

**Transcription ICANN Los Angeles
Translation and Transliteration Contact Information PDP WG meeting
Monday 13 October 2014 at 15:00 local time**

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/#oct>

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

Julie Hedlund: Welcome, everyone, to the translation and transliteration of contact information working group - PDP Working Group, sorry. That's just, after a while, a lot to say. My name is Julie Hedlund - I'm with ICANN staff and welcome to this meeting. And I'll just go ahead at this point and turn things over to our chairs (Chris Dylan) and (Rudy Vonsnick). Thank you.

(Chris Dylan): Thank you very much. This is (Chris Dylan) and I think we should - I'm Co-chair. (Rudy Vonsnick) is my Co-chair sitting next to me. I think we should start by going around the room and then we'll do it Adobe Connect. So (unintelligible) (Edmund), would you like to introduce yourself?

(Edmund Chung): (Unintelligible) good luck, too. (Edmund Chung) from (unintelligible) Asia.

(Jennifer Chung): (Jennifer Chung) - also from (unintelligible) Asia.

Man: (Unintelligible).

Man: (Unintelligible) and programs history.

(Frank): (Frank) (unintelligible) from NCC.

Woman: (Unintelligible) from Tai GAC.

Man: (Unintelligible) from (unintelligible) staff.

Jim Galvin: Jim Galvin from Afiliias.

(Chris Dylan): Thank you very much. And is anybody in Adobe Connect who is - who hasn't just said something? Anybody there?

((Crosstalk))

(Chris Dylan): Okay, well into the next point which is the statement of interest point. And this is a formality really. But we have to ask members of the working group if anyone's statement of interest has changed since we last met. I'm seeing nobody raise hands or anything. I think we can probably move on beyond that agenda point. Now, I'll just give a brief introduction to the meeting.

We're going to start off with an update which basically tells you where the working group is at the moment. There is then a short section in the presentation which is background and that's typically for people who are either new to the group or just need that sort of general information. We will then spend most of the time having a look at the - and note the stressing in the following phrase quite carefully - the draft initial report. So the main thing about it is it is not the initial report, it is only a draft.

This meeting is really a working group meeting. We are - you know, obviously, we welcome members - well, we welcome people outside the working group to attend and to contribute. But fundamentally, it's a meeting, a lot of which, I think, will be about the reasoning which is in the draft initial report. I thought I would just like to stress that before we got the bird off the ground before we got going. Thank you.

Okay, now, have I forgotten anything or can we just get going with the - it feels as if we can get going with the presentation. Let's do just that. Okay, so here are the two main questions which we need to answer these are the questions exactly as the two questions above are the two main questions exactly as they are in the charter. So whether it is desirable to translate contact information to a single common language or transliterate contact information to a single common script.

Now, because it's rather a mouthful, we have tended to use the shorthand transform. And transform covers most of that point. So it's either translating or transliterating. But it's used in the sort of sense of that first point. And then the second one is who should decide who should bear the burden of - but this is exactly (unintelligible) - the burden of translating contact information to a single common language or transliterating contact information to a single common script.

Now we have also spent time on answering other questions. And the three you see at the bottom of this slide are also in the charter. But there were some other questions that we addressed at various points as well and just relevant questions as they came up. Next slide, please. So, bit of an update. We asked for input from SOs and ACs and elsewhere and we have been discussing that in our weekly meetings.

And just a couple of notes - the ICANN publicists study that - so - actually, at the moment, no registries or registrars are transforming registrant data which is interesting and also the protocols are lacking support for IRD -- Internationalized Registration Data. And none of the current tools provide a high level of accuracy and consistency.

Moreover, they cover quite a narrow range of languages when you look at the world's languages as a whole. So then, I wrote a strawman proposal some time ago and in the strawman, basically, what happened was that I put - I basically wrote a transformation should not be mandatory. And that didn't

mean that nobody should do it. That just meant that it shouldn't be mandatory and that anybody who needed to do it should be helped so to do.

And we were also quite interested in practical help for anybody in that situation that's - that was the situation with the strawman. At that point, we became aware that a clear consensus that not emerged. And so more recently, there was the draft initial report. And in the draft initial report, both sides report. So you got the non-mandatory arguments and you got the mandatory arguments and you got recommendations on both sides.

And really the aim of that was to improve the reasoning to get a debate going. And oh, boy, did we succeed. The mailing list came to life and we had a very animated discussion. So I think we must regard that experiment as being a true success. And so already, during one of our last meetings, in fact, we managed just about to get through - oh, now, heavens, I've forgotten exactly where we got - but I think we got through most of the mandatories and non-mandatories and most - and I think we did most of the recommendations as well.

And then, after the call, I then sent an email which tried to pick up what I believed were the main points in the area of the document which we didn't have time to cover. That didn't meet with much comment. So it may be that there are people who want to pick that up at some time this afternoon. That is absolutely fine of course.

So, yes, but anyways, more or less, a description of where I think we now are. And next slide, please. So next up, so here we are in LA having a face-to-face meeting. The aim is to publish the initial report - no more draft - after the meeting. That will be put out for public comment. And so that is the time for formal public comments and obviously the reply period as well.

And the idea is that we're aiming for publication in Morocco or shortly afterwards - various URLs at the bottom of that slide. I think - yes, basically, if

you just go for the Wiki page, I think that may cover all the others. I like Wikis as you probably all know and so I tend to stick absolutely everything in Wikis of one sort or another. Next slide, please. And in fact, we'll go on to the next one because I'll do the brief just beetle through this very quickly so that we've got a certain amount of background as well.

Okay, so the background to this working group is really the gradual internationalization of the domain name system and specifically really the internationalization of world contact information, need for this standardized query of IRD -- Internationalized Registration Data -- ongoing reforms of directory services. And there's an expert working group on that. And so basically, it was necessary to have all of the - as regards that IRD and whether it was desirable for transformation to take place or not.

