to be an easy task as we know that we are influenced by a lot of other activities that are going on and especially what the expert working group is doing but I'm also referring to the study that highlights a lot of things that we have to consider also and that we will probably be - the biggest task to produce this matrix so that it is - James has been also putting in his slides, make it simple and understandable because, otherwise, we will have a lot of discussions on the understanding and because this was already in our discussions in the beginning, a critical issue what are we looking at, what is our mission, what is our task and where are the limits of our task.

So that's probably the work that we will try to finish for June so that we have something we can really work on to start producing maybe an interim recommendation to the board or to GNSO.

Chris Dillon: Thank you very much for that one final comment before we go to the welcome ceremony.

I think everything is in the wiki. I mean, obviously, we'd like to add some text from the two presentations today but - so particularly for people who don't, you know, don't participate very often in the group or, you know, every now and then, for new people, you know, if you look in the wiki, you will see everything. You know, there is nothing else - as far as I know, there's nothing major.

Thank you very much indeed.

Julie Hedlund: And I'll just thank you everyone for joining and also for those in the Adobe Connect Room. There is some food outside. Grab something on your way to the opening ceremony which is in (unintelligible), which is just over off the raft. And thank you again.

END