And so, yes, as we had said, there is the EWG and those being regular communication between us and them thanks to Jim to a large extent. Next slide, please. So more background - so we've got, you know, the who is system, basically using ASCII - the Latin alphabet. And so no standard implementations for character sets other than ASCII. But very, very large numbers of internet users not using our ASCII.

So that's really the general background. And then an anticipation that the next domain name registration data service may not be ASCII-centric. Just make a comment here that what I've seen of it so far is it is ASCII-centric and we may have opinions about that. But at the moment, I suppose there is a longer-term hope that there may be a system that is not ASCII-centric and that it may be possible to move straight from who is to a system which isn't without going via another system which certainly, personal opinion, I think there's very little point in replacing a system twice; smoother to do it just once, sort of slightly after-subject - not - I see it rather strongly.

Next slide, please. Okay, and then we've got definitions of these terms which we use the whole time -- translation - yes, conveying meaning basically. So

with a translation, you can translate the same word and get more than one result. So if I say the word - or the Japanese word (foreign language spoken) - we could translate that as right or we could translate it as correct; it doesn't really matter. But the issue there is that with translation, you can get more than one thing and it is perfectly acceptable.

Transliteration - this is the switching from one script to another and it should be done using a particular system. So there, as long as the system is sufficiently developed, the result actually should be the same although the rules to use it may be complicated. The other complication is that again, if we take Japanese just as an example, you can actually have more than one transliteration system for a language. So Japanese has about two in common use and then you get sort of various people doing their own things as well.

So yes, anyway - but this is just the rendering. And sometimes that can even be done automatically if it's an alphabetic script. Okay, next slide. And then, a definition of contact information which are - let me think - sort of names, organizations, postal addresses, registered name holders - oh, wait a minute. So organizations and postal addresses and it's the name holder technical contact and administrative contact. Then something down there about the RAA specifying data elements which much be provided by registrars but with no requirement for those elements to be transformed.

And next slide - which I thought was the end - good, good. Any questions about any of that?

(Rudy Vonsnick): (Rudy) speaking for the transcript. Maybe it's good for those that are new in the group and have not had the chance to look into all of the documents that we have been looking into. In the actual data model, there is no space foreseen for having - for an address to different languages in the data model, no space to put in. And for instance, a Chinese address and an English address for the state registrant, we thought we have seen in the - coming

from the expert working group also is that in the future data model there is also no space foreseen.

It's quite important to know that if we would go in the direction of needing translation and data to be stored into the database of the Whois there is a need for extra fields -- data fields -- so that you can capture the information. And it's important that if that is not available, you cannot even go into a translation. That's a limit.

(Crosstalk)

(Edmund Chung): (Edmund Chung) here. I'm not sure I necessarily agree with that. I may be mistaken. If I'm not mistaken, when, for example, EPP - during the working group business discussed and there - EPP does allow multiple addresses to be submitted. Not necessarily all the registries support it; you know, that is given. But the protocols do not disallow it. So there is a difference.

So, some of the - not all of the registries support currently, but the protocol does. So, you know - and I think, Jim, perhaps you could add to that because we - the expert working group talked a little bit about this - well, I think, wait.

Man: Yes, thank you very much.

Jim Galvin: Jim Galvin - so, yes, sure. I mean, think about the fact that from a straight up technical point of view, you can have different types of contact, technical contact, linked contact - those kinds of things. So to that extent, you could certainly have multiple sets of contact information. What's not in the protocol is a relationship between two of them.

So for example, if you were to have two technical contacts and you wanted to say in the protocol in the data model that these things were related or intended to be equivalent in some way, there's no way to say that. So that's something that would have to be done over and above all of that. The other

thing that I would offer is if we go down the path of translation and transliteration and you're going to have equivalent information - two sets of equivalent information.

The other thing that we get from the solutions study that was done is also to realize that we will then have to add some additional meta information that will have to be tracked because you'll want to know which one has priority if they should ever not become equivalent. You'll have to talk about what the equivalency rules are. So, you know, who's responsible for ensuring that they're equivalent and then, you know, what do you do if they suddenly fall out of being equivalent for some reason or another?

And you'll want to probably note when the transformation took place and log the timing information that goes with it. So there's some very real - not necessarily overly-complicated - but certainly a bit of challenge on a technical level to deal with transformation of data that will need some attention if we're going to recommend that it needs to occur. Thank you.

(Chris Dylan): Thank you, Jim. I wanted just to pick up the word equivalent and information in what you were saying because in a sense, when transformation goes well, the best case is an equivalent situation. But very often, you will look at original data and then you will look at transformed data and find that some information has actually been lost. So - oh, yet again, if we use Japanese as an example - Japanese tends to come to me very easily.

So for example, you might have a Japanese address and you've got, then, the transliterated version of it. But you actually can't get back to the original because you don't know which character made that pronunciation. So that would be an example of they're really not quite equivalent. They - you know, the transliterated one just about works but it's not quite as good. And then, you mentioned, also which has priority.

And I think it's quite difficult to design a system where a transformed version would have the priority because usually, if you look at the cases, it will be the original which will have additional features. Jim, would you like to add to that?

Jim Galvin: No, I guess I see - I don't want to repeat things. So, no thanks.

(Chris Dylan): Okay. (Unintelligible), did you want to say something? I thought you might just - (Edmund).

(Edmund Chung): (Edmund Chung) here. So, I'm - I asked kind of a little bit of this question in the GNSO session over the weekend. I'm ask - I'm trying to ask it in a better form, I guess. I ask - I wanted to ask the group and (unintelligible) what do we think is the - is really the objective of - if we're going to do transformation - what is the objective of that? And also, you know, in relation to that, how do we define accuracy?

The reason why I say that is I tend to think that whatever is transformed, a transformed version of it, we could easily add a disclaimer and says, you know, this is not - you know, this is a transformed version and, you know, we can't guarantee anything. But this is for your use so you can read it, you can, you know, more easily copy it on a piece of paper. Does that achieve what we want to do? And if that does, then it's a very different question than what we're asking.

So, how do we define accuracy and what is it we're trying to, you know, really achieve?

(Chris Dylan): Thank you. I shall be slightly naughty and say to you that rather than answer your question about - why transform now, I would like to leave that till later because I'm hoping we're going to go into it in a huge amount of detail. And if I try and do a brief version now, it sort of stalling what we'll do later which would not be ideal. Now, the accuracy thing, yes, I think - so if you can imagine looking at two addresses - so one is the original address and one is

a transformed address - you could ask somebody - a native speaker from the country, perhaps or, better, somebody working for the post service.

They're very used to addresses. You can say is this an accurate transformation? And you know, he or she can say yes or no. It may also be possible in some cases to do it mechanically and say, you know, has this address being transliterated correctly from the Cyrillic - yes or no? So those would be accuracy questions. The interesting thing is that you could have an accurate rendering which may not be consistent.

So if you had a rendering which had been done using the (Nit Bonsky) transliteration, it would look quite different from one using the Hepburn transliteration. But they're both accurate but they look different.

(Rudy Vonsnick): (Rudy) speaking. I think it's also important to keep in mind that the address information is essentially necessary and used to send posts to their address. Doesn't have any other sense than being able to send a document or contact the organization or the body that is mentioned in a physical way because, otherwise, you don't need an address. So I think it's important that if you start transforming data of an address into whatever other language or script, it's quite risky that the postal service would not be able to deliver the document due to the fact that that doesn't exist in that language in that country.

(Chris Dylan): Yes, I think that's - I mean, the example I was getting was postal service that is extremely used to Japanese Romanization. I mean the likelihood is your letter will get there. In other countries where Romanizations or transliterations - Romanizations a sort of transliteration - where there aren't really standards of transliteration. There really is a serious risk the letter will get lost. Now, I think there were a couple of hands going up there and (Pat), would you like to get...

(Manny): First, I think we can say - to correct that point - that we agree that there must be priority to the original as you're saying when you - I'm not fully - agree that

the address are just used for postal services, but when it comes to postal services -- local postal services -- of course. And if it comes to local dispute to identify. And it's - I mean, it's - that's the normal when you want to do business agreements and you translate them and you use the translator forms in a general matter.

But there's a small text not readable for everyone if you don't have your glasses on. But stating that the original text is the English one, if there are any differences. So I think that's one point that we have to consider other hand. But what - how I see it and from RPC point of view, again, it's not just for postal services. It's for actually to in a easy way identify as safe as possible the order of a domain name. And then, of course, if you then want to send out a letter or anything like that, you have to do the second step.

But in order to have some general global translation or transliteration of the contact information, something that's still necessary and important.

(Chris Dylan): (Manny), thanks. And I think my head has become rather full of postal workers and it's not helpful. Okay, (Edmund), would you...

(Edmund Chung): (Edmund Chung) here. So would you - yes, I think I agree very much that the - I guess what is the original form needs to be the primary form and the accuracy for which when we talk about who is accuracy and whether we can reach the person, that form needs to take precedence. And in terms of that form, we need to remember - at least I think we should reaffirm one of the fundamental goals of this whole exercise is that a user, let's say, from China registering a Chinese IDN, they could use their Chinese address and not be discriminated against that when they register a, you know, a GTLD domain.

So that's a fundamental goal. Then, now we go - in terms of the - any form of transformation, I don't think I agree - actually, in fact, I'm pretty sure I disagree that the aim is to create a form where we could universally transform it into - let's use English as a - for example. I don't think the goal is to

translate it into something that we can put on an envelope and expect it to be sent correctly because if we want to just achieve that, we can always use the original form, right.

To print something on an envelope, you can use the Chinese address and you can send it from the U.S. to China without a problem. One thing particular in the expert working group that we, you know, I think we retained is the country code. As long as you have the country code, the letter will get sent to that place without a problem. If - of course, if the address is correct, right. So as long as they say CN, the postman will throw it over to the CN pile and once it gets to China, it gets delivered.

So there isn't any problem in terms of postal or if the original is correct. Now, on the transformed version, I think, you know, from some kind of disclaimer is important - that we cannot guarantee the accuracy. But what (Chris) mentioned in terms of accuracy versus consistency, I think, merits some sort of a discussion. I think, in that case, consistency is more important than accuracy because - why I say that is because a lot of times these data is being analyzed by law enforcement or other people.

And if - the consistency of that transformation allows people to match, you know, oh, this address registered so many domain names. The accuracy is less important. The accuracy is important in the original form. The consistency I think is more important in the transformed form. That's, I think, my views on it.

(Chris Dylan): Thank you very much. This - some very interesting points made there. I think we - all of us - agree that it is very important to have a system where Chinese people and other people using non-Latin script are able to enter their data in their own languages. I couldn't agree more with that. Now, as regards transformation, yes, there are some people who - don't believe in transformation tool. They just think the original should be used; that's one standpoint.

And, you know, that's one of the opinions that we'll be having a look at later in the meeting. Jim...

Jim Galvin: Jim Galvin, for the record. A question that I have as we explore this question of whether or not transformation should be mandated - one of the things that I hope that we answer is why we think that's important? What's the motivation for doing that transformation? (Edmund) had, in part, asked that question early on in one of his early comments here today but we certainly asked it of ourselves along the way. And that is what is the purpose of doing the transformation?

And I - we've talked about a number of the issues that go with performing transformation - you know, what it means to have done that or not to do it and the effects that it has in other areas. But I am concerned that we haven't spent enough time talking about why we think it's important to do it or why we think it's important not to do it. I guess we should give equal time to both of those questions. Thank you.

(Chris Dylan): Thank you very much. I hope we can go into that in much greater detail a little later.

(Izumi): ...related to the point that Jim had mentioned. So I don't know if you want to - oh okay. So, hi. This is (Izumi) from (unintelligible). Well, I very much agree with the points that Jim has made. And to add to a little bit to that, maybe we can try to think of specific cases. How would those translated or transliterated information will be used instead of just, you know, think in terms of ideology? So if we think that maybe law enforcement agencies will use in certain way, we can see whether, you know, they actually need the translated data or maybe they have somebody to do this kind of work.

And so I think it would be good to have more specific examples of how the information will be used.

(Chris Dylan): Thank you very much. Absolutely.

(Rudy Vonsnick): (Rudy) for the transcript. And I think it's good if we end up in having a final recommendation, for instance, going in one or the other direction that it would be helpful if we could give some guidance in any case to know whether you have any guidance of existing tools or platforms where - that people can use for doing translation or transliteration so that if we say it's not mandatory, it's - we are not going to do it - that we still give the community the possibility and a guidance on how to best translate contact information that you find and the ways that can also be a helpful thing.

I am taking your example from the Chinese address that's valid as long as that address is in China. I just could imagine that that address would be written in Chinese but is an address in the U.S. I'm afraid that the U.S. Post will not be able to deliver without translating or transliterating it. So, you see it's - we have to use the right context. It's okay if it's in China, but what if the domain name registrant is living in the U.S. and is putting the contact data in Chinese - what will happen then?

So there are a lot of cases as a few have mentioned. I think we need to have an overview of all the possible cases so that we can give a right and a correct answer to what is asked from us.

(Chris Dylan): Thank you. (Edmund).

(Edmund Chung): (Edmund Chung) here. I guess in response to that question, I think that is somewhat out of scope of what this - we can certainly talk about it. And I think that's out of scope because that's a matter of who is accuracy, right. That's a - you know, right now, there is a requirement that who is contact information is accurate which means, you put it on an envelope, it could be set. So that's - you know, this - we can talk about it but I think it's slightly out of scope; it is an accuracy issue.

And if that's - if you provided a U.S. address in Chinese, it literally means that's an incorrect, inaccurate who is data which, well then, you're violating the rules at this point.

(Chris Dylan): Thank you very much. Jim...

Jim Galvin: So, Jim Galvin, for the record. So there is this other expert working group - Internationalized Registration Data - requirements for the submission and display of Internationalized Registration Data Expert Working Group. And to add to the point that (Edmund) was just making, currently, that group is going down the path of suggesting that registration contact data in particular that the language or script that used needs to be appropriate for the region in which that language or script is located.

You know, that's one of the options that's being discussed in that group. It's - there were three folks you'd remember from the preliminary report - interim report that was published about six months ago that it was one of the options. And we're currently, you know, headed down that path with that particular recommendation. I mean, if that continues, it's important to understand that the particular example that (Edmund) just gave of, you know, U.S. address written in Chinese would simply be wrong.

It's not a valid address and would have to be rejected. You know, there are interesting requirements or constraints that follow from that principle of going with your script having to be appropriate for the region. But it'd be useful to keep that in mind as we talk about transformations, what we want to do here. Thank you.

(Chris Dylan): Thank you very much for that and especially the wording which is, I think, very useful to us as well. All right, I wonder if we could perhaps display the document. And specifically, perhaps, roundabout Page 11, I think, may be a good time to start. (Unintelligible), would you like to ask a question?

Woman: Hi, before we go into the report, I just have to - have some comments with (Edmund) as well. I think it's quite a primary answer we have to find out if why we are transforming this and also, according to Jim. If I may, I share what the Thai government trying to do is the accuracy of the (unintelligible) or any contact information is not accurate is not for the who is the card itself only but any (unintelligible) or anything to make sure that you get the correct entity you are talking to and your parcel go through all the way by mailing list.

Or that, I think, also the weight is also the validation more than what is the language displaying. So, given that if the address is already validated somehow, then it's quite easy to transform to any language. You can see something entirely - just copy it in Google Translate and you have the idea of what is it going to be like in English. So it really depends on why we are translating this.

All right, so if you do translate for trying to validate the information that already translated, then it is something. But if you already validated that information, then the transform is just for the display. So it is - quite depends for us. Internally, we already have to make these kind of information valid. So we had some of the system that we have the regular check.

If you can, like - if you are a service provider, you can send a bill to the address and it's still reach, he pay the bills, then we call it - the post to tell us that this address is active for this person. And then we go from that already validated address to do whatever - you can transform, you can do whatever. So the question will be if the address is already okay and if I can require that you need to transform it into English to put it into the whole system, then we do the transform.

And also, if we want to transform, there should be some standardized way to transform. So for us, we kind of base on some of the (unintelligible) system. So the - some of the characters will become "p"; some of the characters will

become “i” like my name. So that will be some sort of expected behavior of the transformation as well. So that’s roughly how we are trying to do. And we do believe that if there is any cause happen, it should be minimal.

So maybe it’s the role of the government as well that have to set up this kind of system. It’s quite huge that you have to put everybody - every stakeholder down and say, okay, this is going to be the standardized system of our country to transform - to validate and then you go from there.

(Chris Dylan): Thank you very much for that. We may need to follow up a bit on this after the meeting, in fact. But anyway, much appreciated. Eleven, I thought. Now, I’m plunging into page eleven. If somebody feels there’s something earlier they would like to pick up, then they are most welcome so to do. But I think a lot of the really meaty things start roundabout - okay, yes. Okay, I was just going to say that’s not - ah, yes, that’s actually fairly good, isn’t it.

Okay, and so this - these are the arguments supporting mandatory transformation. So this is - so within this draft document, at least, there are the two sides. And this is start of the argument for mandatory transformation. And - pardon me - we’ve, already in the group, spoken about this to some extent but I think it’s important that we revisit them because certainly we are aware that some of these arguments are stronger than others.

Okay, well, if you scroll - sorry, Jim.

Jim Galvin: So, Jim Galvin, for the record; just a quick comment here. I am a little cautious about use of the phrase easily searchable because the truth of that statement depends on the choice of script that you use and language. Okay - and so there’s a little bit of technical detail behind that. I mean, another important point, of course, is easily searchable is depended on a canonical form of whatever you put. And what I mean by that is this is much more of an issue in languages other than English, of course.

But you often have, as you look at the Unicode code points and you'll be familiar with this, sometimes you'll have a particular character as a single code point and you also have multiple code points that you can combine to make a character. And the question is, when I search, do I look for both of those? And if I don't, you know, I mean, in order to ensure that searching works, you know, you have to canonicalize one or the other of those into the other, so to speak; you have to pick one of those two to be primary.

This gets into the whole issue of - I also add to this discussion about variants and stuff like that. Anyway, I'm reacting to the use of the phrase easily searchable and I just want to make sure that that's on the record here that I don't think that's quite the right phrase to use there. There's a lot of caveats that go with a phrase like that. But for the purposes I think of this discussion, it's probably fine for right now. I just don't want to lose that detail at some point down the road. Thank you.

(Chris Dylan): Yes, thank you very much for that. Yes, I mean, I suppose it is true though, if you have the choice between searching in the database which have every language in it. And searching in the database which only has some sort of Latin script version, there are transformations that can be done using the Unicodes so that even if there are different ways of representing characters, they can be used standardly represented in the database in a certain form.

And so I feel - yes, it probably does need to be there but, Jim, pick me up.

Jim Galvin: Yes, so I - now you just reminded me. I thought there was one other point I wanted to make here. And also the assertion that having unlimited number of languages or scripts creates a greater burden, I'm not sure I agree with that either because, strictly speaking, it doesn't really matter. You think about it, if I'm going to search, I'm going to enter the characters I want to search for, okay. And that's just a token I'm going to put and I want to look that up in the database.

Frankly, a question that I would ask is does it matter what the language or script of that is? I'm simply going to look in the database for that sequence of code points and whether they exist or not and then it'll come back out to me. All right, so in that respect, the number of languages or scripts that might be in the database is not really the issue. And I don't think that has any bearing on searchability.

The only real question that I have about searchability from a technical point of view is whether or not the data is normalized or canonicalized - that's the issue that's important, okay, because you can't search for something that doesn't exist in the database even if you know as a person who would look at visually or as a person who would understand it semantically, you know they're the same thing. But, you know, computers are fussy like that, you know. Is it - the representation is what matters.

(Chris Dylan): Okay, I certainly agree with you on the normalization front- I think we may need to agree to defer on the script front. (Larz), would you like to...

(Larz Hoffman): Yes - sorry it's (Larz Hoffman) for the record. There's no question really - I mean, when you search the who is database or any future database of contact information, don't you usually just search for a domain name to see who has registered that - meaning that you need the - or you need the - the address that you have anyway, you have an address of for whom you would likely - the owner or the person. And so you look for that and then something comes - I don't know; something comes out.

So whether that's an IDN address - if it's an IDN address, it is an IDN address; you have to look for IDN anyway. If it's - if an ASCII address, you have to type in the first ASCII. And the result, then, you know, you're going to really search for first name or street name - I don't know.

(Chris Dylan): Okay, thank you. (Unintelligible).

Man: I'm talking from different kind of identities - personal identities. Whether it's a private search for information as it - families which domain disputes and when it comes to new assisting clients, if this domain name is free to register or if there is, well, searchability issue or to - just to, again, to see who is holding similar domain names. So I can't say from all these different kind of use that if you'll say that you search for the address or you search for the domain main.

In certain cases, of course, it's if you start with a domain name and then the second step is to see who is the holder? And sometimes, of course, it's more important to start to see - if you know a specific individual or a specific company. I mean, for instance, sometimes you got new clients stating that can you just check out what kind of - what is our domain is? We're trading globally and our subsidiaries have been dealing with 12 or 15 different kind of companies that have sold them domain names.

And now we need to get everything in one specific spot. And those kind of searches, actually, I found they are the most problematic ones because then you have everything from okay, there's a list of their trademarks to see if they are all registered or it's a list of their company names and their subsidiaries and look for those in specific countries to get - well you can hardly get 100% say, for full list - but as much as possible, to identify to see if this is related to that specific company or (unintelligible).

(Chris Dylan): Okay, this idea of what we're searching for certainly sounds to be important now. I think it was (Edmund) and, yes.

(Edmund Chung): Okay, (Edmund Chung) here. So in response to (Larz's) question, I think - and also building on that - I - most of the cases, it depends on how you define what is a searchable who is, right, and also where you're searching from. And we're talking about searching from the registry or registrar website. In that case, I think exact matches or for domains or exact matches for contact are really currently the only searches we allow. So - but it seems to me that this is trying to address - so even, you know, a more macro set of possibilities.

So this data might be reciting somewhere that that eventually someone tries to search. But in that area, I think, you know, again, if we only look at what registries and registrars should provide, that narrows our scope a little bit better and, you know, our (unintelligible) agree. And right now, there only requires a complete match, so some of this argument might not be valid in that context. It's valid in the context of, let's say, some trademarked lawyer creates a database with all the who is information, this might be useful, yes.

But that's probably true. But that would be some database separate from the registry/registrar interface directly. And that would be somewhat outside of our scope, I think. So if we focus on that, you know, what we need - registries and registrars or within the ICANN community too that might be more useful, I think. And on that, I'm jumping ahead a little bit but it seems like the last point you have there is also about searching contact information and probably should combine.

And the other thing is, I think out of all the points, I'm going to be very drastic and say that I think only the point - the second-last point is valid. All the other ones are not something that this - you know, as ICANN, if we think about a policy at ICANN direction, it's probably not relevant for us. But I'm being a bit drastic in that.

(Chris Dylan): Thank you for that. The reason I was looking slightly puzzled earlier was that I wasn't expecting searching to come up right so (unintelligible). So, yes - so yes, there may well be this duplication there. Okay, now, we had question from Jim here.

Jim Galvin: I - you know, just to close this issue of searchability, which you can and can't search - the register agreement is very clear on what you need to provide if you're going to offer searching abilities for GTLDs. And the new register agreement lays out quite precisely what you have to do partial matches on, what you have to do to exact matches on and total matches on. So that's

actually pretty well-specified so I don't think we need to get ourselves, you know, hung up on that particular question.

You know, I really do - just to bring this back around again, to focus on the technical implications - the technical implementation implications of choosing to transform the data and what that means. And so normalization is my comment. Thank you.

(Chris Dylan): Thank you. Okay, and now I think it's (Clause).

(Clause): Maybe I'm too naïve or maybe I'm as (unintelligible) as other people. Who is has a very, very strict function as it says on the name -- who is behind the domain? And that would be the only concern. All other concerns -- how can I search, who can I search -- is on - is none of the business. And that's why a lot of these arguments are just simply not valid.

We have to go back to the basics. It's there to find out who is behind the domain. And then we have to see if it - if that address is how we can make it available for people to use other kind of scripts (unintelligible).

(Chris Dylan): Thank you very much. Okay, now, just working our way down, the other point. Oh, yes, Jim.

Jim Galvin: Okay, so, Jim Galvin, again for the record. So I'll comment on the second one there. It's - I'd appreciate if someone could expand a bit for me. It's not immediately obvious to me how that second bullet point supports mandatory transformation because it's saying here that, you know, if you don't properly transform, then you might lose clarity - well, sure, you know, I mean if you don't do it right then you don't get the right results.

But it's not clear to me why that's a point that supports mandatory transformation. So anyone could expand on that - I'd appreciate it. Thanks.

(Chris Dylan): Can't help you there. No, I think, as I was saying earlier, you know, there was an attempt to list the arguments quite simply similar (unintelligible) than others.

(Rudy Vonsnick): (Rudy) for the transcript. Maybe it's good to mention that the text is a basis of discussion and is not the reporting itself; it's just a collection of all the information we have been gathering. And we've put it in the document in order not to forget that this has been mentioned. It's not at all - it's an intation to report, I would say, rather than a report and in order not to forget valid or important information. That's why it's there.

It's not what's going to be there at the end, it's just a collection of information that we don't want to lose in the track of trying to get a good report.

(Chris Dylan): Thank you.

Jim Galvin: Okay, so - Jim Galvin again. So, you know, I mean, I understand that; I appreciate that this is just sort of a first draft. And so, I'm commenting, you know, in that context and I get all of that. So I'm just highlighting the fact that I think that we're trying to find places to put all of this data in the comments that we got. I mean, the first one made sense to me.

If we want to step through these, I pretty much have a comment about each of these and how it's not immediately clear to me that they actually do support the idea of mandatory transformation. And I - you know, I just put that in here for the record because we're talking about our initial first draft of it, that's all. And I'm just looking for some discussion on these points because, ultimately, when we come to our final report, if we can't expand on these and explain them better, I think they, obviously, have to be moved or taken out or something. So...

(Crosstalk)

(Chris Dylan): Yes, thank you.

(Crosstalk)

(Izumi): (Izumi) - I'm not sure if this - I haven't really fully read the report so maybe it's already covered. But in the last meeting, I commented to consider translating or transliterating the content and there's also an option of just translating the field. And that might really make a huge difference for what the registries or registrars are required to do. If it's just their field, it's very easy - once, you know, even if they don't have the ability themselves to translate the field, they can just, like, outsource it and that's just done once.

And they can, you know, put that in as a default option where if they have to - I mean, I think there's another argument of who's going to do that - but, you know, that would actually make it much easier. It doesn't involve each of the registrants having to translate or transliterate or worry about registrars doing that. But that would allow people, what are the contents of that information without having to go through that if they think that certain information is necessary, they can do the translation or transliteration themselves to reach to the people they need.

(Chris Dylan): Thank you; that is in the document as a possibility.

(Izumi): Okay.

(Chris Dylan): All right, so - might as well just keep on working down through them. So if we get to number three, this is pointing out the lack of IRD functionality. Not completely clear why that's here if you think it through because - yes, it needs to be in there somewhere but not necessarily there. I think - and, Jim, yes.

Jim Galvin: Yes, so - thank you. I mean, I'll just make my one comment again. On each - I think on each of these remaining bullet points, I just have the standing comment that it's not immediately obvious to me how this supports

mandatory transformation. And if we're going to use this as a basis point for supporting mandatory transformation, I think we need more information here and some more detail here - at least I need it.

So if someone could fill that out for me, that'd be great. And I'll not comment - I'll not make that comment on any of your other bullet points so that you can move along from here. Thank you.

(Chris Dylan): Thank you. Okay, so just check - I'm just having a quick look to see if - so unnecessary legal actions, yes, okay. Oh, sorry, it's (Edmund).

(Edmund Chung): (Edmund) here. I think the point on the not currently has no IRD functionality, I think it's kind of self-defeating. I mean, that's the reason why we're going through this...

(Chris Dylan): Yes.

(Edmund Chung): ...macro (unintelligible).

(Chris Dylan): Yes.

(Edmund Chung): So I don't think it needs to be here at all.

(Chris Dylan): No, it needs to be in the document, but probably not here. I think...

(Edmund Chung): Right, but not - the next one, in terms of avoiding unnecessary legal actions, I tend to think the other way. You know, it's actually a con for this because if you mandatory - you have mandatory transformation and the transformation happens to be inaccurate, then you are attracting unnecessary legal actions. So I'm - I think it's the contrary. As I mentioned earlier, I think the second-last point - the second-last bullet where it says, you know, to facilitate communication among stakeholders who don't share the same language, I

think that's the single, you know, most important factor for - you know, for this.

That's my personal opinion, again.

(Chris Dylan): I mean, even with that, okay, I mean, there's almost a certain bias that we're thinking of the internet as it is now. Now, you know, once you - once the internet expands and more and more people who don't know English use it, then this becomes weaker and weaker.

Man: Just to comment here as it was from (Clause's) initial comment - and of course, I mean, if there are two-hour technical problems to get it, that's another thing. But the initial thought here was that, again, to find not 100% safe system, but a system where you could actually identify not to send posts, but to identify that this seems to be the same holder. And this is obviously not the same holder of the domain name - just to get that basic initial search.

And I mean, I've seen some comments I've had - some comments about oh, lawyers pay, they can just invoice their client service - no problem. They can use another system. But I would say it's not just for huge companies are doing these searches; it's also for your individuals, private persons in the individual companies that have actually a need to get a global initial informative system. And then, of course, if you choose to take the next step or if you find it necessary to take another step to send a letter or something like that to the holder, then you have to pay for someone to get that further information and add a local attorney, etcetera, etcetera.

But it's also actually to try to avoid that - these initial expenses and work to - if it's possible - to have the initial searching system that is globally readable, so to speak.

(Chris Dylan): Thank you very much. Now, we'll take (Clause) and then (Jim).

Man: (Larz).

(Chris Dylan): Oh, was (Larz) before?

Man: No.

(Clause): (Unintelligible) (Clause) about the last point - searching contact information is easier in one language. This doesn't make sense of me because what are we looking for? We are not supposed to search the contact information, so why are we saying searching contact information? So that point doesn't make any kind of sense.

(Chris Dylan): Okay.

(Rudy Vonsnick): (Rudy) speaking. I think in generally, the wish is this information that is not publically searchable as such it can be searched through a lot of systems but it's not for the general public. So to the point of how can I have a - or look into - what organization or the - for instance, just the name of the registrant and using the name of registrant to see how many domain names that registrant has registered is an action that you cannot do today. You can't; it's not available.

So we have to be careful in addressing the point of making searchable means that we are commanding that you have a database that you can publically search and I think that's not really correct. We have to avoid to put this in a public context because we are going to oblige to make the who is data publically searchable and that's quite dangerous I think because we're going to break a lot of rules. I can be wrong.

(Chris Dylan): Actually, let me just jump in for a moment. I've just realized this is - I think this is Version 8 of the document. There is a Version 8A as well. One of the improvements in Version 8A was the deletion of that last point. Jim...

Jim Galvin: So, two quick points. Again, just a reminder, the register agreement specifies exactly what needs to be searchable and what you have to search on and it is contact information because you have to be able to search on a name - the registrant's name, organization name - on an address line on the city/state, you know, province - those fields have to be searchable, you know, to pull them up. Second point I want to make coming back, I was just reminded of something again here in this discussion about searching.

The system to date in the legacy GTLD systems is such that when a registrar, you know, creates a new domain name and goes to register it, the typical thing that is done is they create a new contact object and associate it with that domain name. So the ability to search on a name is actually quite valuable because that contact information block could exist especially for people who are domain-holders and have large portfolios might exist 100 times or 1,000 times in your database.

The same thing - it'll be a - obviously, a very different specific registrant ID or it'll be a different ID in the registrant database. But, it, in fact, is the same person because a new object was created. The new agreement changes - the new GTLD program changes the business model for that kind of stuff. And that stuff is likely to go away now because when you validate, okay, you don't want to have to validate again.

All right, since you never had validation before, you could always just create a new contact object - who cares? But now, since you have to do validation, registrars will be motivated to create that contact object once, do the validation that they have to do to it, and then the next time you want to get a domain name in that same contact object, you're going to reuse the contact object because if you create a new one, you got to do validation again.

That's what the rules require and that's the way all of that works. So the needs for searching are sort of different than they were in the past. So I'm

just trying to, again, point out the fact that we shouldn't get stuck in this discussion of what are the purposes of searching and why you have to search. Again, it's in the contract. And then, two, the whole business flow for this kind of stuff is changing anyway under the new program.

And, you know, we shouldn't focus on the purpose of searching; just acknowledge that it exists and - but there is a technical issue. I'm going to go back to my normalization comment earlier. Thank you.

(Chris Dylan): Thank you, Jim. That's also sounds like a powerful argument or the other - you know, the other side to the argument that we're just about to look for. And I don't think it's in there so many thanks. Okay, well shall we have the other - oh, question from (Edmund) first and then the other side of the argument.

(Edmund Chung): (Edmund) here. So, want to make sure that (Peter's) point is taken. And I think, you know, it's missing here in the list of bullets. I think - and I think it's a very good and valid point which is that a very good use of this is, you know, a regular user - I'll use English as an example for now. It's very difficult to compare to who is results whether it's the same guy but that's, you know, very useful for many cases.

So this is useful so that I can compare whether these few domains are actually registered by the same guy. Yes, you can probably cut and paste and do some comparisons, but if it is - it transformed, it's much easier to see and read and, you know, somehow get a sense that it's coming - you know, it's registered from the same address or from the same guy. So I think that that is actually a good argument to be - to add it and why we want to support some sort of transformation.

(Chris Dylan): Okay, I must be getting tired. I just can't - for me, I have to admit, in a situation like that, if I was verifying that - two set of information with the same - I'd really want it to be an original because, you know, if I'm in Japanese, I'm

seeing characters; I can tell it's the same person. So I don't really understand.

(Edmund Chung): But you're a - you know, let's say you only speak English and you're working for a law firm and you want to just take a look at these 20 domains and see if they are, you know, the same guy, right. And it's supposed to take you, I don't know, 15 minutes to do it. And if you're looking at those Arabic characters, it's very typical to know that they're the same. Yes, you can try to cut and paste and see if it - you know, do a - you know, but that's a little bit more technical.

But - so, again, to - it actually ties into facilitating communication. So again, I still think that's the main point for any transformation.

(Chris Dylan): Okay, sorry (unintelligible).

Man: Just short - and again, what we're talking about here is the basic information. So, I mean, as you say, to find that they - there seems to be the same addresses. And again, the next step is to proceed with a original address.

(Chris Dylan): Thank you, (Larz).

(Larz): (Edmund), I'm just wondering - your example. So you're talking about I only speak English, I want to know who these 20 addresses are in. First result comes back; it's Arabic. And then you're talking about the other 19 are also Arabic and I can't - it takes me a while to figure out that they're all the same because then I could, theoretically, once I have the first result in Arabic and I know which one of the lines is the name, I could obviously then do another search with that name and then all the domains would come up, no.

I'm just asking; I'm not saying this is a solution. I'm just trying to understand.

(Edmund Chung): Okay. Just - (Edmund) here - so yes, that is the case. But it's just if you walk through, let's say, what I would normally do when I'd try to investigate, you

know, certain things, right. When you try to investigate, the first thing you do, you search these names and you take a look at, you know, do they look the same. And then you go to the next step and actually put, you know, cut and paste and do all those. It is easier for the first step to happen.

And not saying that it is - like, it's definitive or anything. Yes, no, no, no. So that's the (unintelligible). And if it's transformed into English ASCII so I can, you know, quickly look at it and say, ah, they cut them, they're probably they same. Okay, let's do more investigation.

(Rudy Vonsnick): Well, (Rudy) here. It's quite - yes, it should work as long as the registrant has used the same methodology to add the contact information. You can easily put the number of the house before the street or behind the street and it can still be the same address but it doesn't look the same to - just giving a sample case. So it would work based on as Jim was mentioning. If you have an object that is the registrant's contact details and whatever domain is registered for that person, it would use that object to define the contact details.

But still, as this today, it's not the case. It's - I can register domain names with 20 different forms - formats of registering the - my contact details. So it's more complex than that I think. I saw (Izumi).

(Izumi): I very much agree with (Rudy's) point. And so this might be getting a little out of scope of this working group. So we want to identify whether it's the same person, I don't know in the domain world do we use the (unintelligible). It can search the people by the (unintelligible) and if it's the same, then you can - and it's the - an alphabet. I don't know about who is but - so I think that would give you the identity.

And sometimes people have the same name. So I think that would be more objective way of identifying. But I can see your point, (Edmund), if some people are looking at various different who is and one in Arabic, one in

English. And although it's not 100%, maybe want to get a feel of, then maybe it can help to assist to an extent.

Man: Yes, and also it's not - unfortunately 100% clear if you have identified the first address because, I mean, you have to check out translated or not - the same person could actually spell it with the first name first and the last name last and the next domain, they exactly the same holder could be with the reverse. And also, as you said, the street address with the number first or the number after or sometimes, I've got - I remember, I had one dispute case where I got actually ten different disputes.

And it was up to me to decide whether it was the same holder. And I had to go through the addresses and see if, okay, the name (unintelligible) and in some of them there was a different telephone number and then some there was a different street number. But then I, certainly, after some work, figured out that it seems to be one of the same two 95% safe. But still, there were differences. And these addresses were perfectly readable in English.

So it's - you - sometimes you actually have to see each address and each complete information in order to make the identification as safe as possible.

(Chris Dylan): Thank you very much. Which I think means we can have a very quick look at the other side. Okay, I've, personally, very little to say about this. I think this is the version that still had - oh yes, right at the bottom, there's talk of registration information data which I think should probably be contact information but that is probably really small stuff. I mean, I don't know other...

(Rudy Vonsnick): (Ruby)...

(Chris Dylan): Yes.

(Rudy Vonsnick): (Rudy) speaking. I think that in the context of saying that there is the registration information data is, in fact, all the elements being contact details

of the registrant, contact details of the technical contact, the administrative contact. That's the full set of registration data that's maybe what is mentioned - or meant here.

(Chris Dylan): Okay, I mean, it - you know, that may well be the case but I don't know, I think there's almost an instinct within me just to sort of stick to - with our remake if it may be the case. But actually, what we're really dealing with is contact information. And in fact, at the end of the page, is precisely the area where (Izumi's) point is made about the data field names. That's where it is. Jim, you had...

Jim Galvin: I just - Jim Galvin for the record - just want to say that I agree with (Chris). I really have no idea what that first bullet is trying to say. I don't find that it's a comment in support of not - in support of not having mandatory transformation. Trying not to use too many negatives there - I think I said that right.

(Chris Dylan): I think the idea is just that to - you know, say there were to be transformation, that it would be very, very difficult to transform all of the languages. That in itself is difficult because we literally don't have the tools in many cases. But, you know, then, within each language - you know, to give the Japanese example again - we're talking about using Hepburn, (unintelligible), some other system. You know, each language has to have quite strict rules.

The sort of consistency aspect of that needs adjusting.

Jim Galvin: So isn't that point the second bullet is making?

(Chris Dylan): Yes.

Jim Galvin: That's why the first bullet doesn't really make sense to me.

(Rudy Vonsnick): But again, it's a collection of all things we have together. And in that order, I would say the fact that there are different bullets don't mean that they are all those that we have to - are going to keep. It's just a list of - and we don't want to lose things that have been mentioned and brought up in order to validate our recommendation in the correct way.

(Chris Dylan): Yes. Looks as if that needs revisiting. We are very near the end of the meeting. I would just like to thank you. It's been a very interesting meeting. We discovered quite a few weak arguments which is extremely useful. We also discovered some strong arguments.

I would like to encourage both those in the room and those listening to this and reading the transcript later, particularly, the arguments are for making it mandatory. And a lot of them are looking quite weak at the moment. So if that is what you believe, please get involved with us because we do want to hear you. And that - and if - I think, apart from that, we just need to round off.

Man: Next meeting.

(Rudy Vonsnick): Next meeting, yes - next week. Normally...

(Chris Dylan): Don't play (unintelligible)...

(Rudy Vonsnick): Normally we have weekly calls, but maybe we have to skip a week and go into a two-week session.

(Chris Dylan): I think we should not meet next week. I think there's a lot of work to be done so we'll meet the week after and then we'll decide on frequency then. Anyways, thank you all very much indeed.

(Rudy Vonsnick): Thank you. Recording can be stopped.

(Crosstalk)

END